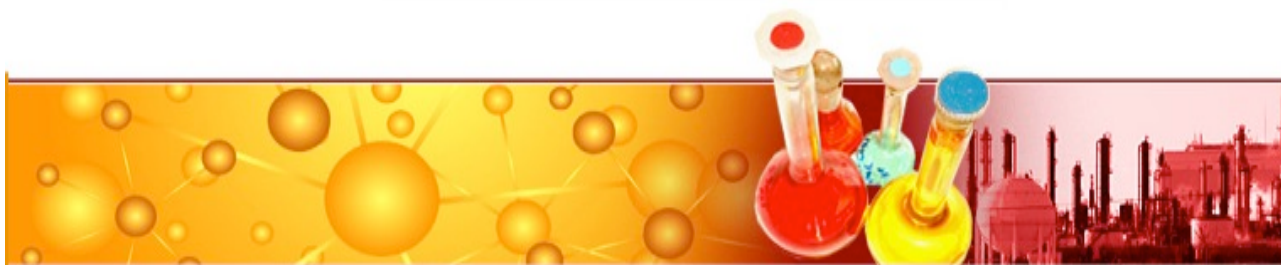


Chimie organique bio-orientée



Luc Patiny
EPFL - SCGC

1	Introduction	1
1.1	Références	1
1.2	Méthode d'évaluation	1
2	Analyse	2
2.1	Spectrométrie de masse.....	2
2.1.1	Isotopes naturels.....	2
2.1.2	Abondance relative: A, A+1, A+2.....	4
2.1.3	Définitions	4
2.1.4	Résolution	4
2.1.5	Précision	5
2.1.6	Formule brute.....	5
2.1.7	Application au proteomics.....	5
2.2	Degré d'insaturation.....	6
2.2.1	Isomères de constitution	6
2.3	Spectroscopie IR	7
2.4	Spectroscopie RMN.....	12
2.4.1	Déplacement chimique	12
2.4.2	Table d'incrément.....	16
2.4.3	Intégration	16
2.4.4	La multiplicité des pics : couplage indirect spin-spin	16
2.4.5	OH et NH	19
3	Acides aminés, peptides et protéines	20
3.1	Acides aminés	20
3.1.1	20 acides aminés principaux dans les protéines	20
3.1.2	Codification IUPAC des atomes.....	23
3.2	Peptides.....	26
3.2.1	Synthèse des polypeptides	26
3.2.2	Nécessité d'avoir des groupes protecteurs :	26
3.2.3	Activation du groupement carbonyle.....	27
3.2.4	Synthèse sélective	27
3.2.5	Synthèse peptidique sur support solide : Merrifield	27
3.3	Protéines	28
3.4	Diagramme de Ramachandran	31
4	Protein Data Bank (PDB)	33
4.1	Statistiques: 24-Feb-2014	33
4.2	Format PDB.....	34

4.2.1	Début et fin.....	34
4.2.2	Littérature / Expérience.....	34
4.2.3	Composition.....	34
4.2.4	Structure secondaire, site important.....	34
4.2.5	Atomes.....	35
4.2.6	Liaison.....	35
4.3	Exemple de PDB.....	35
4.4	Recherche dans la PDB.....	38
4.5	JSmol (http://www.jmol.org).....	39
4.5.1	Instructions utiles.....	39
5	Réactions "biocompatibles".....	41
5.1	Modification de résidus nucléophiles.....	41
5.1.1	Réactions sur RNH ₂	41
5.1.2	Réactions sur RSH.....	41
5.2	Réactions orthogonales (bioconjugaison).....	42
6	Peptide deformylase.....	43
6.1	Introduction.....	43
6.2	Recherche d'un substrat minimal.....	45
6.3	Recherche d'un inhibiteur.....	45
6.4	Détermination du site catalytique par RMN.....	46
6.5	Partie de reconnaissance (sélectivité), partie de réaction (catalyse).....	47
6.6	Structures RX et stabilisation du complexe protéine -inhibiteur.....	49
6.6.1	Ponts hydrogènes.....	49
6.6.2	Effets électrostatiques.....	50
6.6.3	Effets hydrophobes.....	51
7	Conclusions.....	52

1 Introduction

Le travail d'équipe est aujourd'hui nécessaire pour faire de nouvelles découvertes. Dans ces équipes, un mélange de biologistes, chimistes, pharmaciens, théoriciens, ... participent et contribuent, chacun dans leur domaine, à faire avancer le projet.

Un problème majeur apparaît néanmoins rapidement: la communication ! Il faut en effet pouvoir parler un langage commun pour pouvoir se comprendre.

La chimie organique est importante à plusieurs niveaux dans ce contexte pour les biologistes:

- la compréhension au niveau moléculaire des interactions entre macromolécules
- la compréhension et le design d'inhibiteur de protéines
- la synthèse ou la modification chimique de protéines

L'objectif de ce cours est de vous familiariser avec les atomes, les structures chimiques "possibles", les protéines, ...

Dans ce but différents sujets vont être abordés:

- l'analyse structurale, permettant de mieux manipuler les molécules
- la synthèse peptidique, pour permettre d'écrire une séquence d'acide aminé et d'expliquer le problème de chemosélectivité
- la "protein database", pour appréhender les protéines et les interactions stabilisants les structures secondaires
- la chiminformatique, permettant de visualiser et manipuler les protéines
- les interactions faibles, permettant d'expliquer pourquoi un inhibiteur se trouve à un endroit précis
- le "protein engineering", permettant de fixer sélectivement des molécules chimiques sur les protéines

1.1 Références

- Vollhardt, K Peter C, and Neil Eric Schore. *Organic chemistry: structure and function*. WH Freeman, 2011.
- Wikipedia
- Silverstein, Robert, and Francis Webster. *Spectrometric identification of organic compounds*. Wiley. com, 2006.

1.2 Méthode d'évaluation

Test écrit de 1h30 noté sur 4.

Travail de représentation d'une protéine noté sur 2.

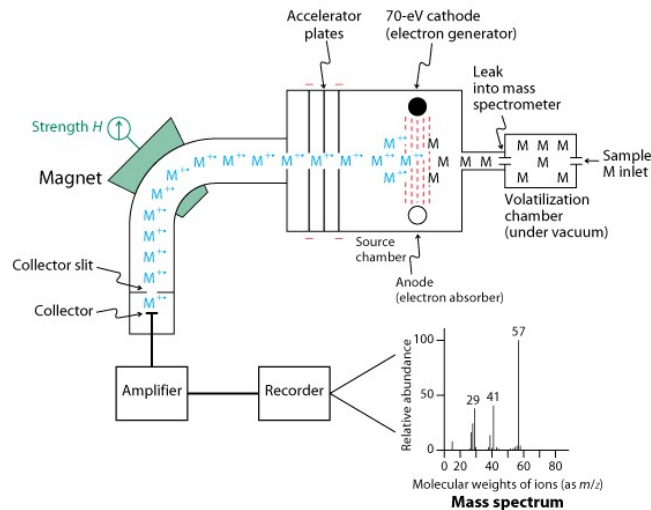
2 Analyse

2.1 Spectrométrie de masse

Spectromètre de masse: appareil permettant de séparer des **ions gazeux** en fonction du **rapport de leur masse à leur charge** (m/z). Une molécule peut être plusieurs fois chargées.

Principe général:

1. **Ionisation des molécules. Si une seule charge: M^+ ou $M+H^+$**
2. Accélération à une vitesse précise
3. Déviation dans un champ magnétique (Force de Lorentz, $q\vec{v} \wedge \vec{B} = \vec{F}$)
4. Détermination de m/z en fonction de la déviation



On mesure donc chaque molécule individuellement !!!

6 C Carbon-12 12.000000	13 C Carbon-13 13.003355	14 C Carbon-14 14.003241	16 O Oxygen 15.994915	17 O Oxygen-17 16.999131	19 F Fluorine 18.998403
28 Si Silicon 27.976927	29 Si Silicon-29 28.976495	30 Si Silicon-30 29.973762	32 S Sulfur 31.972071	35 Cl Chlorine 34.968852	37 Cl Chlorine-37 36.965903

Ne pas confondre le MW (masse moléculaire, Molecular Weight) et la masse de chacun des isotopes !

Exercice: j'observe une masse de 30.0460 dans le spectromètre: SiH_2 , CH_2O ou C_2H_6 ?

2.1.1 Isotopes naturels

Références:

- NIST: <http://www.nist.gov/pml/data/comp.cfm>
- ChemCalc <http://www.chemcalc.org/>

	Masse nominale	Abondance relative (%)	Abondance (%)	Masse isotopique (u)	Masse atomique
H	1	100	99.99	1.007825	1.008
	2	0.01	0.01	2.014101	
C	12	100	98.93	12.000000	12.011
	13	1.08	1.07	13.003354	
N	14	100	99.64	14.003074	14.007
	15	0.37	0.36	15.000108	
O	16	100	99.76	15.994914	15.999
	17	0.04	0.04	16.999131	
	18	0.20	0.20	17.999161	
P	31	100	100	30.973761	30.974
S	32	100	94.99	31.972071	32.065
	33	0.79	0.75	32.971458	
	34	4.47	4.25	33.967866	
	36	0.01	0.01	35.967080	
Si	28	100	92.22	27.976926	28.086
	29	5.08	4.69	28.976494	
	30	3.35	3.09	29.973770	
F	19	100	100	18.998403	18.998
Cl	35	100	75.76	34.968852	35.453
	37	32.00	24.24	36.965902	
Br	79	100	50.69	78.918337	79.903
	81	97.28	49.31	80.916290	
I	127	100	100	126.904473	126.904

2.1.2 Abondance relative: A, A+1, A+2

Element	A (masse)	A (%)	A+1 (masse)	A+1 (%)	A+2 (masse)	A+2 (%)	A+3 (masse)	A+3 (%)
H	1	100	2	0.01				
C	12	100	13	1.08				
N	14	100	15	0.37				
O	16	100	17	0.04	18	0.20		
P	31	100						
S	32	100	33	0.79	34	4.47	35	0.01
Si	28	100	29	5.08	30	3.35		
F	19	100						
Cl	35	100			37	32.00		
Br	79	100			81	97.28		
I	127	100						

Exercice: Dessiner le spectre de masse de Cl, Br, S, N, ...

2.1.3 Définitions

Isotopomères: 2 molécules ayant le même nombre de chacun des isotopes à des positions différentes

Composition isotopique majoritaire (isotopologue majoritaire): composition isotopique la plus abondante

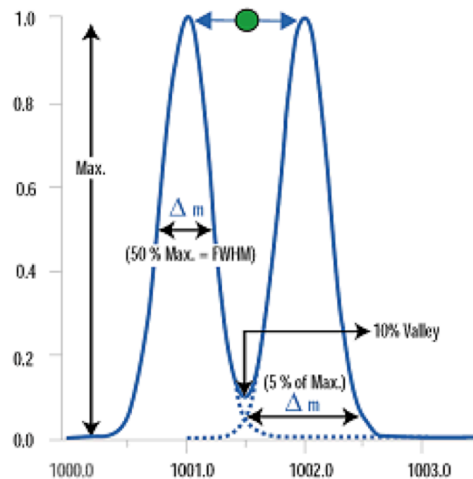
Distribution isotopique: ensemble des isotopomères avec leur masse et intensité relative respective

Masse monoisotopique: masse calculée en utilisant les masses exactes des isotopes majoritaires de chaque élément

Exercice: Dessiner la distribution isotopique de Cl₂, Br₂, BrCl

2.1.4 Résolution

Masse divisée par la largeur du pic à une hauteur relative donnée (x) par rapport à la hauteur totale du pic. En général, **50% (FWHM : Full Width of the peak at Half its Maximum height)**, 5% ou 0.5%.



2.1.5 Précision

La différence qui est observée entre la masse théorique et la masse mesurée

$$\Delta m = m_{\text{theory}} - m_{\text{experimental}}$$

$$\text{ppm} = \frac{\Delta m}{m_{\text{experimental}}} 10^6$$

En fonction de la précision on va pouvoir décider du nombre de formule brute possible pour une masse monoisotopique donnée.

2.1.6 Formule brute

- Au départ de la distribution isotopique on peut déterminer la présence de Cl, Br et S
- Au départ d'une masse isotopique de haute précision (masse « exacte », 4 décimales) on peut déterminer la formule brute. On va considérer l'erreur expérimentale de plus ou moins 1 sur la troisième décimale.

Démonstration de ChemCalc

Problème de plusieurs charges

Problème de la masse de l'électron et du H⁺

Problème de la précision

2.1.7 Application au proteomics

Proteomics can be defined as the qualitative and quantitative comparison of proteomes under different conditions to further unravel biological processes.

L'objectif du proteomics est de déterminer la composition en protéines dans un organisme particulier à un moment particulier dans des conditions particulières.

Un moyen de déterminer la séquence d'une protéine est de:

1. séparer le mélange pour avoir une protéine
2. cliver avec la trypsin
3. analyser les différents peptides obtenus, par exemple par masse exacte
4. comparer avec des bases de donnée de protéines existantes

La trypsin va cliver exclusivement en C-terminal des arginines et lysines. Exemple sur la cyclophilin :

```
MET VAL ASN PRO THR VAL PHE PHE ASP ILE ALA VAL ASP
GLY GLU PRO LEU GLY ARG VAL SER PHE GLU LEU PHE ALA
ASP LYS VAL PRO LYS THR ALA GLU ASN PHE ARG ALA LEU
```


SER THR GLY GLU **LYS** GLY PHE GLY TYR **LYS** GLY SER CYS
 PHE HIS **ARG** ILE ILE PRO GLY PHE MET CYS GLN GLY GLY
 ASP PHE THR **ARG** HIS ASN GLY THR GLY GLY **LYS** SER ILE
 TYR GLY GLU **LYS** PHE GLU ASP GLU ASN PHE ILE LEU **LYS**
 HIS THR GLY PRO GLY ILE LEU SER MET ALA ASN ALA GLY
 PRO ASN THR ASN GLY SER GLN PHE PHE ILE CYS THR ALA
LYS THR GLU TRP LEU ASP GLY **LYS** HIS VAL VAL PHE GLY
LYS VAL **LYS** GLU GLY MET ASN ILE VAL GLU ALA MET GLU
ARG PHE GLY SER **ARG** ASN GLY **LYS** THR SER **LYS** **LYS** ILE
 THR ILE ALA ASP CYS GLY GLN LEU GLU

On trouvera donc des séquences du type :

- HPheGlySerArgOH 465.233583
- HAsnGlyLysOH 317.16992
- HThrSerLysOH 334.185236
- ...

2.2 Degré d'insaturation

Le degré d'insaturation est la somme des nombres de cycles et de liaisons π présents dans la molécule.

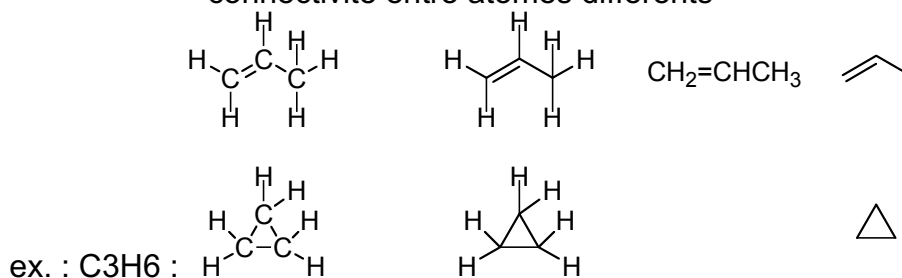
Degré d'insaturation = $(2 n_C - n_X - n_H + n_N + 2) / 2$ (on ne tient pas compte des O et S).

Exemples :

- C_3H_6
- C_7H_8
- $C_6H_{15}N$
- $C_4H_8O_2$

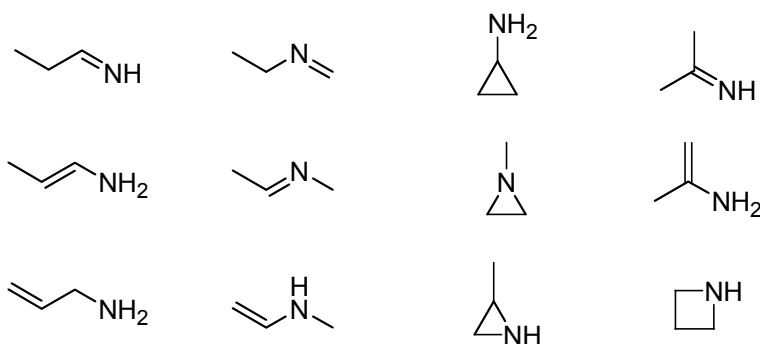
2.2.1 Isomères de constitution

Isomères de constitution ou **isomères de structure** : même composition d'atome mais connectivité entre atomes différents



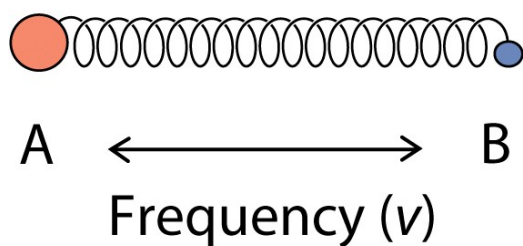
Vite très compliqué :

C_3H_7N



2.3 Spectroscopie IR

Les molécules absorbent la lumière infrarouge et subissent une excitation vibrationnelle.



Loi de Hooke.

$$\bar{\nu} = k \sqrt{f \frac{m_1 + m_2}{m_1 m_2}}$$

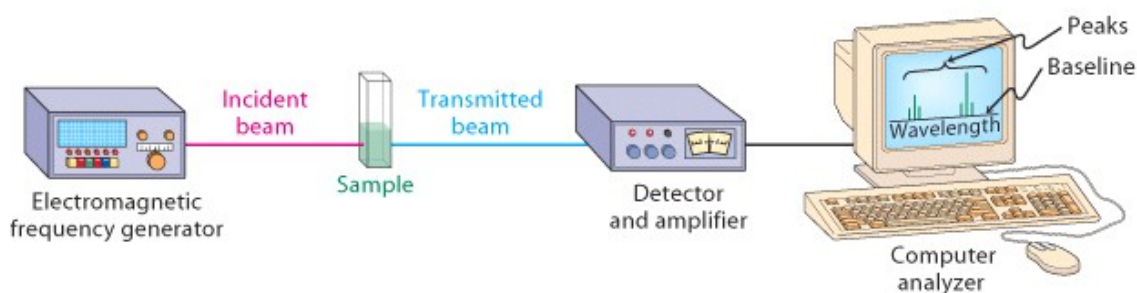
où :

$\bar{\nu}$: fréquence vibrationnelle en nombre d'ondes (cm^{-1})

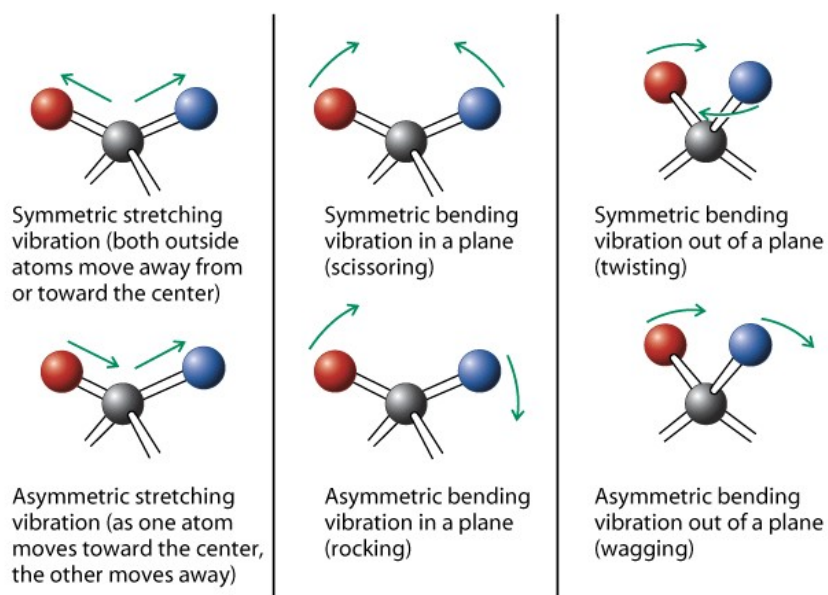
k = constante

f = constante de force, indiquant la raideur du ressort (de la liaison)

m_1, m_2 = valeurs des masses attachées au ressort (masses des atomes unis)

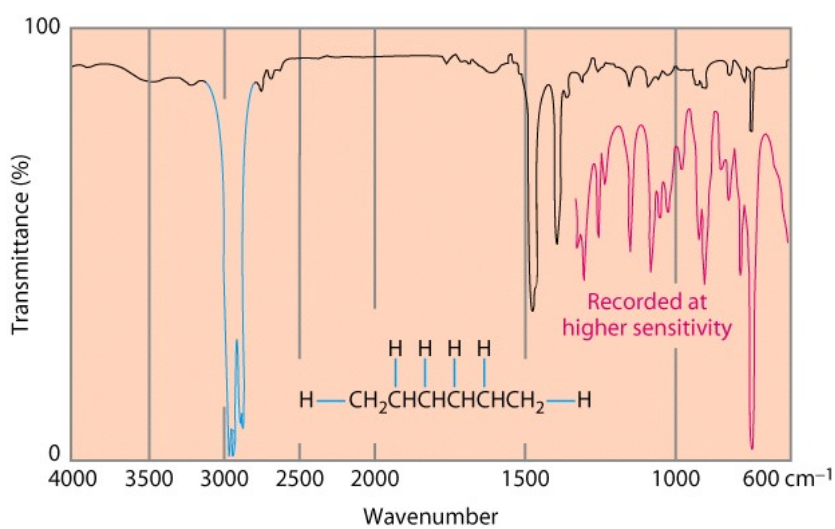
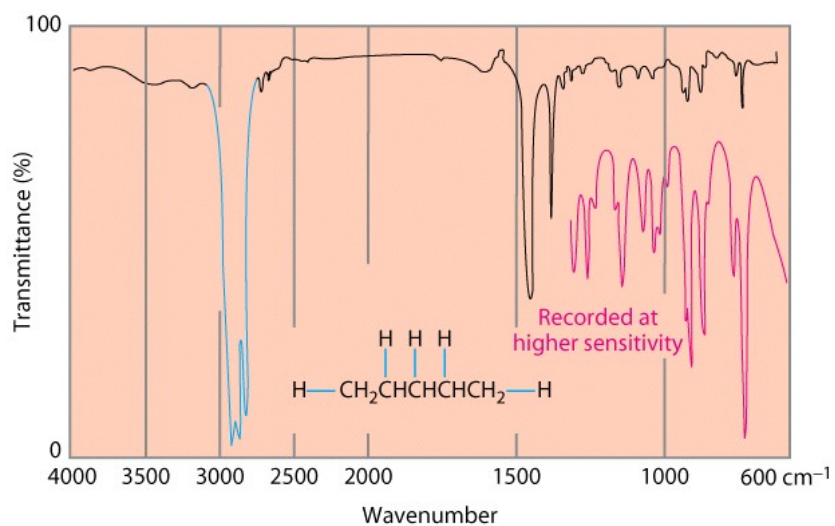


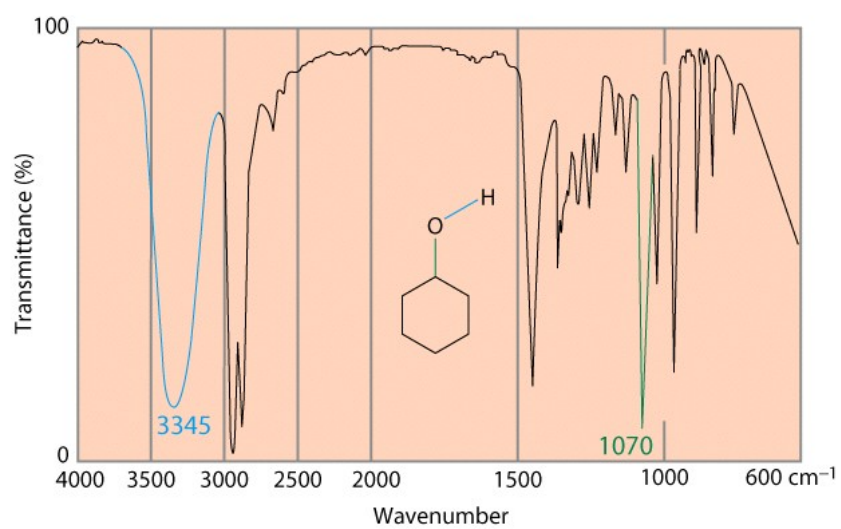
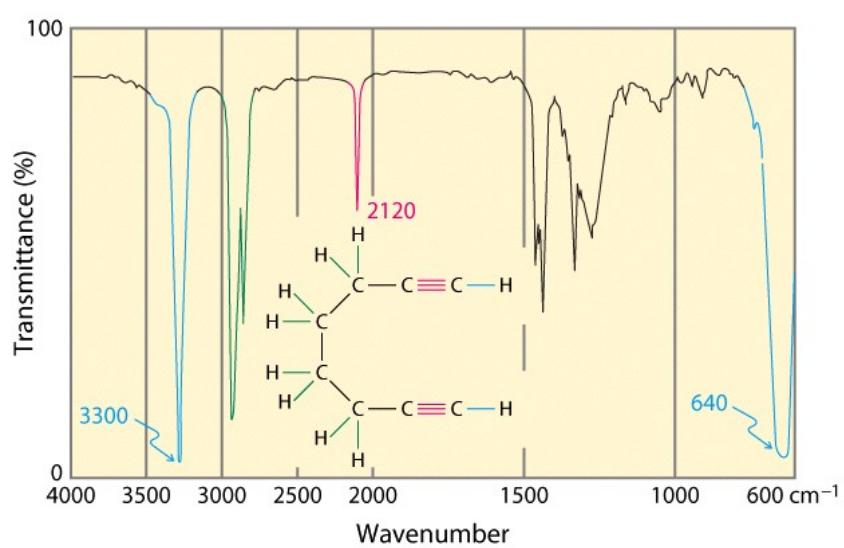
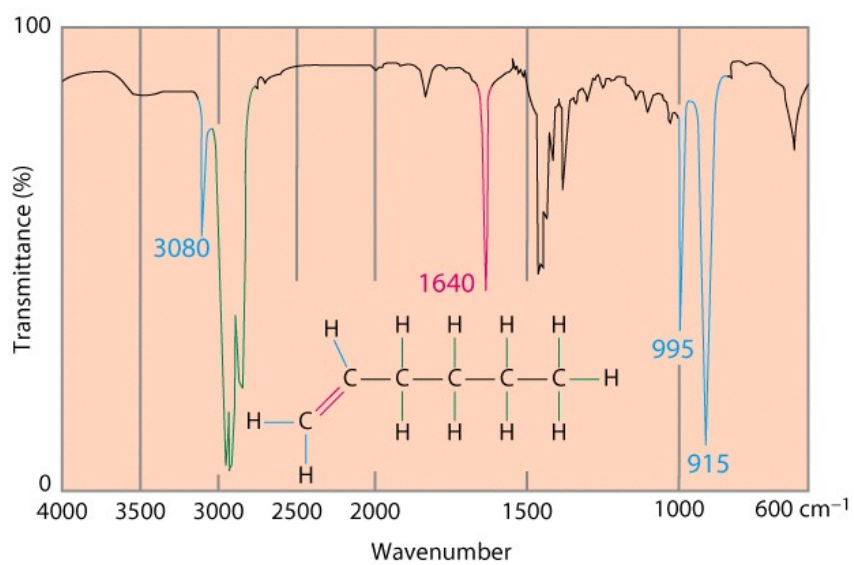
Pas seulement des élongations mais également des vibrations de déformation angulaire (souvent de plus faible intensité).

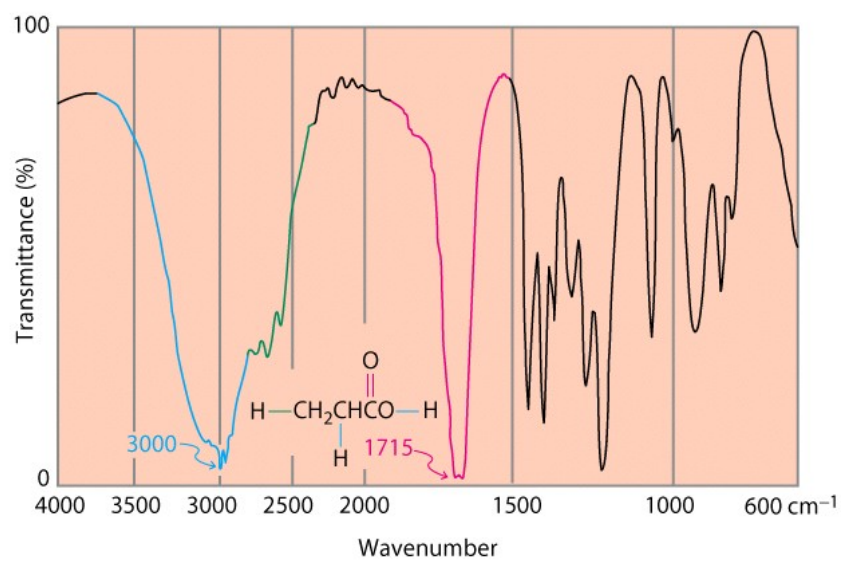
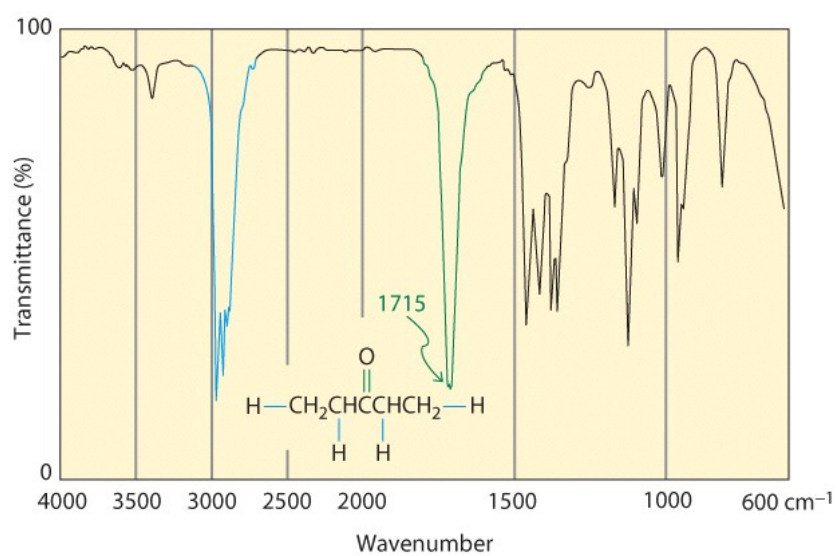
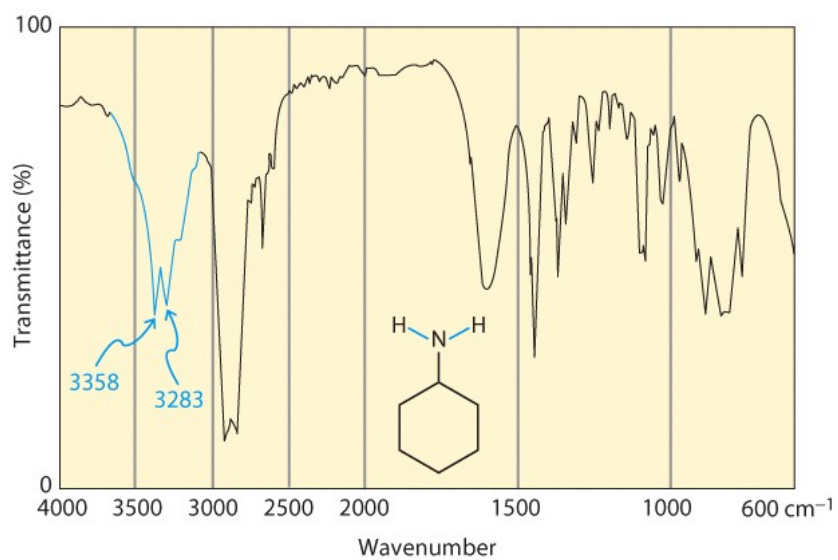


Spectres infrarouges : 600 à 4000 cm^{-1}

Zone de 600 à 1300 cm^{-1} = "fingerprint".





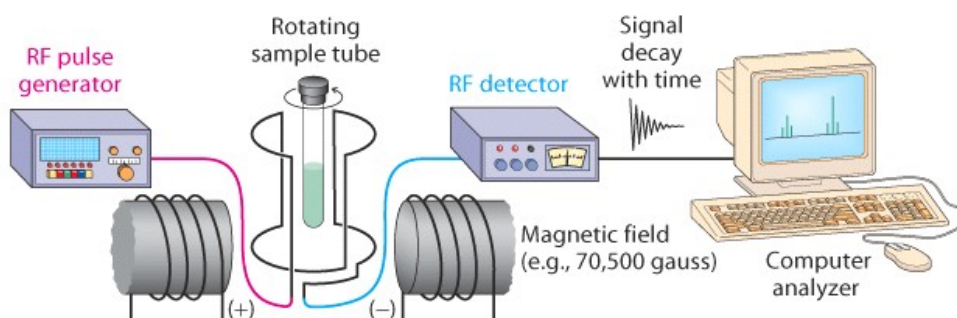


Les groupes fonctionnels donnent lieu à des absorptions infrarouges caractéristiques.

TABLE 11-4 Characteristic Infrared Stretching Wavenumber Ranges of Organic Molecules

Bond or functional group	$\tilde{\nu}$ (cm^{-1})	Bond or functional group	$\tilde{\nu}$ (cm^{-1})
RO—H (alcohols)	3200–3650	RC \equiv N (nitriles)	2220–2260
$\begin{array}{c} \text{O} \\ \parallel \\ \text{RCO—H} \end{array}$ (carboxylic acids)	2500–3300	$\begin{array}{c} \text{O} \quad \text{O} \\ \parallel \quad \parallel \\ \text{RCH, RCR}' \end{array}$ (aldehydes, ketones)	1690–1750
R ₂ N—H (amines)	3250–3500	$\begin{array}{c} \text{O} \\ \parallel \\ \text{RCOR}' \end{array}$ (esters)	1735–1750
RC \equiv C—H (alkynes)	3260–3330	$\begin{array}{c} \text{O} \\ \parallel \\ \text{RCOH} \end{array}$ (carboxylic acids)	1710–1760
$\begin{array}{c} \diagup \quad \diagdown \\ \text{C}=\text{C} \\ \diagdown \quad \diagup \\ \text{H} \end{array}$ (alkenes)	3050–3150	$\begin{array}{c} \diagup \quad \diagdown \\ \text{C}=\text{C} \\ \diagdown \quad \diagup \end{array}$ (alkenes)	1620–1680
$\begin{array}{c} \\ \text{—C—H} \\ \end{array}$ (alkanes)	2840–3000	$\begin{array}{c} \\ \text{RC—OR}' \\ \end{array}$ (alcohols, ethers)	1000–1260
RC \equiv CH (alkynes)	2100–2260		

2.4 Spectroscopie RMN



Certains noyaux atomiques se comportent comme des toupies en rotation : ils possèdent un état de spin nucléaire (spin I différent de 0). Par exemple ^1H , ^{13}C , ^{15}N ($I=1/2$).

Ces noyaux peuvent prendre dans un champ magnétique différentes énergies :

$$E = - m \gamma \hbar B_0$$

où

γ : rapport gyromagnétique (caractéristique du noyau)

m : $-I, -I+1, \dots, +I$

\hbar : constante de Planck / 2π

B_0 : champ magnétique externe

Ainsi, pour le proton, deux niveaux énergétiques sont possibles et la spectroscopie RMN va observer la transition entre ces deux niveaux. On peut dès lors calculer l'énergie de cette transition ainsi que la différence de population entre ces deux niveaux d'énergie.

2.4.1 Déplacement chimique

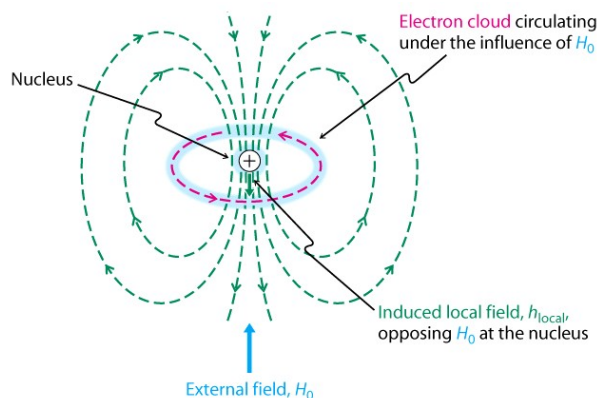
Si l'ensemble des protons passait d'un état à l'autre en absorbant la même quantité d'énergie, la spectroscopie RMN serait assez inutile. Heureusement, l'énergie absorbée dépend de l'environnement du proton.

Les déplacements chimiques des noyaux vont varier en fonction de la densité électronique autour de ce noyau. En effet, les électrons « blindent » le noyau de sorte que le champ magnétique réel au niveau du noyau est généralement plus faible que le champ magnétique externe.

$$B_{\text{eff}} = B_0 - \sigma B_0 = (1 - \sigma) B_0$$

où σ est la constante de blindage

Cet effet de blindage est un phénomène physique bien connu (Lenz, 1804-1865). Sous l'effet d'un champ magnétique externe, les électrons vont circuler de sorte à générer un champ magnétique induit de direction opposée.



Généralement, les spectres RMN sont référencés par rapport au tétraméthylsilane (TMS) et les différentes transitions sont représentées sur une échelle δ en ppm. Cette échelle est définie comme :

$$\delta = \frac{\nu_{\text{échantillon}} - \nu_{\text{référence}}}{\nu_{\text{référence}}} 10^6$$

Exemple sur un spectromètre travaillant à 90 MHz :

- TMS : 90 000 000 Hz : 0 ppm
- CH₃Br : 90 000 237 Hz : 2.63 ppm
- CH₂Br₂ : 90 000 441 Hz : 4.90 ppm
- CHBr₃ : 90 000 614 Hz : 6.82 ppm

Pour le proton, il est cependant difficile de prévoir les déplacements chimiques. En général on pourra cependant tenir compte de l'électronégativité des substituants, de l'hybridation de l'atome de carbone, de cycles aromatiques, ... Il existe cependant des tables qui vont permettre d'avoir des « fourchettes ». Ainsi on saura qu'un proton de type alcane sortira aux environs de 1 ppm, un proton en α d'un oxygène aux environs de 4 ppm et un proton aromatique aux alentours de 7 ppm.

Dans les éthènes, éthyne et noyaux aromatiques il y a une circulation électronique au niveau des liaisons qui provoque un champ magnétique local plus important.

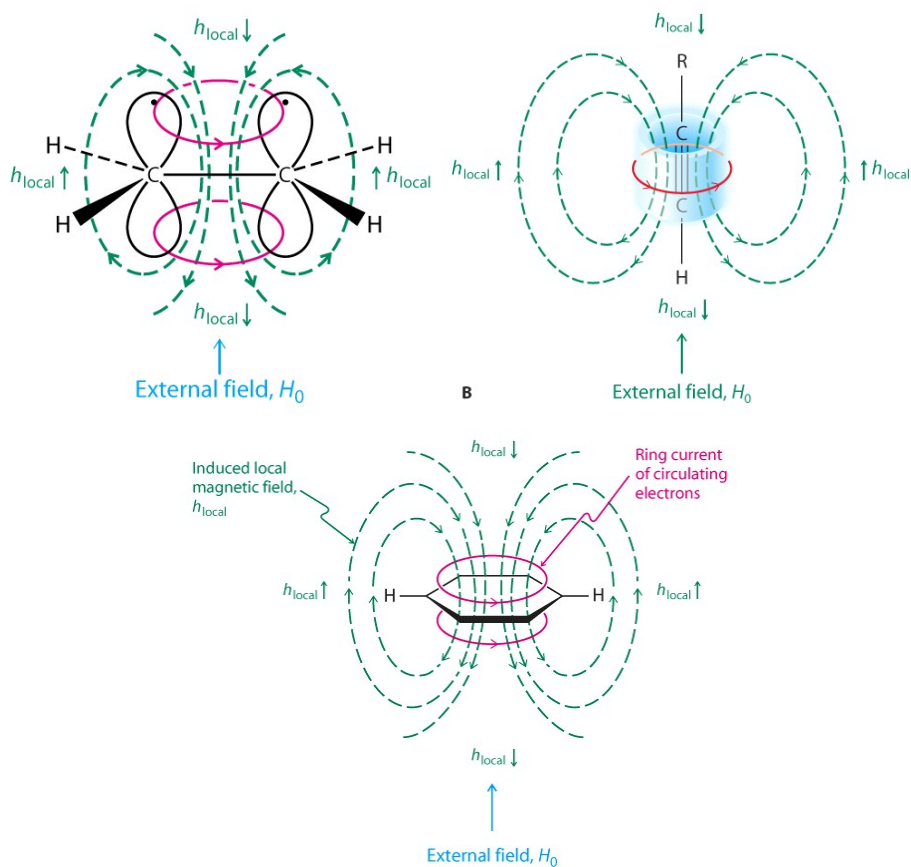


TABLE 10-2 Typical Hydrogen Chemical Shifts in Organic Molecules

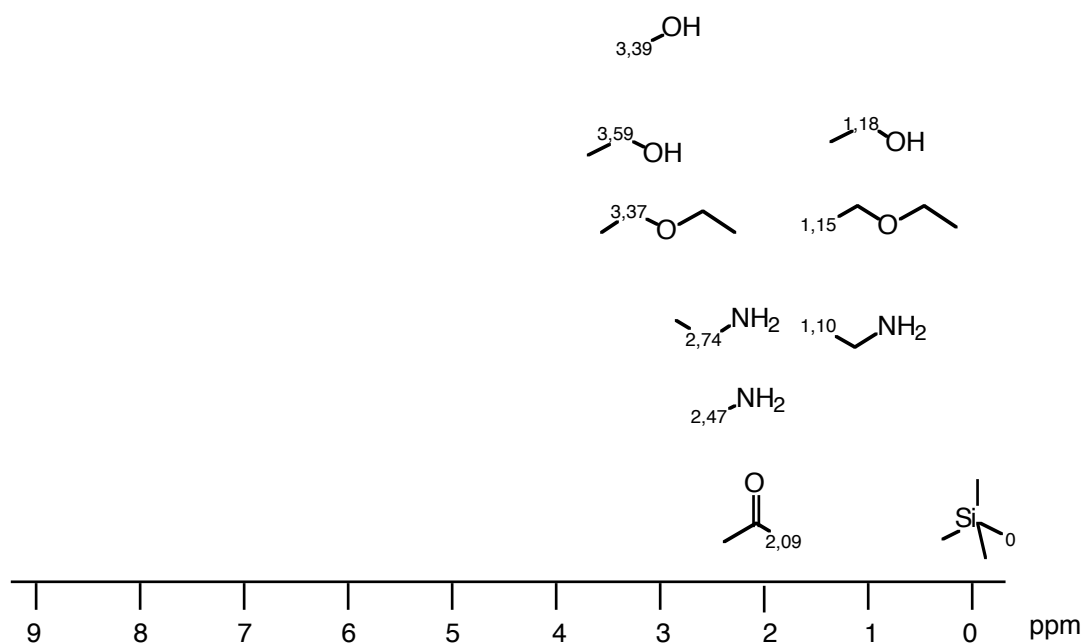
Type of hydrogen ^a	Chemical shift δ in ppm	
Primary alkyl, RCH_3	0.8–1.0	Alkane and alkane like hydrogens
Secondary alkyl, $\text{RCH}_2\text{R}'$	1.2–1.4	
Tertiary alkyl, R_3CH	1.4–1.7	
Allylic (next to a double bond), $\text{R}_2\text{C}=\text{C}(\text{CH}_3)\text{R}'$	1.6–1.9	Hydrogens adjacent to unsaturated functional groups
Benzylic (next to a benzene ring), ArCH_2R	2.2–2.5	
Ketone, RC(=O)CH_3	2.1–2.6	

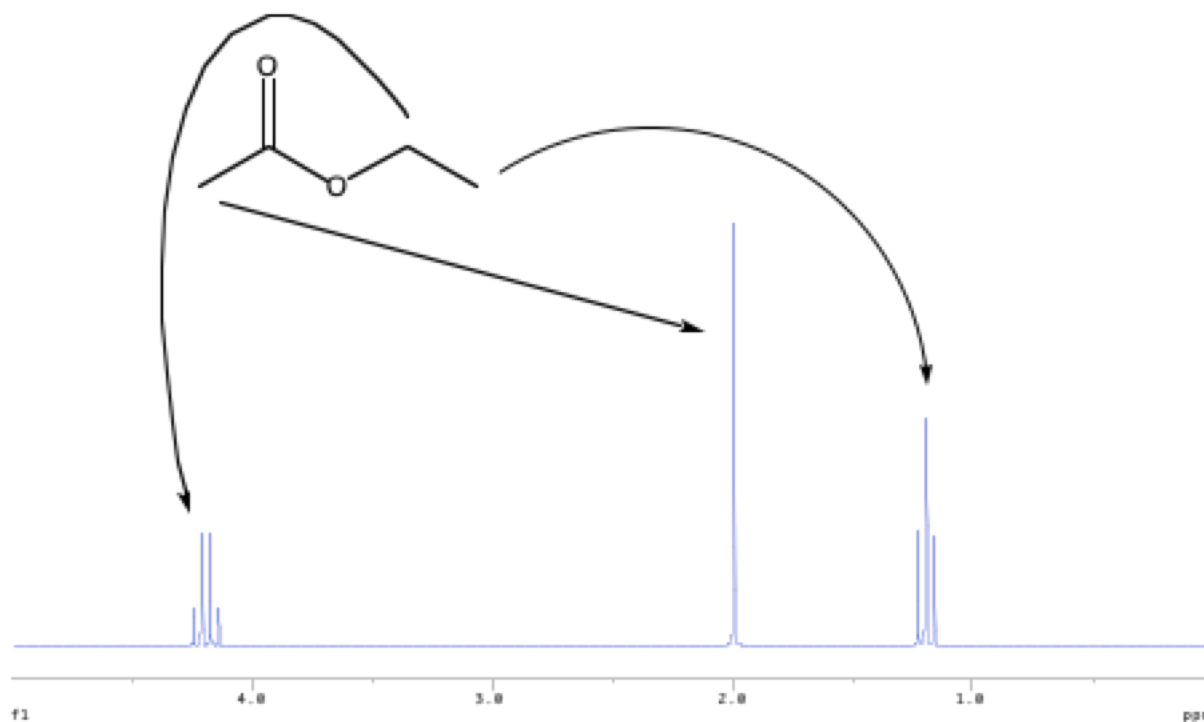
^aR, R', alkyl groups; Ar, aromatic group (not argon).

TABLE 10-2 Typical Hydrogen Chemical Shifts in Organic Molecules

Type of hydrogen ^a	Chemical shift δ in ppm	
Alkyne, $\text{RC}\equiv\text{CH}$	1.7–3.1	
Chloroalkane, RCH_2Cl	3.6–3.8	Hydrogens adjacent to electronegative atoms
Bromoalkane, RCH_2Br	3.4–3.6	
Iodoalkane, RCH_2I	3.1–3.3	
Ether, $\text{RCH}_2\text{OR}'$	3.3–3.9	
Alcohol, RCH_2OH	3.3–4.0	
Terminal alkene, $\text{R}_2\text{C}=\text{CH}_2$	4.6–5.0	Alkene hydrogens
Internal alkene, $\text{R}_2\text{C}=\underset{\text{R}'}{\text{CH}}$	5.2–5.7	
Aromatic, ArH	6.0–9.5	
Aldehyde, $\text{RCH}=\overset{\text{O}}{\parallel}$	9.5–9.9	
Alcoholic hydroxy, ROH	0.5–5.0	(variable)
Thiol, RSH	0.5–5.0	(variable)
Amine, RNH_2	0.5–5.0	(variable)

^aR, R', alkyl groups; Ar, aromatic group (not argon).





2.4.2 Table d'incréments

CH et CH₂

éthène

benzène

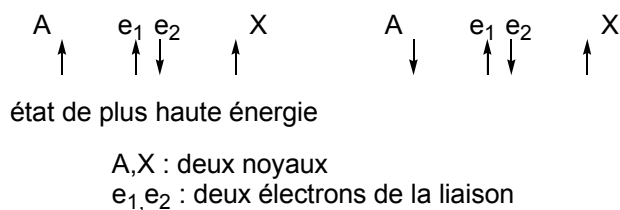
2.4.3 Intégration

En proton la surface des pics est proportionnelle au nombre d'hydrogène

2.4.4 La multiplicité des pics : couplage indirect spin-spin

L'autre information importante est la multiplicité des signaux. En effet, comme on peut le constater sur le spectre de l'acétate d'éthyle, le CH₂ n'apparaît pas comme un seul signal mais comme un triplet.

La multiplicité des signaux est due à la présence de noyaux possédant un spin non nul liés par des liaisons. Le spin nucléaire (A) est en interaction avec les électrons de la liaison qui seront en interaction avec le spin nucléaire (X). Si A et X sont antiparallèle l'énergie est plus basse que si ils sont parallèle.



2.4.4.1 Influence du rapprochement des pics : effet de toit

Système AX : deux protons couplés distants de plus de 10 fois la constante de couplage

Système AB : deux protons distants de moins de 10 fois la constante de couplage

Système A₂ : deux protons avec le même déplacement chimique

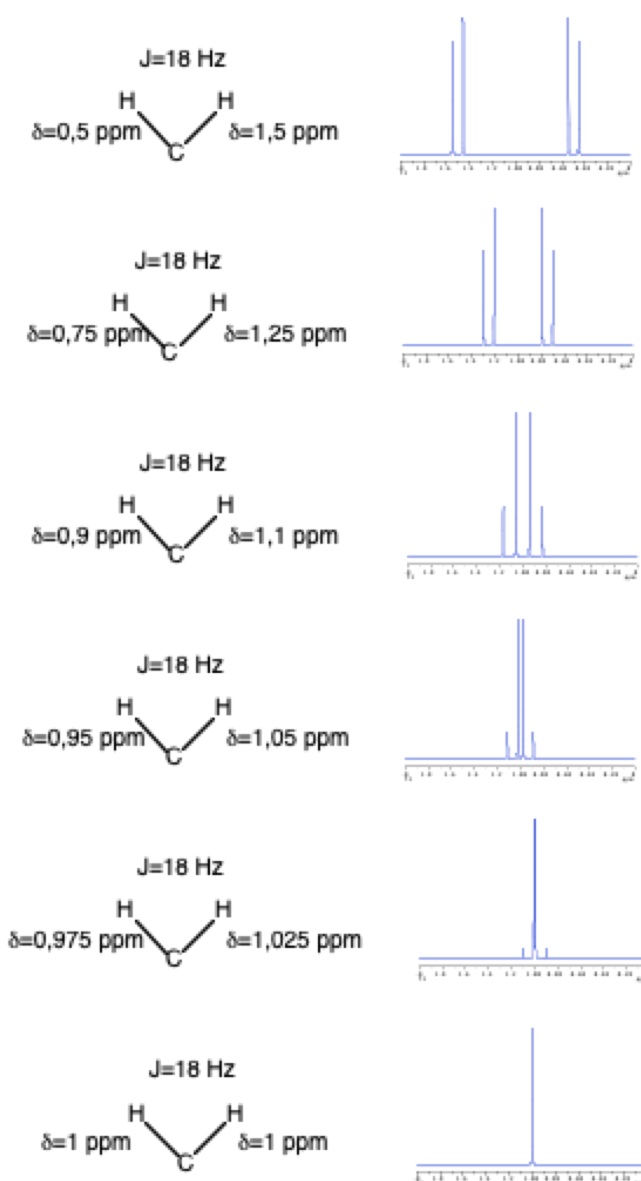
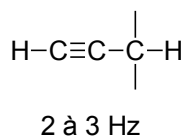
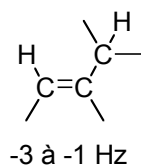
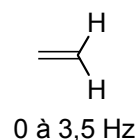
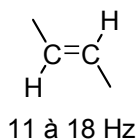
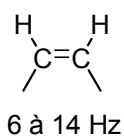
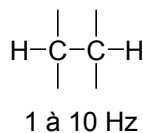
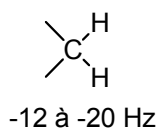


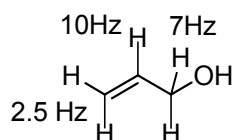
Schéma 1 : exemples de spectres protons à 200 MHz illustrant la structure de deux doublets en fonction du déplacement chimique. Du système AX au système A₂ via le système AB.

2.4.4.2 Constantes de couplage en proton

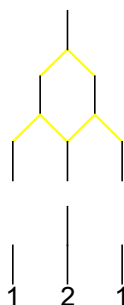
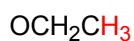
Dans le cas des protons, la multiplicité va être liée au nombre de protons présents à une distance de deux 2J , trois 3J ou éventuellement quatre (4J) liaisons et qui ont un déplacement chimique différent.



trans : 16 Hz



Constante de couplage pour un CH₂ libre : **environ 7Hz**



s	1
d	1 1
t	1 2 1
q	1 3 3 1
quint	1 4 6 4 1
hex	1 5 10 10 5 1
hept	1 6 15 20 15 6 1
oct	1 7 21 35 35 21 7 1
non	1 8 28 56 70 56 28 8 1

- Exemple de quadruplet
- Exemple d'un dd ($J=16$ Hz et 5 Hz)
- Exemple d'un ddd (16, 5, 1)
- Exemple d'un dt (16, 4) (correspond à un ddd 16 4 4)
- Exemple d'un dt (8, 4)

2.4.4.3 Système plus complexe

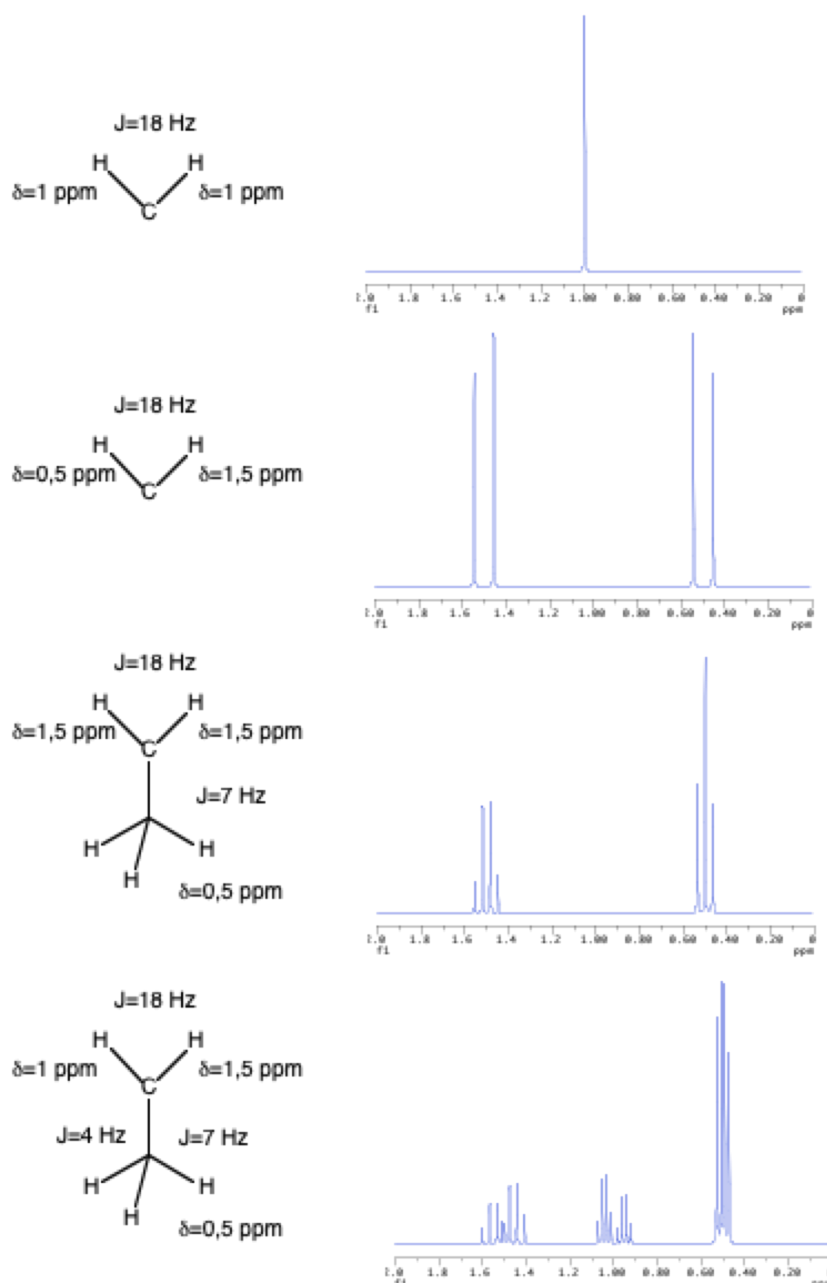


Schéma 2 : exemples de spectres protons à 200 MHz illustrant les constantes de couplage

2.4.5 OH et NH

Les protons sur des OH et NH ont la possibilité d'échanger. La conséquence est de ne plus observer les couplages et le pic du proton porté par N ou O devient large.

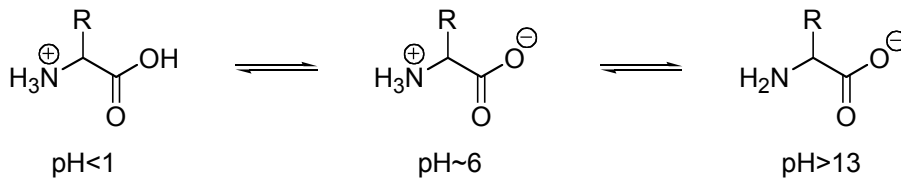
3 Acides aminés, peptides et protéines

Objectifs :

- visualiser et comprendre la structure des protéines
- visualiser un site catalytique et un inhibiteur

3.1 Acides aminés

- A la fois une fonction acide carboxylique et un groupe amine.
- Dans la nature, généralement des α -acides aminés (2-aminoacides).
- Configuration S dans la plupart des cas (une exception).
- Formule générale $RCH(NH_2)COOH$
- Il existe un code à 3 lettres et un code à une lettre.
- Les acides aminés sont à la fois acides et basiques. Dans le cas de la glycine, à pH neutre la forme majoritaire est le zwitterion. A pH acide, amine protonée, à pH basique, acide déprotoné.

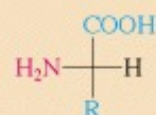





- Au point isoélectrique (**pI**) les charges se neutralisent par compensation.

3.1.1 20 acides aminés principaux dans les protéines

TABLE 26-1 Natural (2S)-Amino Acids							
R	Name	Three-letter code	One-letter code	pK _a of α -COOH	pK _a of α -NH ₃ ⁺	pK _a of acidic function in R	Isoelectric point, pI
H	Glycine	Gly	G	2.3	9.6	—	6.0
Alkyl group							
CH ₃	Alanine	Ala	A	2.3	9.7	—	6.0
CH(CH ₃) ₂	Valine ^a	Val	V	2.3	9.6	—	6.0
CH ₂ CH(CH ₃) ₂	Leucine ^a	Leu	L	2.4	9.6	—	6.0
CHCH ₂ CH ₃ (S) CH ₃	Isoleucine ^a	Ile	I	2.4	9.6	—	6.0
H ₂ C-	Phenylalanine ^a	Phe	F	1.8	9.1	—	5.5
	Proline	Pro	P	2.0	10.6	—	6.3

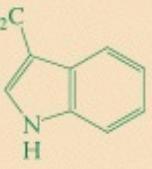
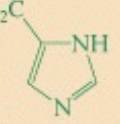
Continued

TABLE 26-1 Natural (2*S*)-Amino Acids

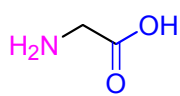
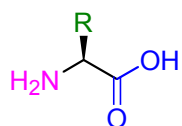
R	Name	Three-letter code	One-letter code	p <i>K</i> _a of α-COOH	p <i>K</i> _a of α- ⁺ NH ₃	p <i>K</i> _a of acidic function in R	Isoelectric point, pI
Hydroxy containing							
CH ₂ OH	Serine	Ser	S	2.2	9.2	—	5.7
CHOH (R) CH ₃	Threonine ^a	Thr	T	2.1	9.1	—	5.6
H ₂ C—  —OH	Tyrosine	Tyr	Y	2.2	9.1	10.1	5.7
Amino containing							
 CH ₂ CNH ₂	Asparagine	Asn	N	2.0	8.8	—	5.4
 CH ₂ CH ₂ CNH ₂	Glutamine	Gln	Q	2.2	9.1	—	5.7
(CH ₂) ₄ NH ₂	Lysine ^a	Lys	K	2.2	9.0	10.5 ^c	9.7
(CH ₂) ₃ NHCNH ₂	Arginine ^a	Arg	R	2.2	9.0	12.5 ^c	10.8

Continued

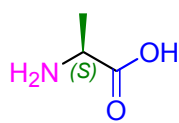
TABLE 26-1 Natural (2*S*)-Amino Acids (continued)

R	Name	Three-letter code	One-letter code	p <i>K</i> _a of α-COOH	p <i>K</i> _a of α- ⁺ NH ₃	p <i>K</i> _a of acidic function in R	Isoelectric point, pI
Amino containing (continued)							
H ₂ C— 	Tryptophan ^a	Trp	W	2.8	9.4	—	5.9
H ₂ C— 	Histidine ^a	His	H	1.8	9.2	6.1 ^c	7.6
Mercapto or sulfide containing							
CH ₂ SH	Cysteine ^a	Cys	C	2.0	10.3	8.2	5.1
CH ₂ CH ₂ SCH ₃	Methionine ^a	Met	M	2.3	9.2	—	5.7
Carboxy containing							
CH ₂ COOH	Aspartic acid	Asp	D	1.9	9.6	3.7	2.8
CH ₂ CH ₂ COOH	Glutamic acid	Glu	E	2.2	9.7	4.3	3.2

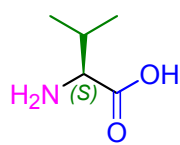
^aEssential amino acids. ^bEntire structure. ^cp*K*_a of conjugate acid. ^dThe stereocenter is *R* because the CH₂SH substituent has higher priority than the COOH group.



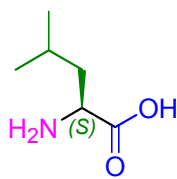
Glycine
Gly
G



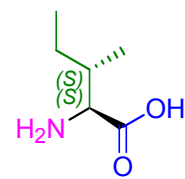
Alanine
Ala
A



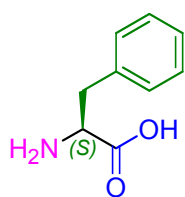
Valine
Val
V



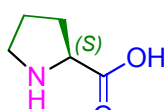
Leucine
Leu
L



Isoleucine
Ile
I



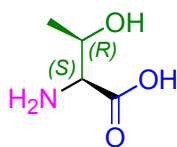
Phenylalanine
Phe
F



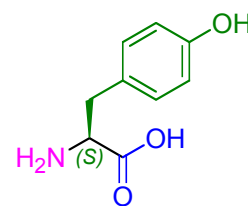
Proline
Pro
P



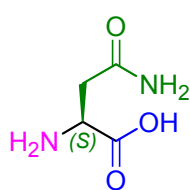
Serine
Ser
S



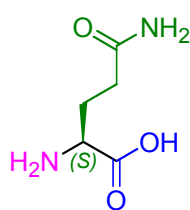
Threonine
Thr
T



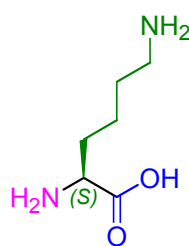
Tyrosine
Tyr
Y



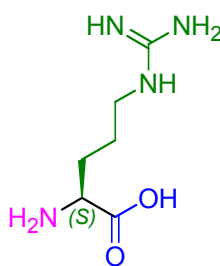
Asparagine
Asn
N



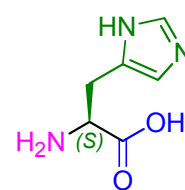
Glutamine
Gln
Q



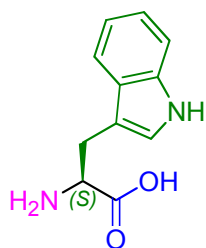
Lysine
Lys
K



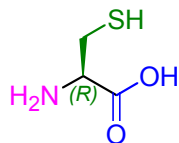
Arginine
Arg
R



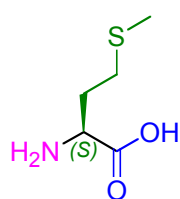
Histidine
His
H



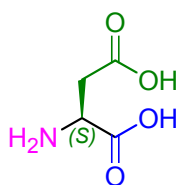
Tryptophan
Trp
W



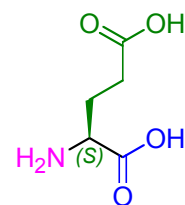
Cysteine
Cys
C



Methionine
Met
M

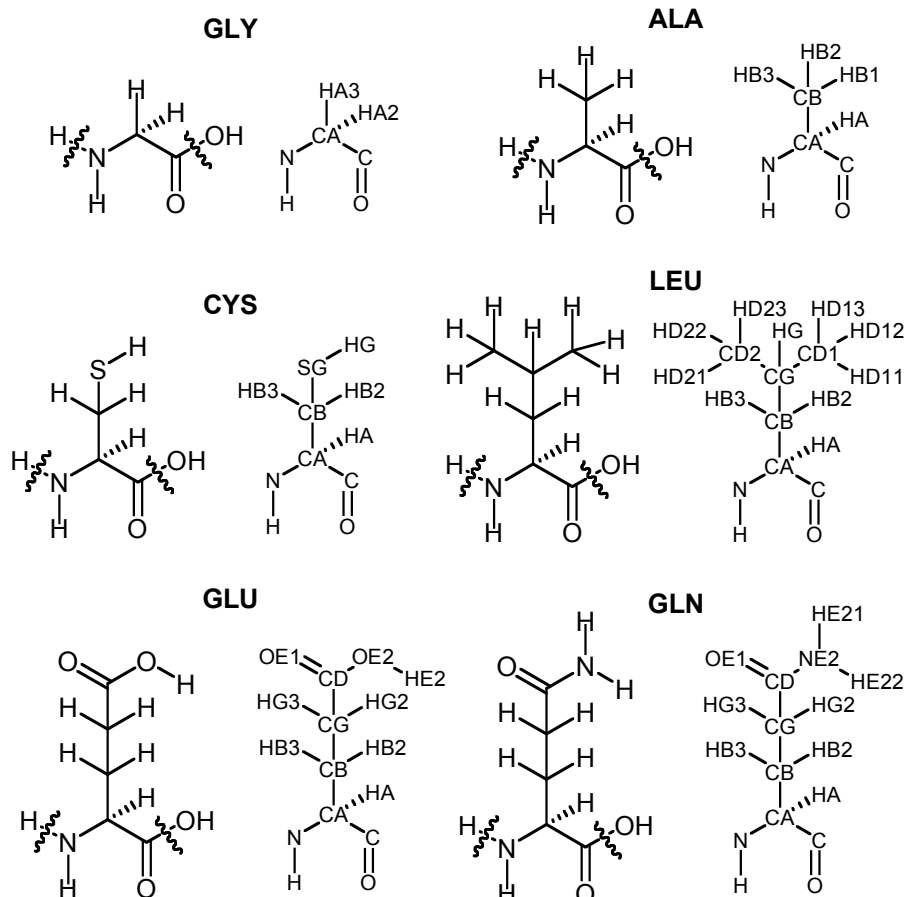


Aspartic acid
Asp
D



Glutamic acid
Glu
E

3.1.2 Codification IUPAC des atomes



Amino Acid	IUPAC	PDB
ALA	H	H
ALA	HA	HA
ALA	HB1	1HB
ALA	HB2	2HB
ALA	HB3	3HB
ALA	C	C
ALA	CA	CA
ALA	CB	CB
ALA	N	N
ALA	O	O
ARG	H	H
ARG	HA	HA
ARG	HB2 (pro-R)	2HB
ARG	HB3 (pro-S)	3HB
ARG	HG2 (pro-S)	2HG
ARG	HG3 (pro-R)	3HG
ARG	HD2 (pro-S)	2HD
ARG	HD3 (pro-R)	3HD
ARG	HE	HE
ARG	HH11 (Z)	1HH1
ARG	HH12 (E)	2HH1
ARG	HH21 (Z)	1HH2
ARG	HH22 (E)	2HH2
ARG	C	C
ARG	CA	CA
ARG	CB	CB
ARG	CG	CG
ARG	CD	CD
ARG	CZ	CZ
ARG	N	N
ARG	NE	NE
ARG	NH1 (Z)	NH1
ARG	NH2 (E)	NH2
ARG	O	O
ASP	H	H
ASP	HA	HA
ASP	HB2 (pro-S)	2HB

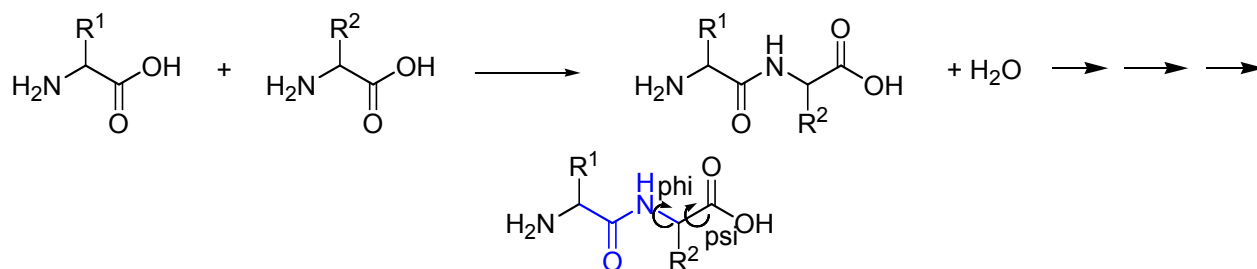
ASP	HB3 (pro-R)	3HB
ASP	HD2	HD2
ASP	C	C
ASP	CA	CA
ASP	CB	CB
ASP	CG	CG
ASP	N	N
ASP	O	O
ASP	OD1	OD1
ASP	OD2	OD2
ASN	H	H
ASN	HA	HA
ASN	HB2 (pro-S)	2HB
ASN	HB3 (pro-R)	3HB
ASN	HD21 (E)	1HD2
ASN	HD22 (Z)	2HD2
ASN	C	C
ASN	CA	CA
ASN	CB	CB
ASN	CG	CG
ASN	N	N
ASN	ND2	ND2
ASN	O	O
ASN	OD1	OD1
CYS	H	H
CYS	HA	HA
CYS	HB2 (pro-S)	2HB
CYS	HB3 (pro-R)	3HB
CYS	HG	HG
CYS	C	C
CYS	CA	CA
CYS	CB	CB
CYS	N	N
CYS	O	O
CYS	SG	SG
GLU	H	H
GLU	HA	HA
GLU	HB2 (pro-R)	2HB

GLU	HB3	(pro-S)	3HB	LEU	HB3	(pro-S)	3HB
GLU	HG2	(pro-S)	2HG	LEU	HG		HG
GLU	HG3	(pro-R)	3HG	LEU	HD11		1HD1
GLU	HE2		HE2	LEU	HD12		2HD1
GLU	C		C	LEU	HD13		3HD1
GLU	CA		CA	LEU	HD21		1HD2
GLU	CB		CB	LEU	HD22		2HD2
GLU	CG		CG	LEU	HD23		3HD2
GLU	CD		CD	LEU	C		C
GLU	N		N	LEU	CA		CA
GLU	O		O	LEU	CB		CB
GLU	OE1		OE1	LEU	CG		CG
GLU	OE2		OE2	LEU	CD1	(pro-R)	CD1
				LEU	CD2	(pro-S)	CD2
GLN	H		H	LEU	N		N
GLN	HA		HA	LEU	O		O
GLN	HB2	(pro-R)	2HB				
GLN	HB3	(pro-S)	3HB	LYS	H		H
GLN	HG2	(pro-S)	2HG	LYS	HA		HA
GLN	HG3	(pro-R)	3HG	LYS	HB2	(pro-R)	2HB
GLN	HE21	(E)	1HE2	LYS	HB3	(pro-S)	3HB
GLN	HE22	(Z)	2HE2	LYS	HG2	(pro-R)	2HG
GLN	C		C	LYS	HG3	(pro-S)	3HG
GLN	CA		CA	LYS	HD2	(pro-S)	2HD
GLN	CB		CB	LYS	HD3	(pro-R)	3HD
GLN	CG		CG	LYS	HE2	(pro-S)	2HE
GLN	CD		CD	LYS	HE3	(pro-R)	3HE
GLN	N		N	LYS	HZ1		1HZ
GLN	NE2		NE2	LYS	HZ2		2HZ
GLN	O		O	LYS	HZ3		3HZ
GLN	OE1		OE1	LYS	C		C
				LYS	CA		CA
GLY	H		H	LYS	CB		CB
GLY	HA2	(pro-R)	2HA	LYS	CG		CG
GLY	HA3	(pro-S)	3HA	LYS	CD		CD
GLY	C		C	LYS	CE		CE
GLY	CA		CA	LYS	N		N
GLY	N		N	LYS	NZ		NZ
GLY	O		O	LYS	O		O
HIS	H		H	MET	H		H
HIS	HA		HA	MET	HA		HA
HIS	HB2	(pro-S)	2HB	MET	HB2	(pro-S)	2HB
HIS	HB3	(pro-R)	3HB	MET	HB3	(pro-R)	3HB
HIS	HD1		HD1	MET	HG2	(pro-S)	2HG
HIS	HD2		HD2	MET	HG3	(pro-R)	3HG
HIS	HE1		HE1	MET	HE1		1HE
HIS	HE2		HE2	MET	HE2		2HE
HIS	C		C	MET	HE3		3HE
HIS	CA		CA	MET	C		C
HIS	CB		CB	MET	CA		CA
HIS	CG		CG	MET	CB		CB
HIS	CD2		CD2	MET	CG		CG
HIS	CE1		CE1	MET	CE		CE
HIS	N		N	MET	N		N
HIS	ND1		ND1	MET	O		O
HIS	NE2		NE2	MET	SD		SD
HIS	O		O				
ILE	H		H	PHE	H		H
ILE	HA		HA	PHE	HA		HA
ILE	HB		HB	PHE	HB2	(pro-R)	1HB
ILE	HG12	(pro-R)	2HG1	PHE	HB3	(pro-S)	2HB
ILE	HG13	(pro-S)	3HG1	PHE	HD1		HD1
ILE	HG21		1HG2	PHE	HD2		HD2
ILE	HG22		2HG2	PHE	HE1		HE1
ILE	HG23		3HG2	PHE	HE2		HE2
ILE	HD11		1HD1	PHE	HZ		HZ
ILE	HD12		2HD1	PHE	C		C
ILE	HD13		3HD1	PHE	CA		CA
ILE	C		C	PHE	CB		CB
ILE	CA		CA	PHE	CG		CG
ILE	CB		CB	PHE	CD1		CD1
ILE	CG1		CG1	PHE	CD2		CD2
ILE	CG2		CG2	PHE	CE1		CE1
ILE	CD1		CD1	PHE	CE2		CE2
ILE	N		N	PHE	CZ		CZ
ILE	O		O	PHE	N		N
				PHE	O		O
LEU	H		H	PRO	H2	(pro-R)	H2
LEU	HA		HA	PRO	H3	(pro-S)	H3
LEU	HB2	(pro-R)	2HB	PRO	HA		HA

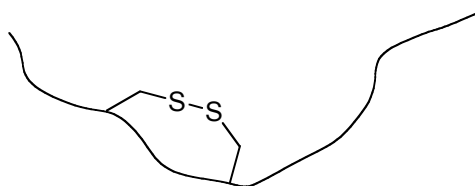
PRO	HB2	(pro-R)	2HB	TYR	CE2	CE2
PRO	HB3	(pro-S)	3HB	TYR	CZ	CZ
PRO	HG2	(pro-S)	2HG	TYR	N	N
PRO	HG3	(pro-R)	3HG	TYR	O	O
PRO	HD2	(pro-S)	2HD	TYR	OH	OH
PRO	HD3	(pro-R)	3HD			
PRO	C		C	VAL	H	H
PRO	CA		CA	VAL	HA	HA
PRO	CB		CB	VAL	HB	HB
PRO	CG		CG	VAL	HG11	1HG1
PRO	CD		CD	VAL	HG12	2HG1
PRO	N		N	VAL	HG13	3HG1
PRO	O		O	VAL	HG21	1HG2
				VAL	HG22	2HG2
SER	H		H	VAL	HG23	3HG2
SER	HA		HA	VAL	C	C
SER	HB2	(pro-S)	2HB	VAL	CA	CA
SER	HB3	(pro-R)	3HB	VAL	CB	CB
SER	HG		HG	VAL	CG1 (pro-R)	CG1
SER	C		C	VAL	CG2 (pro-S)	CG2
SER	CA		CA	VAL	N	N
SER	CB		CB	VAL	O	O
SER	N		N			
SER	O		O			
SER	OG		OG			
THR	H		H			
THR	HA		HA			
THR	HB		HB			
THR	HG1		HG1			
THR	HG21		1HG2			
THR	HG22		2HG2			
THR	HG23		3HG2			
THR	C		C			
THR	CA		CA			
THR	CB		CB			
THR	CG2		CG2			
THR	N		N			
THR	O		O			
THR	OG1		OG1			
TRP	H		H			
TRP	HA		HA			
TRP	HB2	(pro-R)	2HB			
TRP	HB3	(pro-S)	3HB			
TRP	HD1		HD1			
TRP	HE1		HE1			
TRP	HE3		HE3			
TRP	HZ2		HZ2			
TRP	HZ3		HZ3			
TRP	HH2		HH2			
TRP	C		C			
TRP	CA		CA			
TRP	CB		CB			
TRP	CG		CG			
TRP	CD1		CD1			
TRP	CD2		CD2			
TRP	CE2		CE2			
TRP	CE3		CE3			
TRP	CZ2		CZ2			
TRP	CZ3		CZ3			
TRP	CH2		CH2			
TRP	N		N			
TRP	NE1		NE1			
TRP	O		O			
TYR	H		H			
TYR	HA		HA			
TYR	HB2	(pro-R)	2HB			
TYR	HB3	(pro-S)	3HB			
TYR	HD1		HD1			
TYR	HD2		HD2			
TYR	HE1		HE1			
TYR	HE2		HE2			
TYR	HH		HH			
TYR	C		C			
TYR	CA		CA			
TYR	CB		CB			
TYR	CG		CG			
TYR	CD1		CD1			
TYR	CD2		CD2			
TYR	CE1		CE1			

3.2 Peptides

Les acides aminés s'unissent en formant des liaisons peptidiques pour former des peptides.



Un autre type de liaison peut être présent entre deux cystéines :



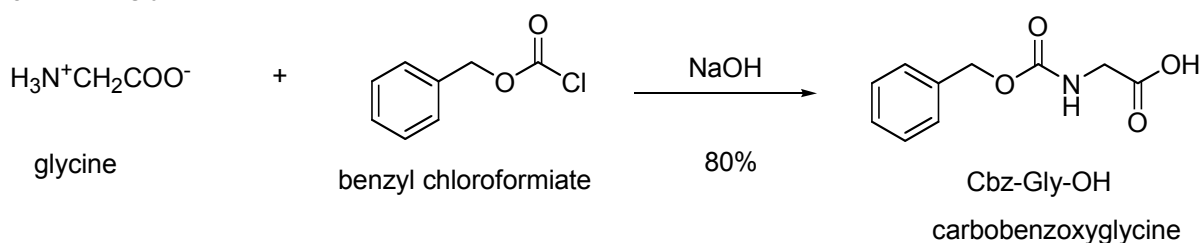
3.2.1 Synthèse des polypeptides

On peut envisager de chauffer 2 acides aminés mais on va obtenir un mélange ...



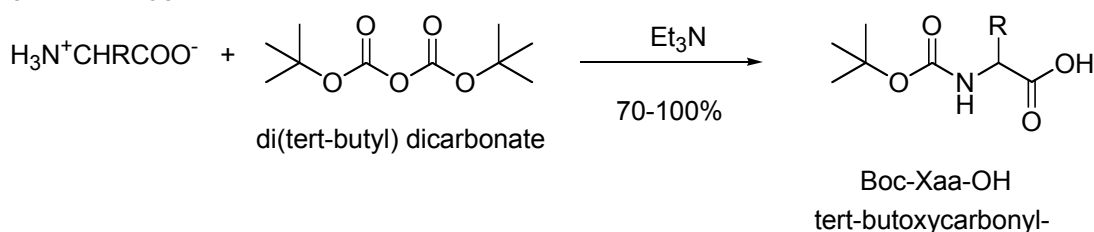
3.2.2 Nécessité d'avoir des groupes protecteurs :

3.2.2.1 Cbz



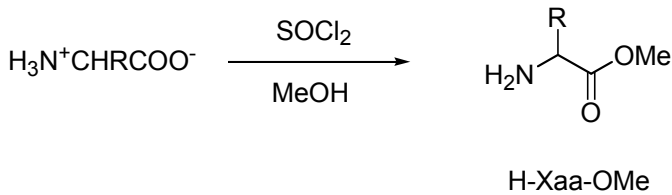
Déprotection : H_2 / Pd-C

3.2.2.2 Boc



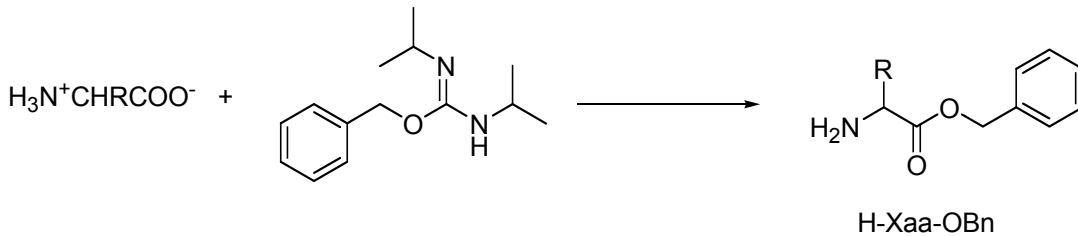
Déprotection : TFA, 25°C

3.2.2.3 OMe



Déprotection : LiOH 1N

3.2.2.4 OBn (benzyl ester)



Déprotection : H₂ / Pd-C

3.2.3 Activation du groupement carbonyle

Activation par DCC (N,N'-dicyclohexylcarbodiimide)

ROOH + RNH₂ + DCC → amide + urée

3.2.4 Synthèse sélective

Boc-Gly-OH + H-Ala-OBn → Boc-Gly-Ala-OBn

Déprotection H⁺/H₂O et H₂ / Pd-C

3.2.5 Synthèse peptidique sur support solide : Merrifield

Merrifield, prix Nobel 1984.

Polymérisation radicalaire du styrène + divinylbenzène : polystyrène

Fonctionnalisation du polystyrène par chlorométhylation électrophile

ClCH₂OEt / SnCl₄

1. fixation de l'acide aminé N-protégé
2. déprotection de l'extrémité amino
3. couplage avec le second aa N-protégé
4. déprotection de l'extrémité amino
5. déconnection du dipeptide par rapport au polymère

3.2.5.1 Avantages :

- rapide
- pas de purification intermédiaire
- large excès de réactifs (réaction plus complète)
- double couplage
- automatisable

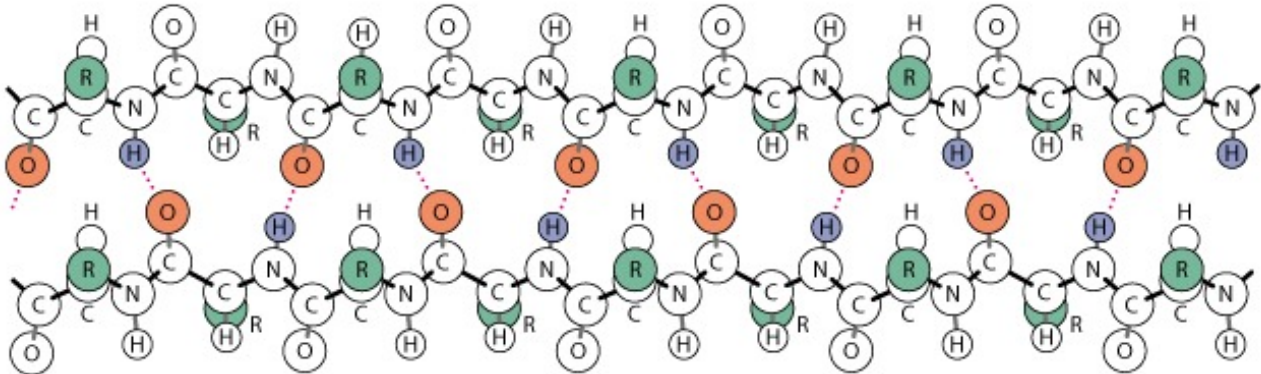
3.2.5.2 Inconvénients

- pas toujours très pur, on peut obtenir un mélange très complexe

3.3 Protéines

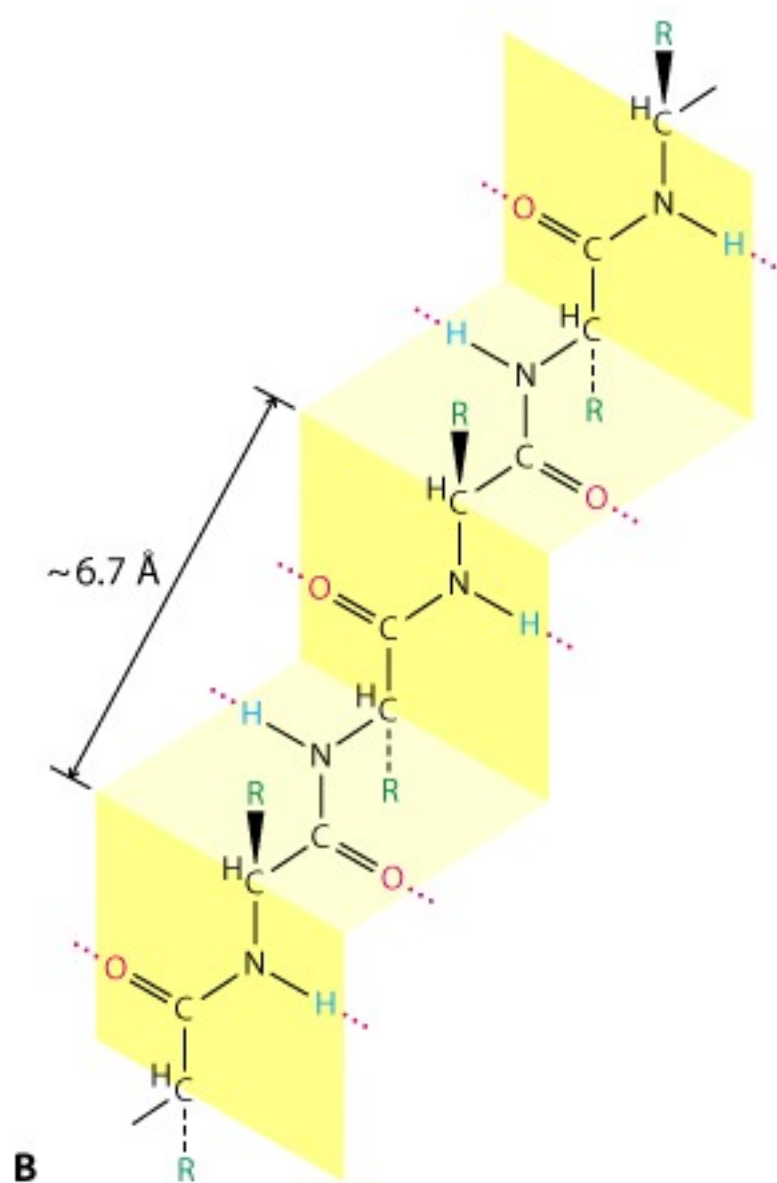
Séquence primaire : séquences des acides aminés.

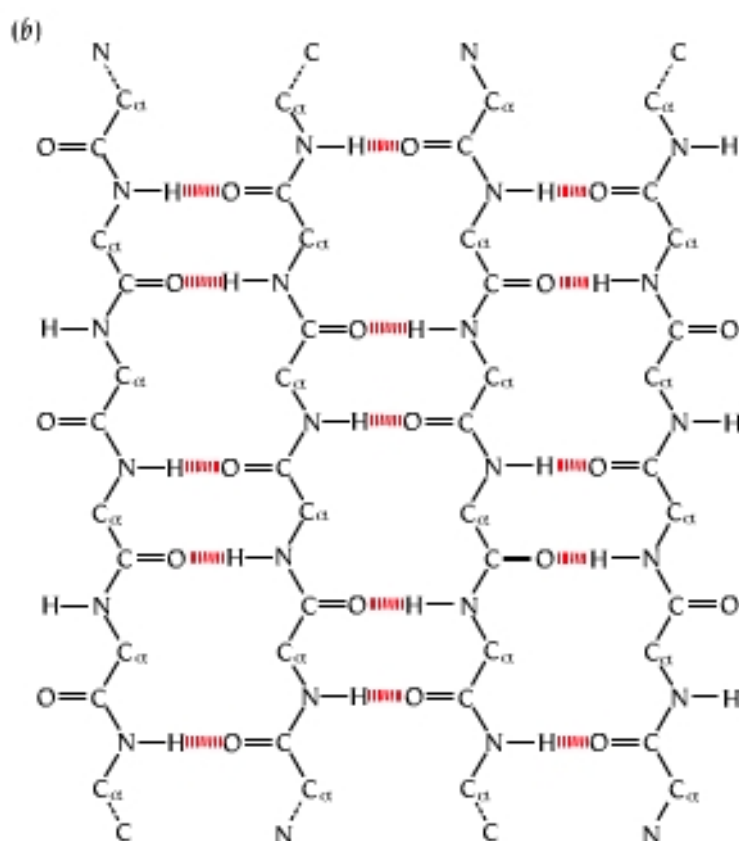
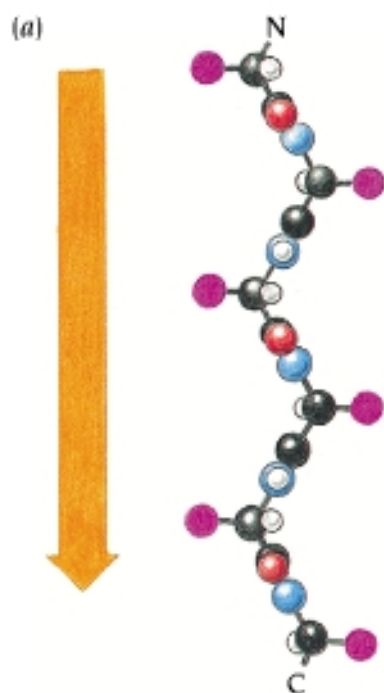
Les protéines se replient sous forme de feuilles plissées ou d'hélices : structure secondaire.



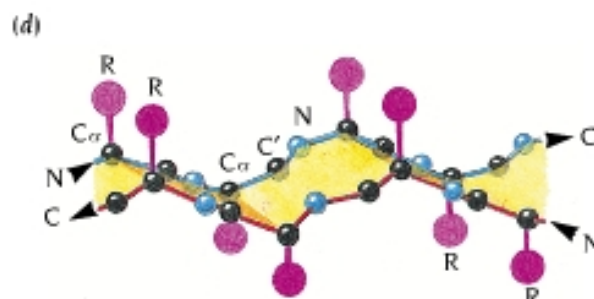
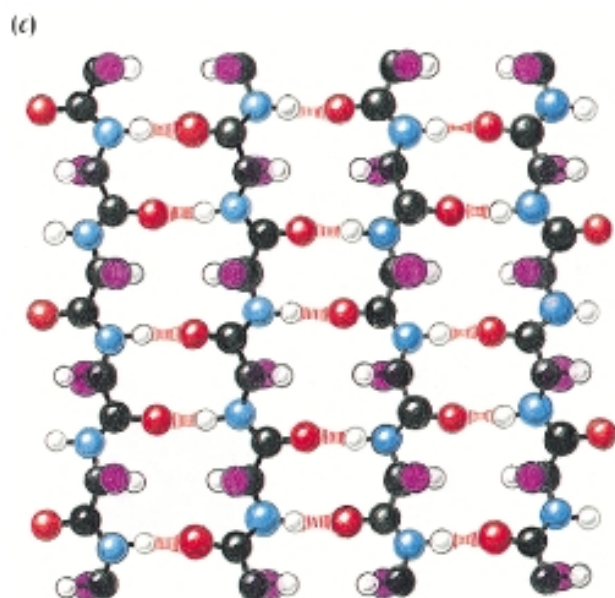
En moyenne 1 pont hydrogène par acide aminé.

Substituants sur le carbone α en alternance au-dessus, en dessous.

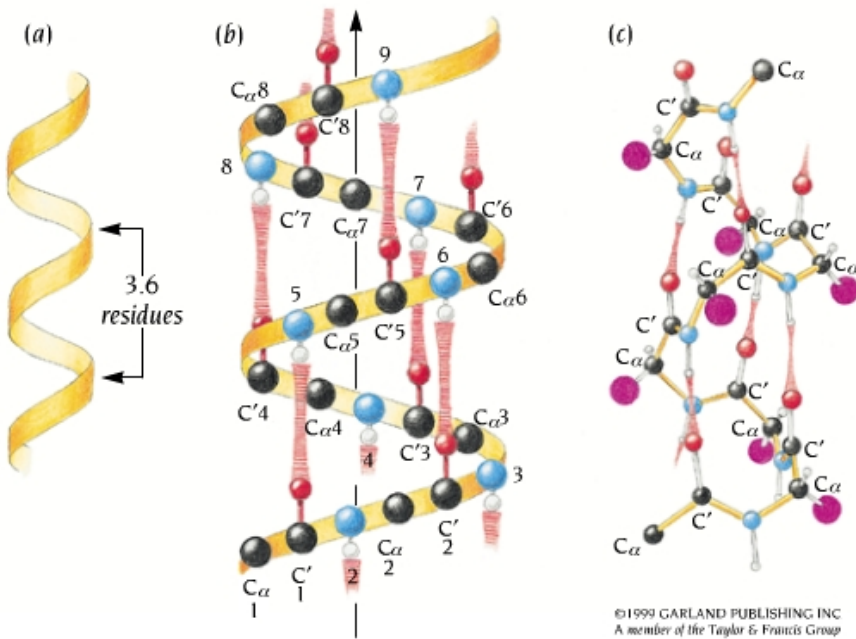
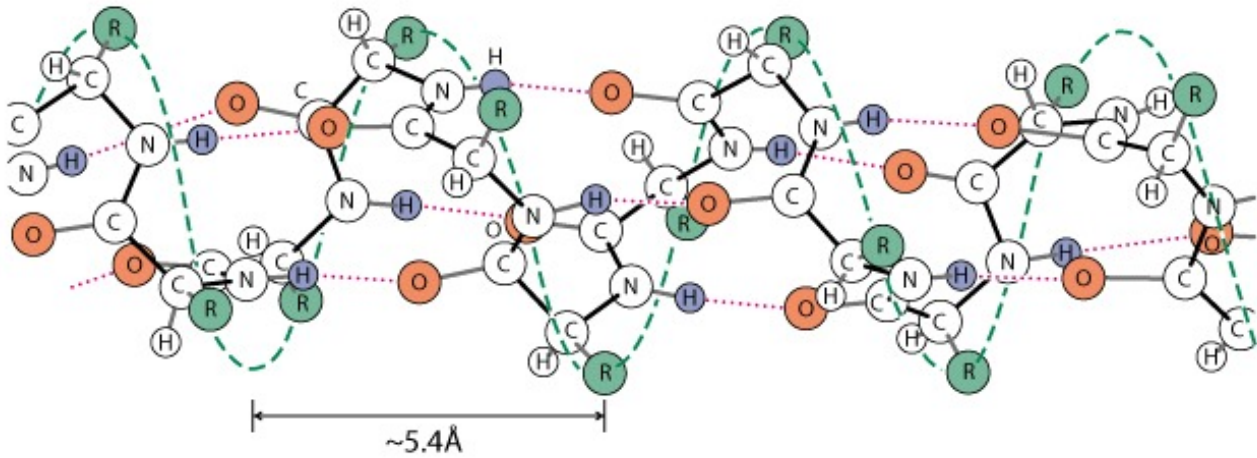




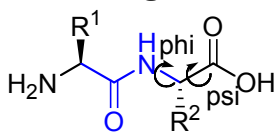
© 1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group



© 1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group



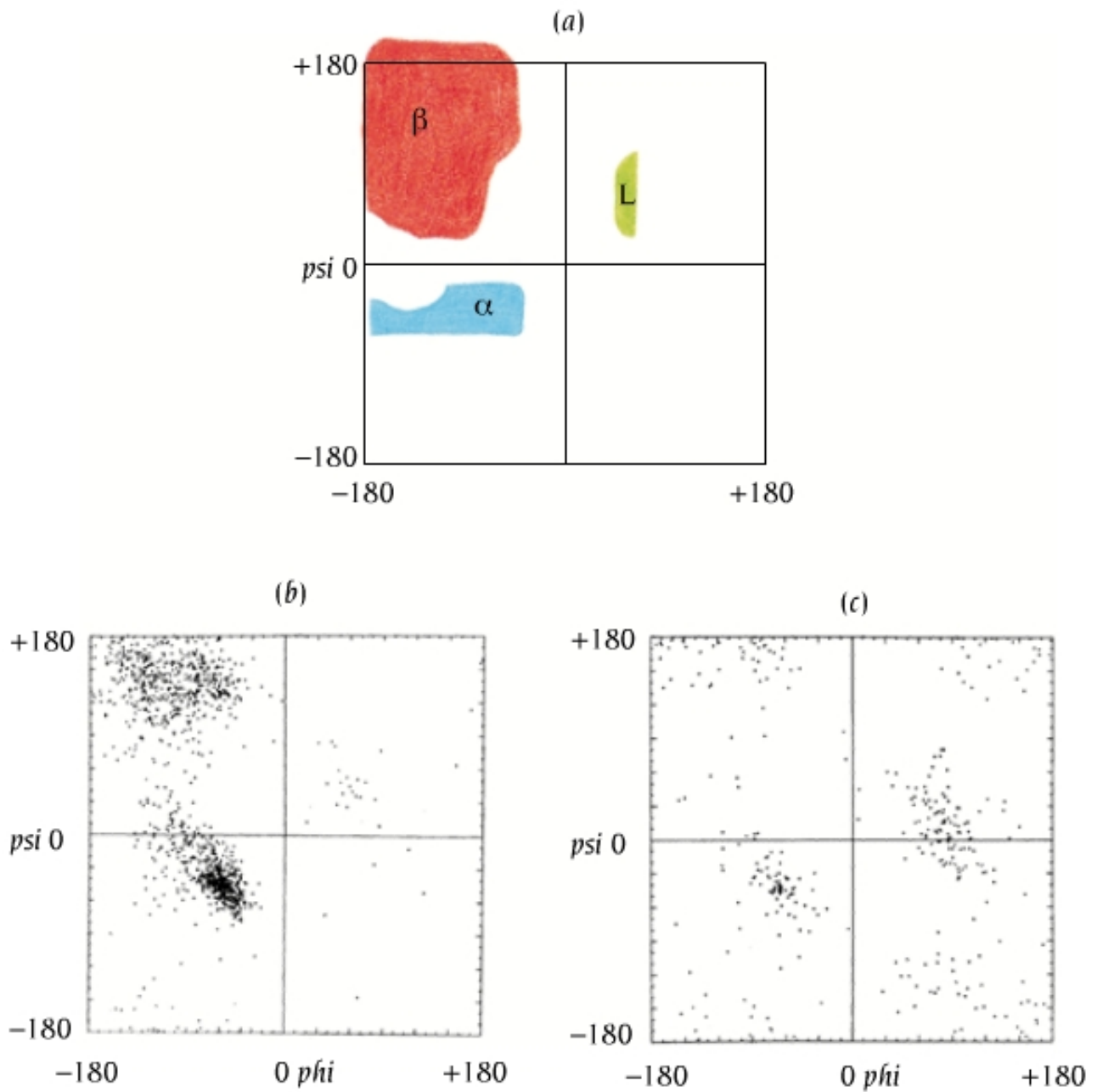
3.4 Diagramme de Ramachandran



α : hélice alpha droite

L : hélice alpha gauche

β : plan beta (feuilles plissées)



©1999 GARLAND PUBLISHING INC.
A member of the Taylor & Francis Group

Structure tertiaire provient d'autres interactions provenant de chaînes latérales (ponts disulfures, interactions hydrophobes, ...). Va donner la structure globulaire des protéines.

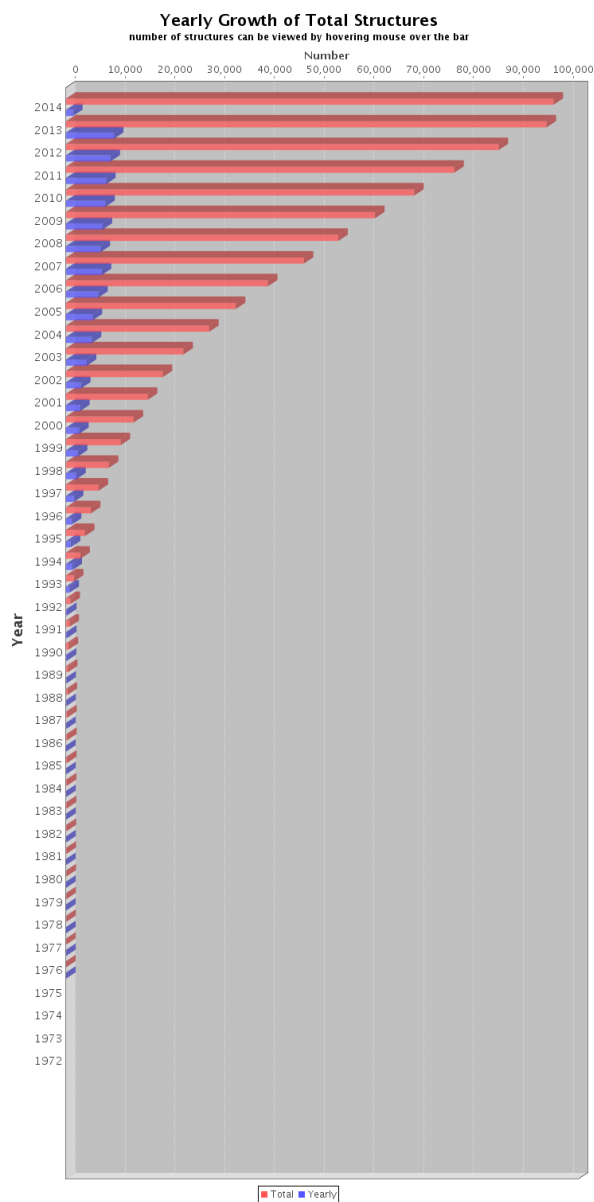
4 Protein Data Bank (PDB)

Structures déterminées par RX ou RMN.

Gratuitement accessible : <http://www.pdb.org>

4.1 Statistiques: 24-Feb-2014

		Proteins
Exp.	X-ray Diffraction and other	80964
	NMR	9038
Tech.	Total	20939



4.2 Format PDB

- Format de fichier permettant de stocker les informations sur les coordonnées et la structure de la molécule.
- Document texte
- Suite de lignes de 80 caractères.
- Les 6 premières colonnes identifient la ligne.
- Description complète: <http://www.wwpdb.org/docs.html>

4.2.1 Début et fin

HEADER Première ligne de l'entrée, contient le ID PDB

END Dernière ligne

4.2.2 Littérature / Expérience

COMPND Description des molécules

SOURCE Provenance de l'échantillon

AUTHOR Liste des auteurs

TITLE Titre de l'expérience

JNRL Référence dans la littérature

KEYWDS Liste de mots clef

REMARK Remarques

4.2.3 Composition

SEQRES Séquence primaire des résidus

HET Groupement ou résidus non-standards

FORMUL Formule brute des résidus exotiques

4.2.4 Structure secondaire, site important

SHEET Identification de feuilles plissées (plan β)

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SHEET "	
8 - 10	Integer	strand	Strand number which starts at 1 for each strand within a sheet and increases by one.
12 - 14	LString(3)	sheetID	Sheet identifier.
15 - 16	Integer	numStrands	Number of strands in sheet.
18 - 20	Residue name	initResName	Residue name of initial residue.
22	Character	initChainID	Chain identifier of initial residue in strand.
23 - 26	Integer	initSeqNum	Sequence number of initial residue in strand.
27	AChar	initICode	Insertion code of initial residue in strand.
29 - 31	Residue name	endResName	Residue name of terminal residue.
33	Character	endChainID	Chain identifier of terminal residue.
34 - 37	Integer	endSeqNum	Sequence number of terminal residue.
38	AChar	endICode	Insertion code of terminal residue.
39 - 40	Integer	sense	Sense of strand with respect to previous strand in the sheet. 0 if first strand, 1 if parallel,

42 - 45	Atom	curAtom	-1 if anti-parallel. Registration. Atom name in current strand.
46 - 48	Residue name	curResName	Registration. Residue name in current strand.
50	Character	curChainId	Registration. Chain identifier in current strand.
51 - 54	Integer	curResSeq	Registration. Residue sequence number in current strand.
55	AChar	curICode	Registration. Insertion code in current strand.
57 - 60	Atom	prevAtom	Registration. Atom name in previous strand.
61 - 63	Residue name	prevResName	Registration. Residue name in previous strand.
65	Character	prevChainId	Registration. Chain identifier in previous strand.
66 - 69	Integer	prevResSeq	Registration. Residue sequence number in previous strand.
70	AChar	prevICode	Registration. Insertion code in previous strand.

HELIX Identification d'hélice α

TURN Tournant

SITE Par exemple site catalytique

4.2.5 Atomes

ATOM Coordonnées atomiques

HETATM Coordonnées atomiques des résidus exotiques, même structure que ATM

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM "	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms.
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms.
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms.
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
73 - 76	LString(4)	segID	Segment identifier, left-justified.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

TER terminaison de chaîne

4.2.6 Liaison

SSBOND Identification des ponts disulfures

CONECT Connection entre atomes. Nécessaire pour les groupes exotiques (HET)

4.3 Exemple de PDB

```

HEADER    COMPLEX (ISOMERASE/IMMUNOSUPPRESSANT)    06-SEP-95    1CWA    1CWA    2
COMPND    MOL_ID: 1;                                1CWA    3
COMPND    2 MOLECULE: CYCLOPHILIN A;                1CWA    4
COMPND    3 CHAIN: A;                                1CWA    5

```

COMPND 4 ENGINEERED: YES; 1CWA 6
 COMPND 5 MOL_ID: 2; 1CWA 7
 COMPND 6 MOLECULE: CYCLOSPORIN A; 1CWA 8
 COMPND 7 CHAIN: C; 1CWA 9
 COMPND 8 ENGINEERED: YES 1CWA 10
 SOURCE MOL_ID: 1; 1CWA 11
 SOURCE 2 ORGANISM_SCIENTIFIC: HOMO SAPIENS; 1CWA 12
 SOURCE 3 ORGANISM_COMMON: HUMAN; 1CWA 13
 SOURCE 4 EXPRESSION_SYSTEM: ESCHERICHIA COLI; 1CWA 14
 SOURCE 5 MOL_ID: 2; 1CWA 15
 SOURCE 6 SYNTHETIC: YES 1CWA 16
 EXPDTA X-RAY DIFFRACTION 1CWA 17
 AUTHOR V.MIKOL,J.KALLEN,M.D.WALKINSHAW 1CWA 18
 REVDTA 1 29-JAN-96 1CWA 0 1CWA 19
 JRNL AUTH V.MIKOL,J.KALLEN,G.PFLUEGL,M.D.WALKINSHAW 1CWA 20
 JRNL TITL X-RAY STRUCTURE OF A MONOMERIC CYCLOPHILIN 1CWA 21
 JRNL TITL 2 A-CYCLOSPORIN A CRYSTAL COMPLEX AT 2.1 ANGSTROMS 1CWA 22
 JRNL TITL 3 RESOLUTION 1CWA 23
 JRNL REF J.MOL.BIOL. V. 234 1119 1993 1CWA 24
 JRNL REFN ASTM JMOBAC UK ISSN 0022-2836 0070 1CWA 25
 REMARK 1 1CWA 26
 REMARK 1 REFERENCE 1 1CWA 27
 REMARK 1 AUTH V.MIKOL,D.DUC 1CWA 28
 REMARK 1 TITL CRYSTALLIZATION OF THE COMPLEX BETWEEN CYCLOPHILIN 1CWA 29
 REMARK 1 TITL 2 A AND CYCLOSPORIN DERIVATIVES: THE USE OF 1CWA 30
 REMARK 1 TITL 3 CROSS-SEEDING 1CWA 31
 REMARK 1 REF ACTA CRYSTALLOGR.,SECT.D V. 50 543 1994 1CWA 32
 REMARK 1 REFN ASTM ABCRE6 DK ISSN 0907-4449 0766 1CWA 33
 REMARK 1 REFERENCE 2 1CWA 34
 REMARK 1 AUTH V.MIKOL,J.KALLEN,M.D.WALKINSHAW 1CWA 35
 REMARK 1 TITL THE X-RAY STRUCTURE OF (MEBM=2=T) ==1==--CYCLOSPORIN 1CWA 36
 REMARK 1 TITL 2 COMPLEXED WITH CYCLOPHILIN A PROVIDES AN 1CWA 37
 REMARK 1 TITL 3 EXPLANATION FOR ITS ANOMALOUSLY HIGH 1CWA 38
 REMARK 1 TITL 4 IMMUNOSUPPRESSIVE ACTIVITY 1CWA 39
 REMARK 1 REF PROTEIN ENG. V. 7 597 1994 1CWA 40
 REMARK 1 REFN ASTM PRENE9 UK ISSN 0269-2139 0859 1CWA 41
 REMARK 1 REFERENCE 3 1CWA 42
 REMARK 1 AUTH C.PAPAGEORGIOU,A.FLORINETH,V.MIKOL 1CWA 43
 REMARK 1 TITL IMPROVED BINDING AFFINITY FOR CYCLOPHILIN A BY A 1CWA 44
 REMARK 1 TITL 2 CYCLOSPORIN DERIVATIVE SINGLY MODIFIED AT ITS 1CWA 45
 REMARK 1 TITL 3 EFFECTOR DOMAIN 1CWA 46
 REMARK 1 REF J.MED.CHEM. V. 37 3674 1994 1CWA 47
 REMARK 1 REFN ASTM JMCAR US ISSN 0022-2623 0151 1CWA 48
 REMARK 1 REFERENCE 4 1CWA 49
 REMARK 1 AUTH G.PFLUEGL,J.KALLEN,T.SCHIRMER,J.N.JANSONIUS, 1CWA 50
 DBREF 1CWA A 2 165 SWS P05092 CYPH_HUMAN 1 164 1CWA 116
 SEQADV 1CWA MET A 1 SWS P05092 MISSING FROM SWS 1CWA 117
 SEQRES 1 A 165 MET VAL ASN PRO THR VAL PHE PHE ASP ILE ALA VAL ASP 1CWA 118
 SEQRES 2 A 165 GLY GLU PRO LEU GLY ARG VAL SER PHE GLU LEU PHE ALA 1CWA 119
 SEQRES 3 A 165 ASP LYS VAL PRO LYS THR ALA GLU ASN PHE ARG ALA LEU 1CWA 120
 SEQRES 4 A 165 SER THR GLY GLU LYS GLY PHE GLY TYR LYS GLY SER CYS 1CWA 121
 SEQRES 5 A 165 PHE HIS ARG ILE ILE PRO GLY PHE MET CYS GLN GLY GLY 1CWA 122
 SEQRES 6 A 165 ASP PHE THR ARG HIS ASN GLY THR GLY GLY LYS SER ILE 1CWA 123
 SEQRES 7 A 165 TYR GLY GLU LYS PHE GLU ASP GLU ASN PHE ILE LEU LYS 1CWA 124
 SEQRES 8 A 165 HIS THR GLY PRO GLY ILE LEU SER MET ALA ASN ALA GLY 1CWA 125
 SEQRES 9 A 165 PRO ASN THR ASN GLY SER GLN PHE PHE ILE CYS THR ALA 1CWA 126
 SEQRES 10 A 165 LYS THR GLU TRP LEU ASP GLY LYS HIS VAL VAL PHE GLY 1CWA 127
 SEQRES 11 A 165 LYS VAL LYS GLU GLY MET ASN ILE VAL GLU ALA MET GLU 1CWA 128
 SEQRES 12 A 165 ARG PHE GLY SER ARG ASN GLY LYS THR SER LYS LYS ILE 1CWA 129
 SEQRES 13 A 165 THR ILE ALA ASP CYS GLY GLN LEU GLU 1CWA 130
 SEQRES 1 C 11 BMT ABA SAR MLE VAL MLE ALA DAL MLE MLE MVA 1CWA 131
 FTNOTE 1 1CWA 132
 FTNOTE 1 VAL C 5 - LEU C 6 OMEGA = 0.00 1CWA 133
 FTNOTE 1 PEPTIDE BOND DEVIATES SIGNIFICANTLY FROM TRANS CONFORMATION 1CWA 134
 HET BMT C 1 13 4-[(E)-2-BUTENYL]-4,N-DIMETHYL-THREONINE 1CWA 135
 HET ABA C 2 6 ALPHA-AMINOBUTYRIC ACID 1CWA 136
 HET SAR C 3 5 SARCOSINE 1CWA 137
 HET MLE C 4 9 N-METHYLLEUCINE 1CWA 138
 HET MLE C 6 9 N-METHYLLEUCINE 1CWA 139
 HET DAL C 8 5 D-ALANINE 1CWA 140
 HET MLE C 9 9 N-METHYLLEUCINE 1CWA 141

HET	MLE	C	10	9	N-METHYLLEUCINE				1CWA	142							
HET	MVA	C	11	8	N-METHYLVALINE				1CWA	143							
FORMUL	2	BMT	C10 H19 N1 O3							1CWA	144						
FORMUL	2	ABA	C4 H9 N1 O2							1CWA	145						
FORMUL	2	SAR	C3 H7 N1 O2							1CWA	146						
FORMUL	2	MLE	4(C7 H15 N1 O2)							1CWA	147						
FORMUL	2	DAL	C3 H7 N1 O2							1CWA	148						
FORMUL	2	MVA	C6 H13 N1 O2							1CWA	149						
FORMUL	3	HOH	*144(H2 O1)							1CWA	150						
HELIX	1	A	PRO	A	30	THR	A	41	1	1CWA	151						
HELIX	2	B	GLU	A	120	LEU	A	122	5	1CWA	152						
HELIX	3	C	MET	A	136	GLU	A	143	1	1CWA	153						
SHEET	1	S1	8	ILE	A	156	GLN	A	163	0	1CWA	154					
SHEET	2	S1	8	PHE	A	53	ILE	A	57	-1	1CWA	155					
SHEET	3	S1	8	MET	A	61	GLY	A	72	-1	1CWA	156					
SHEET	4	S1	8	PHE	A	112	CYS	A	115	-1	1CWA	157					
SHEET	5	S1	8	ILE	A	97	MET	A	100	-1	1CWA	158					
SHEET	6	S1	8	VAL	A	128	GLU	A	134	-1	1CWA	159					
SHEET	7	S1	8	GLU	A	15	LEU	A	24	-1	1CWA	160					
SHEET	8	S1	8	THR	A	5	VAL	A	12	-1	1CWA	161					
TURN	1	T1	VAL	A	12	GLU	A	15	TYPE I'		1CWA	162					
TURN	2	T2	ALA	A	26	VAL	A	29	TYPE I		1CWA	163					
TURN	3	T3	ILE	A	57	PHE	A	60	TYPE II		1CWA	164					
SITE	1	BIN	13	ARG	A	55	PHE	A	60	MET	A	61	GLN	A	63	1CWA	165
SITE	2	BIN	13	GLY	A	72	ALA	A	101	ASN	A	102	ALA	A	103	1CWA	166
SITE	3	BIN	13	GLN	A	111	PHE	A	113	TRP	A	121	LEU	A	122	1CWA	167
SITE	4	BIN	13	HIS	A	126										1CWA	168
CRYST1	36.395	60.718	72.210	90.00	90.00	90.00	P	21	21	21	4	1CWA	169				
ORIGX1	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000						1CWA	170				
ORIGX2	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000						1CWA	171				
ORIGX3	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000						1CWA	172				
SCALE1	0.027476	0.000000	0.000000	0.000000	0.000000	0.000000						1CWA	173				
SCALE2	0.000000	0.016470	0.000000	0.000000	0.000000	0.000000						1CWA	174				
SCALE3	0.000000	0.000000	0.013848	0.000000	0.000000	0.000000						1CWA	175				
ATOM	1	N	MET	A	1	12.696	8.077	38.463	1.00	51.22	1CWA	176					
ATOM	2	CA	MET	A	1	13.292	9.390	38.316	1.00	50.25	1CWA	177					
ATOM	3	C	MET	A	1	12.202	10.251	38.911	1.00	46.70	1CWA	178					
ATOM	4	O	MET	A	1	11.135	9.735	39.266	1.00	47.07	1CWA	179					
ATOM	5	CB	MET	A	1	13.548	9.787	36.814	1.00	54.09	1CWA	180					
ATOM	6	CG	MET	A	1	12.453	10.214	35.788	1.00	57.61	1CWA	181					
ATOM	7	SD	MET	A	1	12.287	11.992	35.397	1.00	61.46	1CWA	182					
ATOM	8	CE	MET	A	1	13.712	12.258	34.373	1.00	61.18	1CWA	183					
ATOM	9	N	VAL	A	2	12.470	11.531	39.035	1.00	41.83	1CWA	184					
ATOM	10	CA	VAL	A	2	11.496	12.430	39.581	1.00	36.64	1CWA	185					
ATOM	11	C	VAL	A	2	11.667	13.653	38.690	1.00	31.31	1CWA	186					
ATOM	12	O	VAL	A	2	12.781	14.007	38.279	1.00	30.86	1CWA	187					
ATOM	13	CB	VAL	A	2	11.865	12.581	41.103	1.00	38.45	1CWA	188					
ATOM	14	CG1	VAL	A	2	13.306	13.075	41.325	1.00	41.32	1CWA	189					
ATOM	15	CG2	VAL	A	2	10.851	13.512	41.722	1.00	38.59	1CWA	190					
ATOM	16	N	ASN	A	3	10.546	14.212	38.253	1.00	24.80	1CWA	191					
ATOM	17	CA	ASN	A	3	10.582	15.401	37.419	1.00	19.94	1CWA	192					
ATOM	18	C	ASN	A	3	11.120	16.578	38.191	1.00	17.67	1CWA	193					
ATOM	19	O	ASN	A	3	10.769	16.731	39.362	1.00	18.08	1CWA	194					
ATOM	20	CB	ASN	A	3	9.206	15.772	36.910	1.00	16.86	1CWA	195					
ATOM	21	CG	ASN	A	3	8.817	14.832	35.800	1.00	15.98	1CWA	196					
ATOM	22	OD1	ASN	A	3	7.653	14.522	35.610	1.00	17.82	1CWA	197					
ATOM	23	ND2	ASN	A	3	9.708	14.269	35.006	1.00	14.16	1CWA	198					
ATOM	24	N	PRO	A	4	12.014	17.383	37.624	1.00	13.15	1CWA	199					
ATOM	25	CA	PRO	A	4	12.558	18.514	38.341	1.00	11.88	1CWA	200					
ATOM	26	C	PRO	A	4	11.520	19.565	38.701	1.00	11.27	1CWA	201					
ATOM	27	O	PRO	A	4	10.519	19.707	37.998	1.00	11.21	1CWA	202					
ATOM	28	CB	PRO	A	4	13.671	18.988	37.425	1.00	11.37	1CWA	203					
ATOM	29	CG	PRO	A	4	13.352	18.418	36.059	1.00	10.90	1CWA	204					
ATOM	30	CD	PRO	A	4	12.747	17.086	36.401	1.00	11.37	1CWA	205					
ATOM	31	N	THR	A	5	11.733	20.289	39.793	1.00	9.85	1CWA	206					
ATOM	32	CA	THR	A	5	10.896	21.377	40.219	1.00	9.19	1CWA	207					
ATOM	33	C	THR	A	5	11.817	22.570	40.264	1.00	9.95	1CWA	208					
ATOM	34	O	THR	A	5	12.890	22.454	40.878	1.00	10.72	1CWA	209					
ATOM	35	CB	THR	A	5	10.344	21.101	41.611	1.00	10.80	1CWA	210					
ATOM	36	OG1	THR	A	5	9.602	19.890	41.550	1.00	14.06	1CWA	211					
ATOM	37	CG2	THR	A	5	9.480	22.248	42.087	1.00	10.15	1CWA	212					


```

ATOM      38  N   VAL A   6      11.411  23.706  39.694  1.00  6.78      1CWA 213
ATOM      39  CA  VAL A   6      12.219  24.913  39.543  1.00  7.42      1CWA 214
ATOM      40  C   VAL A   6      11.424  26.059  40.127  1.00  6.97      1CWA 215
ATOM      41  O   VAL A   6      10.202  25.938  40.199  1.00  9.40      1CWA 216
ATOM      42  CB  VAL A   6      12.487  25.110  38.012  1.00 12.74      1CWA 217
ATOM      43  CG1 VAL A   6      12.754  26.551  37.633  1.00 13.72      1CWA 218
ATOM      44  CG2 VAL A   6      13.685  24.244  37.649  1.00 12.98      1CWA 219
...
ATOM    1260  O   GLU A 165      14.634  17.408  41.991  1.00 24.79      1CWA1435
ATOM    1261  CB  GLU A 165      12.689  18.744  44.103  1.00 28.30      1CWA1436
ATOM    1262  CG  GLU A 165      12.330  17.273  44.213  1.00 38.66      1CWA1437
ATOM    1263  CD  GLU A 165      11.128  17.058  45.129  1.00 44.34      1CWA1438
ATOM    1264  OE1 GLU A 165      10.043  17.556  44.797  1.00 46.41      1CWA1439
ATOM    1265  OE2 GLU A 165      11.289  16.403  46.172  1.00 46.39      1CWA1440
ATOM    1266  OXT GLU A 165      14.097  19.330  41.174  1.00 18.97      1CWA1441
TER      1267  GLU A 165
HETATM  1268  N   BMT C   1         5.177  34.600  19.702  1.00  5.99      1CWA1443
HETATM  1269  CN  BMT C   1         4.611  35.988  19.802  1.00  3.02      1CWA1444
HETATM  1270  CA  BMT C   1         6.388  34.385  18.911  1.00  9.99      1CWA1445
HETATM  1271  C   BMT C   1         7.585  35.165  19.430  1.00 10.72      1CWA1446
HETATM  1272  O   BMT C   1         7.509  35.989  20.346  1.00 12.90      1CWA1447
HETATM  1273  CB  BMT C   1         6.215  34.722  17.334  1.00  7.86      1CWA1448
HETATM  1274  OG1 BMT C   1         6.015  36.130  17.180  1.00  8.51      1CWA1449
HETATM  1275  CG2 BMT C   1         4.985  34.028  16.675  1.00  7.29      1CWA1450
HETATM  1276  CD1 BMT C   1         5.061  32.527  16.900  1.00  3.90      1CWA1451
...
HETATM  1343  C   MLE C  10         1.442  34.383  20.285  1.00 10.43      1CWA1518
HETATM  1344  O   MLE C  10         1.493  35.438  20.958  1.00  9.52      1CWA1519
HETATM  1345  N   MVA C  11         2.366  33.397  20.388  1.00  7.84      1CWA1520
HETATM  1346  CN  MVA C  11         2.294  32.013  19.835  1.00  7.46      1CWA1521
HETATM  1347  CA  MVA C  11         3.500  33.683  21.244  1.00  8.92      1CWA1522
HETATM  1348  CB  MVA C  11         3.478  32.735  22.488  1.00 10.37      1CWA1523
HETATM  1349  CG1 MVA C  11         4.465  33.299  23.491  1.00  9.22      1CWA1524
HETATM  1350  CG2 MVA C  11         2.113  32.701  23.180  1.00 10.87      1CWA1525
HETATM  1351  C   MVA C  11         4.794  33.534  20.442  1.00  7.42      1CWA1526
HETATM  1352  O   MVA C  11         5.399  32.431  20.436  1.00  4.10      1CWA1527
TER      1353  MVA C  11
HETATM  1354  O   HOH      1        13.047  24.867  27.324  1.00 12.88      1CWA1529
HETATM  1355  O   HOH      2        22.613  27.940  32.427  1.00  9.41      1CWA1530
HETATM  1356  O   HOH      3         7.910  25.142  20.393  1.00  4.04      1CWA1531
HETATM  1357  O   HOH      4        12.453  36.127  31.572  1.00 20.00      1CWA1532
HETATM  1358  O   HOH      5        12.146  29.906  22.594  1.00 13.47      1CWA1533
HETATM  1359  O   HOH      6        12.718  31.311  19.707  1.00 20.71      1CWA1534
...
HETATM  1493  O   HOH    140        24.779  34.296  26.774  1.00 37.99      1CWA1668
HETATM  1494  O   HOH    141         1.691  30.122  11.355  1.00 53.47      1CWA1669
HETATM  1495  O   HOH    142        -5.960  32.347  35.204  1.00 47.65      1CWA1670
HETATM  1496  O   HOH    143        -9.158  34.235  29.310  1.00 21.15      1CWA1671
HETATM  1497  O   HOH    144         0.185  43.301  43.162  1.00 50.41      1CWA1672
CONECT  1268 1269 1270 1351
CONECT  1269 1268
CONECT  1270 1268 1271 1273
CONECT  1271 1270 1272 1281
CONECT  1272 1271
...
CONECT  1351 1268 1347 1352
CONECT  1352 1351
MASTER      90      3      9      3      8      3      4      6 1495      2      77      14
END
1CWA1751

```

4.4 Recherche dans la PDB

<http://www.pdb.org/>

Exemple :

1LRV

Vérifier les informations globales :

une chaîne, 167 aa, 1510 atomes

BB2 : actinonine

ZN : zinc

Download : complete with coordinates, format PDB

4.5 JSmol (<http://www.jmol.org>)

Jmol est une applet (programme écrit en "java") qui permet de visualiser des molécules au départ d'une page WEB. Jmol est très performant et permet une très grande interactivité.

La nouvelle version, JsMol ne requiert plus java et fonctionne avec n'importe quel navigateur moderne.

4.5.1 Instructions utiles

4.5.1.1 Selection

Voir [atom expressions]

Sélection:

- [residueType]
- seqRange
- :chainLetter
- .atomName
- _atomType

Wildcards:

* n'importe quel "residueType"

? remplace 0 ou 1 caractère

Exemples:

[ALA].CA	Tous les carbones alpha des alanines
*.?	Tous les "atomName" en une lettre de n'importe quel "residueType"
select within(4, :C) and not :C and not HOH	Tous les atomes à une distance de moins de 4Å d'un des atomes de la "chainLetter" C et qui n'est pas d'un "residueType" HOH ni de la "chainLetter" C.

```
select acid;
select atomno=123;
select within (4.0, (:c)) and not :c;
select within( GROUP, within (3.0, (:c)) and not :c)
select [GLY]
select [GLY]:C
select 1-10:A
select 1-10:A.CA
```

4.5.1.2 Show beta-sheet

```
select ALL; wireframe off; cpk 0%;
select (89-97 or 109-117) and :a and (*.CA or *.C or *.N or *.O);
cpk 20%, wireframe on;
ribbon on;
hbonds calculate; hbonds 0.5; color hbonds yellow;
```

4.5.1.3 Animation

```
zoomto 2 (97:a) 400%;
```

```
delay 2;  
zoomto 2 (all) 100%;
```

4.5.1.4 Ramachandran

```
select :a; rama;  
select [LYS]  
model 1  
model 2
```

4.5.1.5 Transparency

```
color ribbons TRANSLUCENT 0.5 orange
```

4.5.1.6 Radius temperature

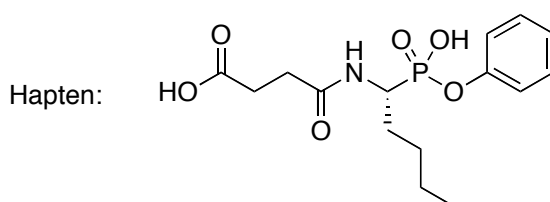
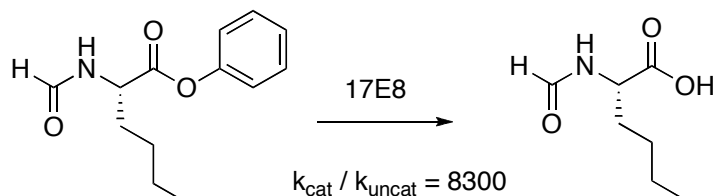
4.5.1.7 Measure

```
set measurements 0.5; set measurements DOTTED; measure  
(atomno=123) (atomno=246); color measure red;
```

5 Réactions "biocompatibles"

Il est souvent intéressant soit de modifier des protéines, pour par exemple fixer un haptène (catalytic antibodies (Nat. Prod. Rep., 1996, 13, 479 - 511, DOI: 10.1039/NP9961300479), vaccins), soit de faire des réactions de couplage entre deux molécules comportant de nombreuses fonctions.

Exemple de synthèse d'un anticorps catalytique:



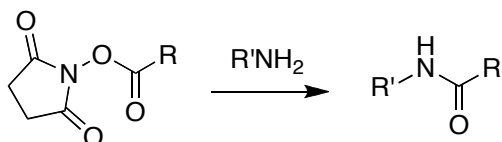
Afin d'atteindre ces objectifs, il est nécessaire d'avoir des synthèses chimiques chémosélectives (on va faire une réaction uniquement avec un type de fonction) et qui se font dans des conditions douces (il ne faut pas détruire la structure des molécules).

5.1 Modification de résidus nucléophiles

Nous allons profiter de la présence de résidus nucléophiles à la surface de protéines pour fixer une molécule.

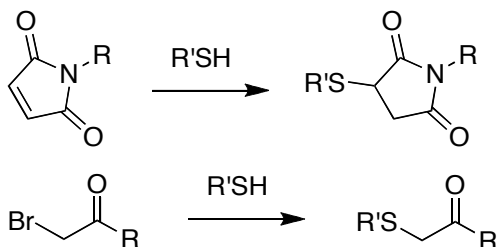
5.1.1 Réactions sur RNH₂

OSu (O succinimide):



5.1.2 Réactions sur RSH

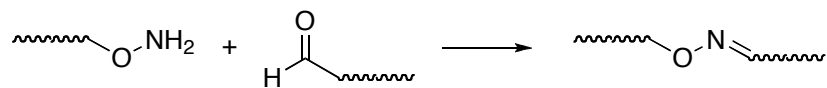
maléimide:



5.2 Réactions orthogonales (bioconjugaison)

Il est possible de trouver des réactions chimiques chémosélective qui sont compatibles avec l'ensemble des fonctions trouvées dans des acides aminés. On parlera ici de "chémosélective ligation". Ceci permet par exemple de placer sélectivement un groupe fluorescent sur une protéine que l'on aura modifié préalablement pour posséder une fonction spécifique.

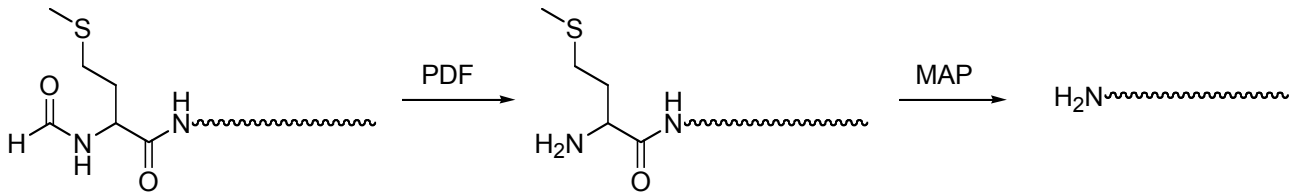
Exemple de réaction chémosélective: la formation d'oxime.



6 Peptide deformylase

6.1 Introduction

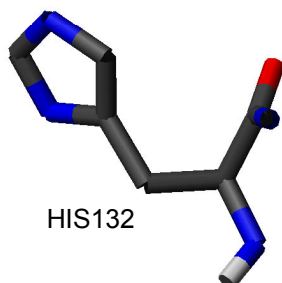
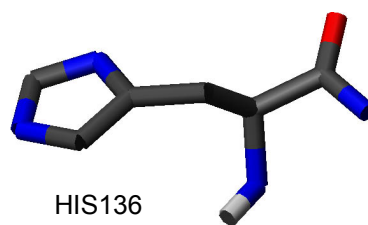
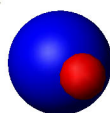
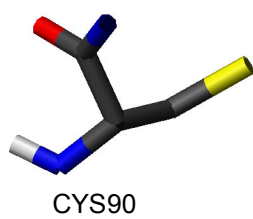
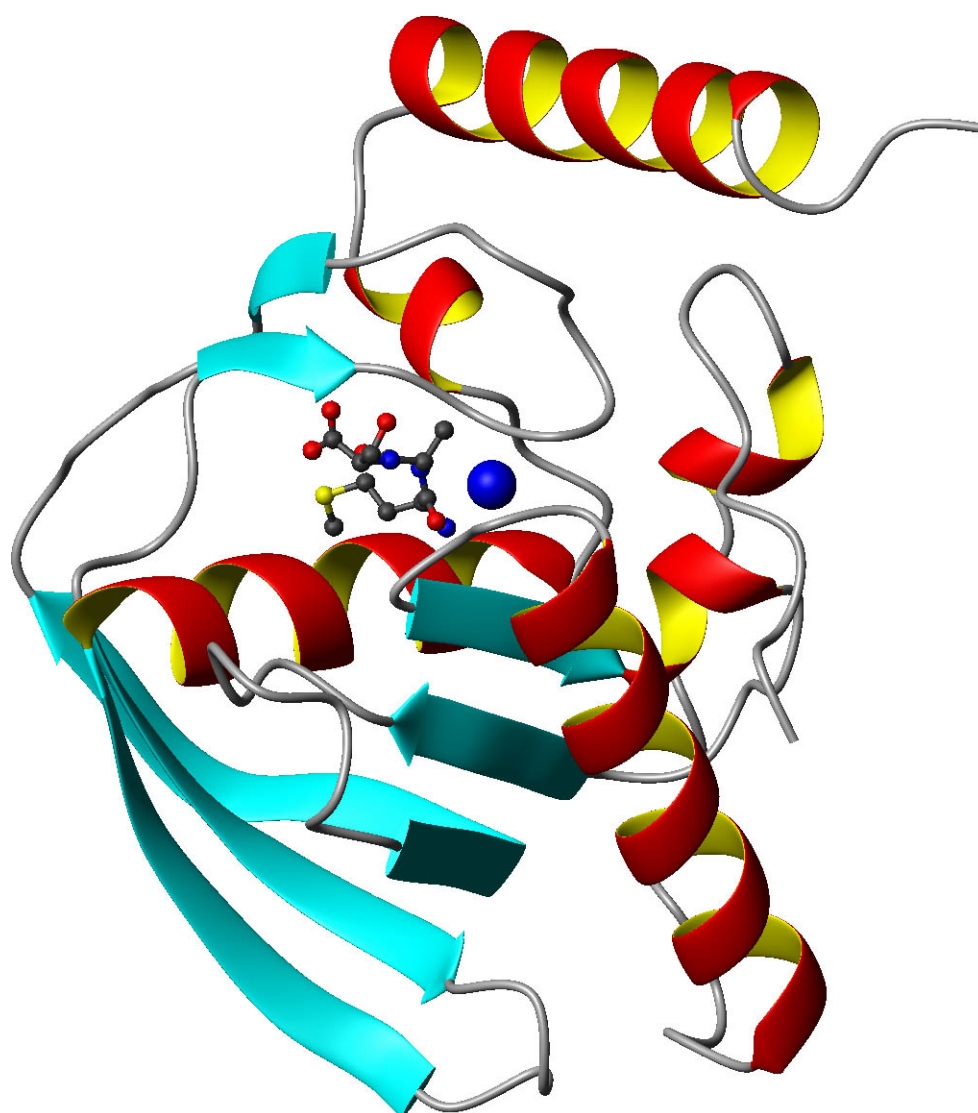
La synthèse protéique chez les procaryotes débute par une N-formylméthionine préalablement obtenue par une étape de formylation du côté N-terminal. La PDF **catlyse** le clivage du groupement formyl de la chaîne peptidique naissante. Ensuite la N-méthionine aminopeptidase (MAP) excise la méthionine N-terminale du peptide.



Ce cycle de formylation et déformylation distingue nettement les bactéries des cellules animales et végétales et fait de la PDF une cible de choix pour la recherche de produits pharmaceutiques à visée antibiotique.

Il arrive souvent que le produit de la réaction soit en partie inhibiteur. Ainsi HMetAlaSerOH est un inhibiteur (faible) de la protéine.

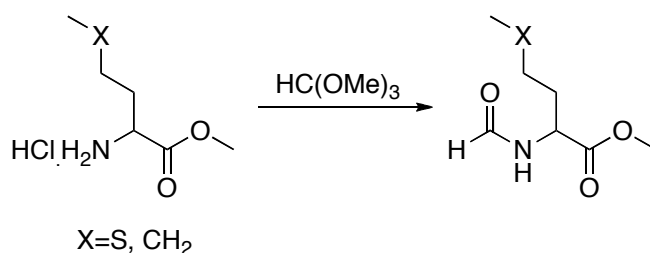
PDF est une métalloprotéine : stabilisation par deux histidines, une cystéine et une molécule d'eau. L'inhibiteur se trouve au voisinage du métal indiquant qu'il participe très certainement à la catalyse.



6.2 Recherche d'un substrat minimal

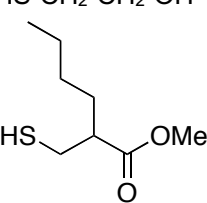
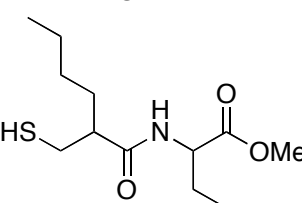
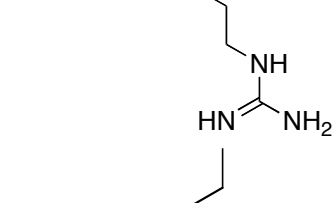
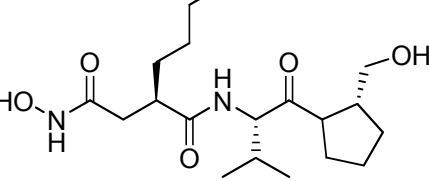
Mesure de la cinétique d'hydrolyse de différents produits (Biochemistry 1999, 4287) :

Peptide	vitesse relative
Fo-Ala-Ala-Ala-OH	0.20
Fo-Met-Ala-Ser-OH	100
Fo-Nle-Ala-Ser-OH	115
Fo-Leu-Ala-Ser-OH	11
Fo-Met-OMe	27
Fo-Nle-OMe	59
Fo-Met-OH	0.16
Fo-Leu-Ala-OMe	267
Fo-Nle-Arg-NH ₂	839

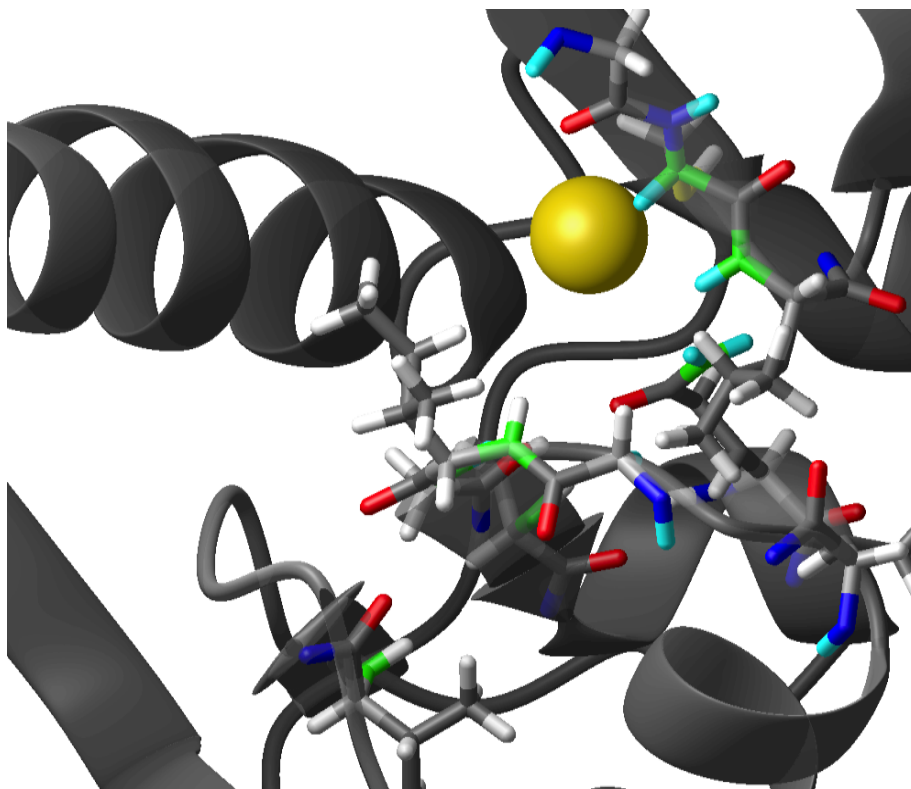


6.3 Recherche d'un inhibiteur

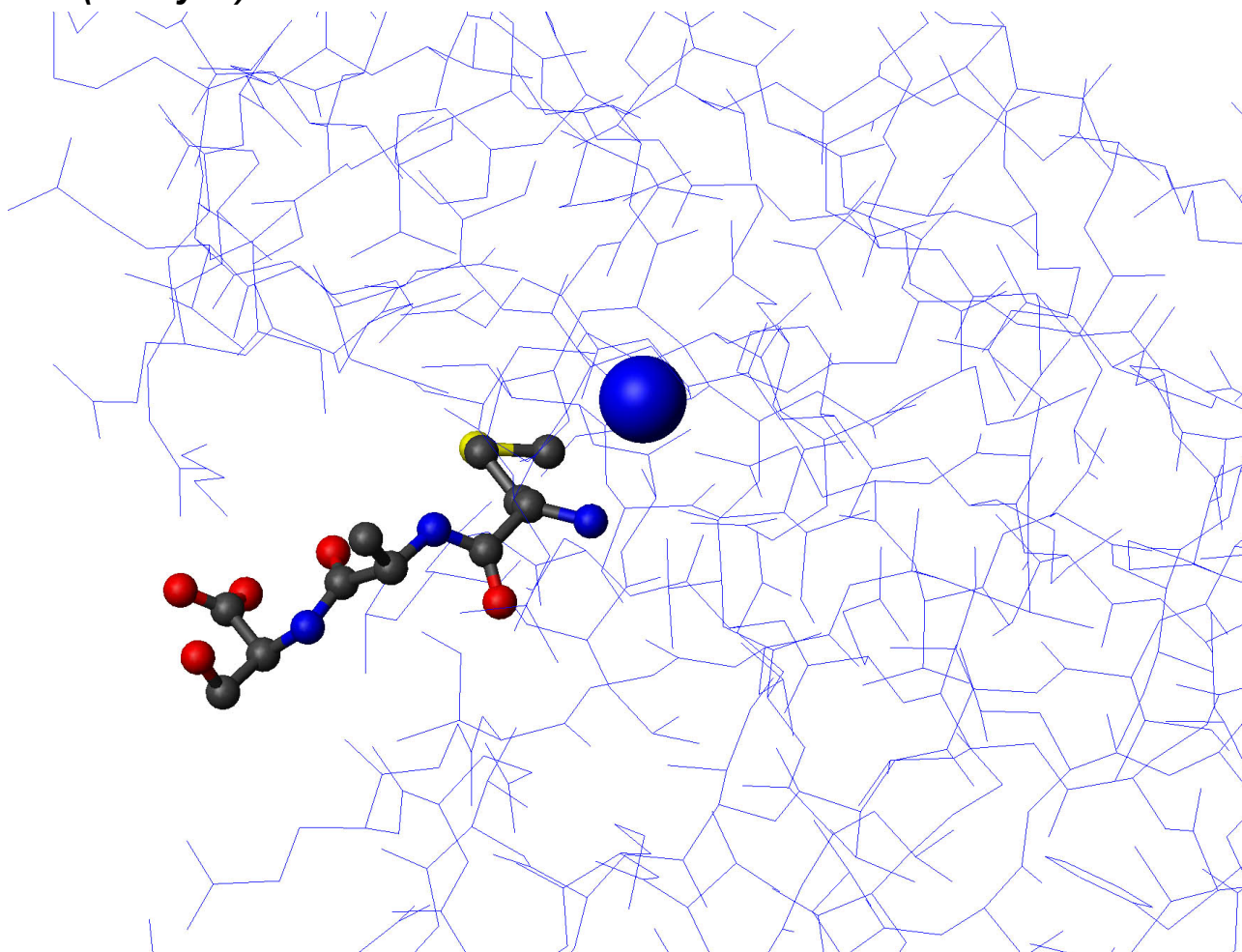
Synthèse d'inhibiteurs : on va donner un bon ligand à l'atome de métal : un RSH ou un RONH₂.

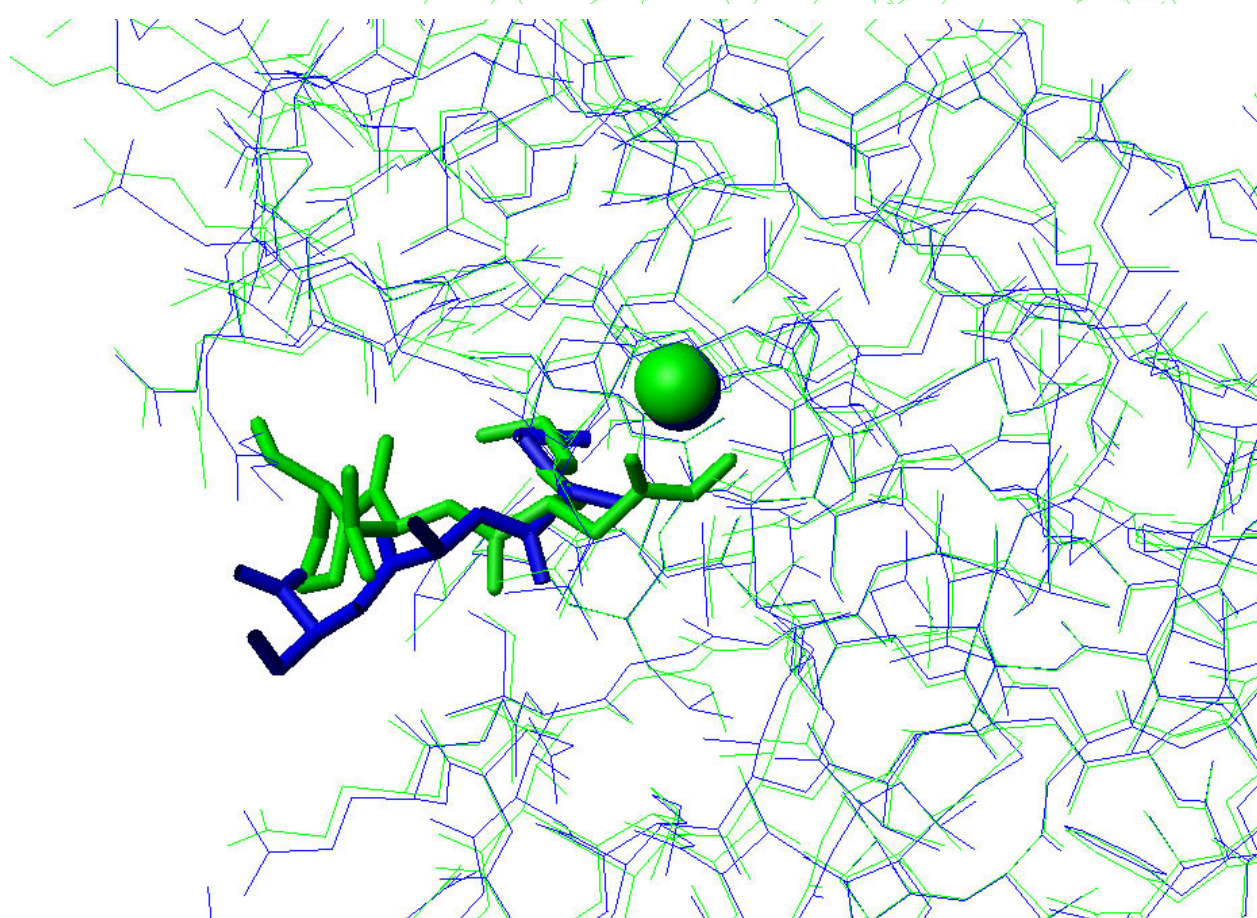
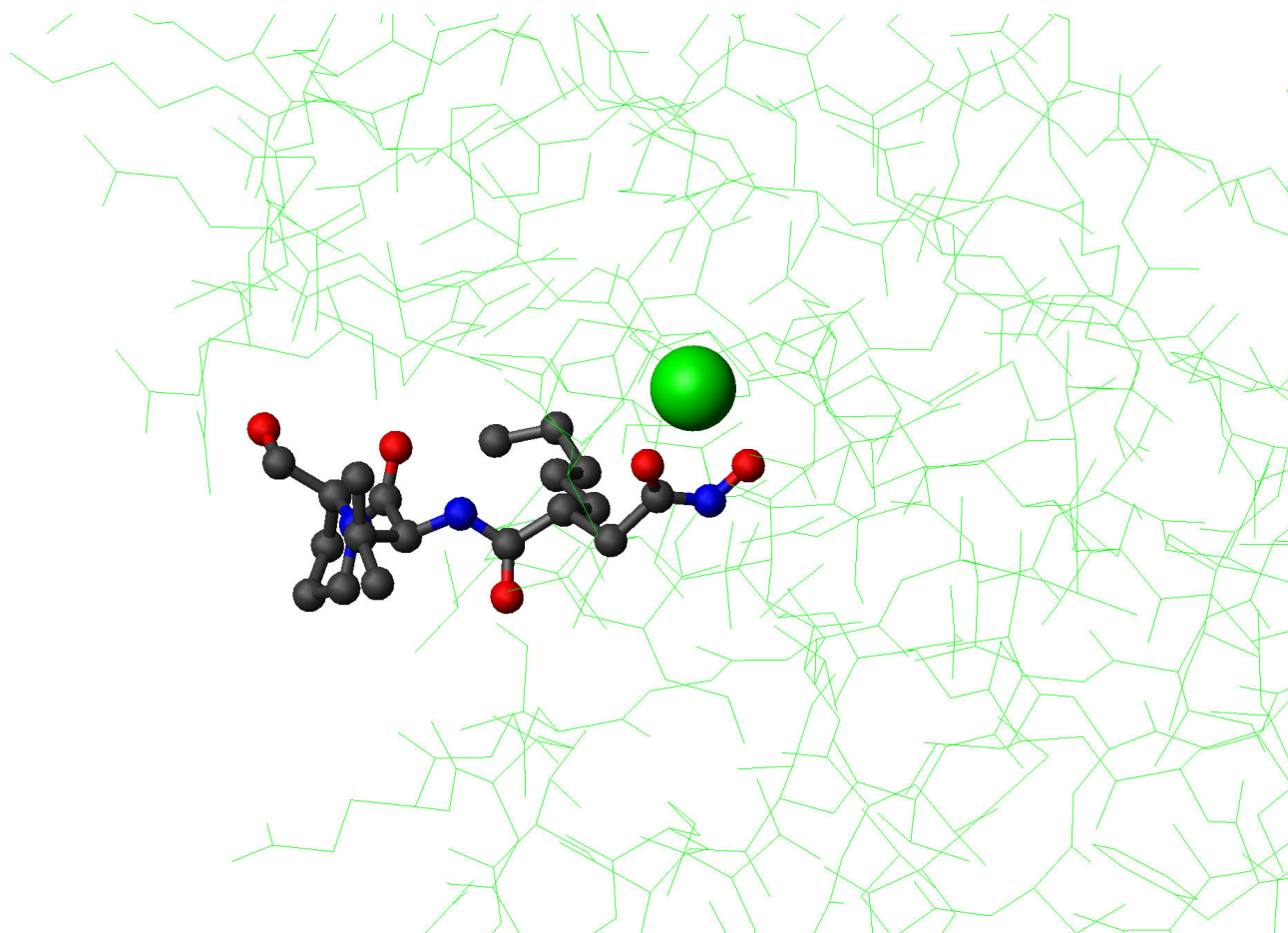
Peptide	K _i (μM)
H-Met-Ala-Ser-OH	53000
H-Met-Arg.OH	3800
HS-CH ₂ -CH ₂ -OH	10000
	52
	2.5
	64
	actinonine*

* Antimicrobial Agents and Chemotherapy 2001, 563-570.



6.5 Partie de reconnaissance (sélectivité), partie de réaction (catalyse)





6.6 Structures RX et stabilisation du complexe protéine - inhibiteur

6.6.1 Ponts hydrogènes

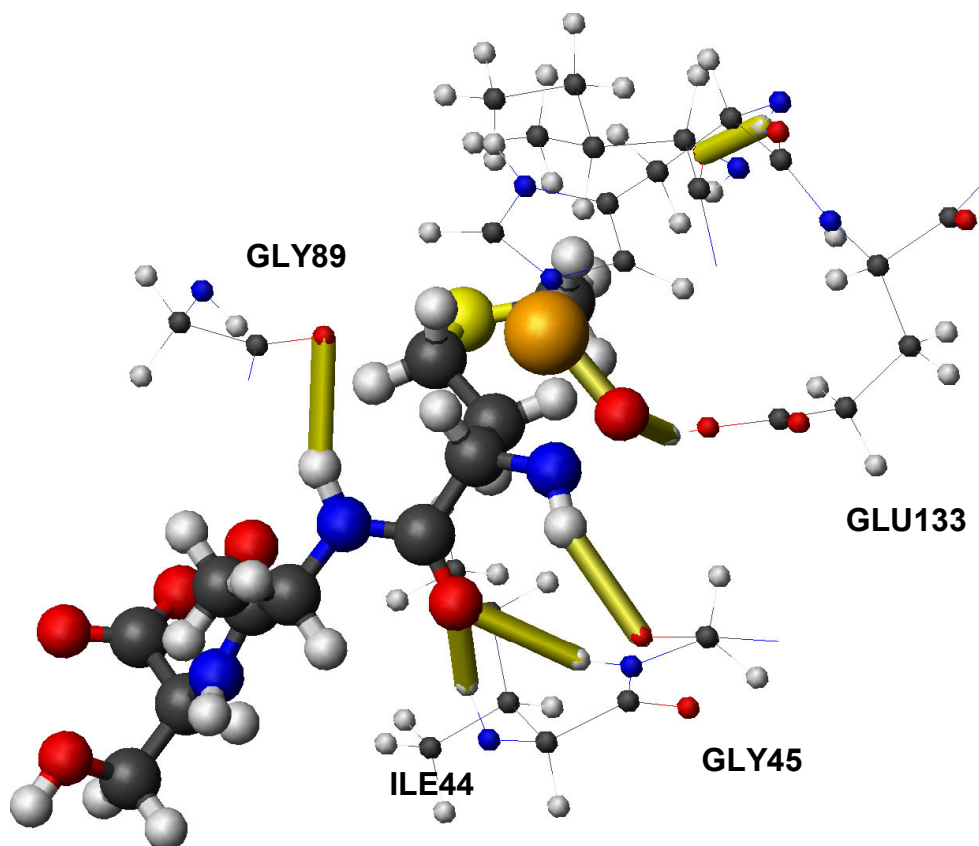
Interaction entre un C=O et un NH ou OH

C=O: amide, ASP, GLU, GLN, ASN

N-H: amide, ASN, GLN, LYS

O-H: SER, THR, TYR

Environ 6 kJ par mole



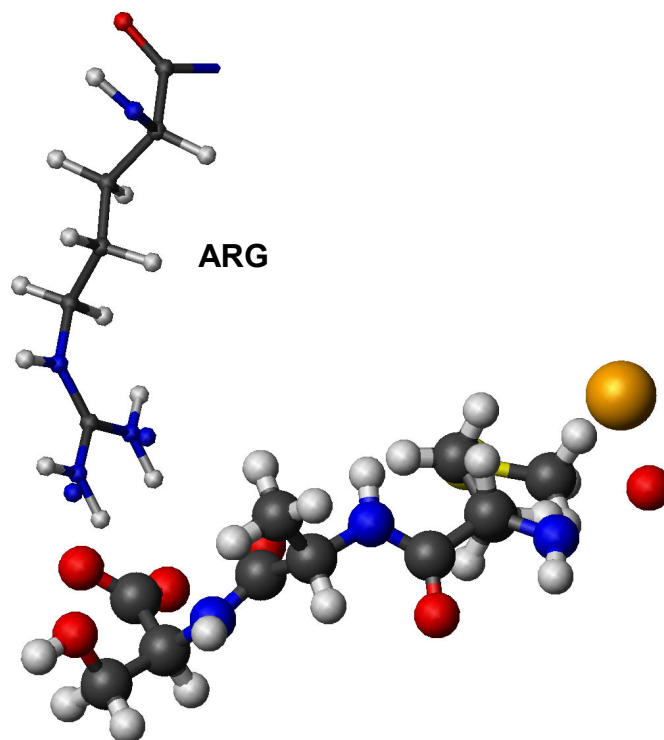
6.6.2 Effets électrostatiques

Une charge positive stabilise une charge négative via une interaction électrostatique.

En surface de la protéine: ~4 kJ/mole

Au coeur de la protéine: ~15 kJ/mole

Exemple : Asp / Arg



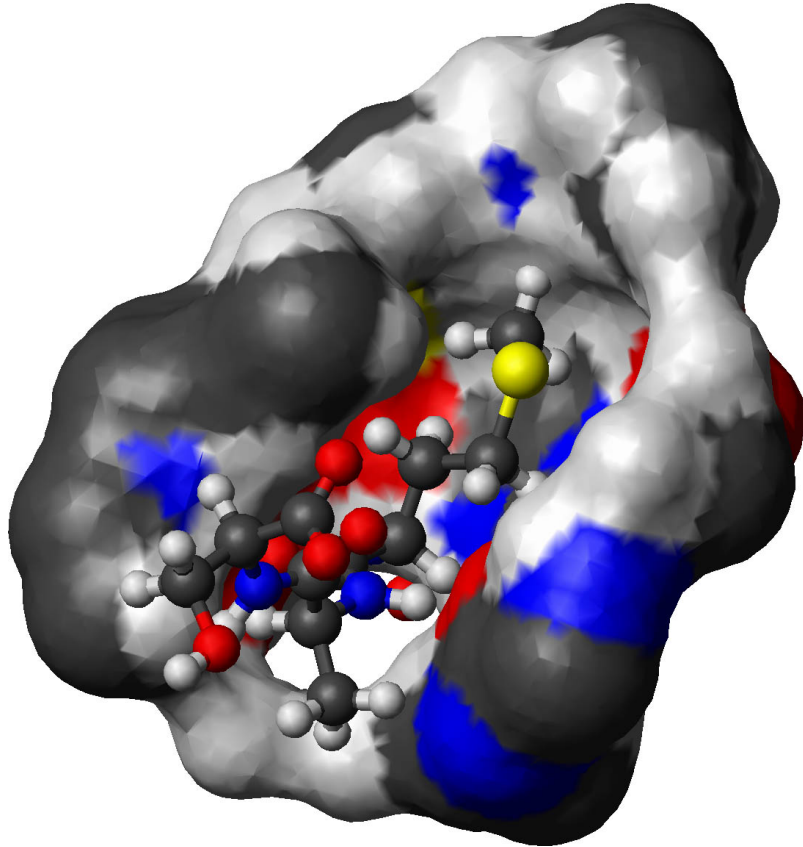
6.6.3 Effets hydrophobes

perte de 1 CH₂ dans une protéine : perte de ~4 kJ/mol

Hartley, G., (1936) *Aqueous solutions of paraffin-chain salts*, Hermann & Cie., Paris, as quoted in Tanford, C. (1973) *The Hydrophobic Effect: Formation of Micelles and Biological Membranes*, p viii, John Wiley & Sons, New York.

"The antipathy of the paraffin chain for water is, however, frequently misunderstood. There is no question of actual repulsion between individual water molecules and paraffin chains, nor is there any very strong attraction of paraffin chains for one another. There is, however, a very strong attraction of water molecules for one another in comparison with which the paraffin-paraffin or paraffin-water attractions are slight."

L'eau a tendance à s'organiser autour des molécules apolaires ce qui diminue l'entropie.
 $\Delta G = \Delta H - T\Delta S$.



7 Conclusions

RMN:

- Déterminer la structure de petites molécules
- Déterminer la structure de protéines ou l'interaction en inhibiteur et protéine

Importance de la structure de la protéine:

- Déterminer le site catalytique
- Comprendre le mécanisme de la réaction (par exemple hydrolase)
- Améliorer un inhibiteur (charge, poche hydrophobe, ...)

Réactions chimiques:

- Problème des groupes protecteurs
- Possibilité de faire des réactions chimiosélective