



HAL
open science

Pourquoi les MATH et l'intelligence artificielle sont important pour comprendre et améliorer le fonctionnement de protéine ?

David Allouche

► **To cite this version:**

David Allouche. Pourquoi les MATH et l'intelligence artificielle sont important pour comprendre et améliorer le fonctionnement de protéine ?. Fête de la science, Oct 2018, Castanet-Tolosan, France. 18 p. hal-02733726

HAL Id: hal-02733726

<https://hal.inrae.fr/hal-02733726>

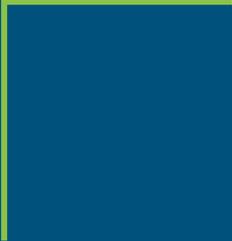
Submitted on 2 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pourquoi les MATH et l'intelligence artificielle sont important pour comprendre et améliorer le fonctionnement de protéine ?

intelligence artificielle/
biologie/chimie / physique présentation journée
sciences porte ouvertes inra 2018



pourquoi des proteines ?



adaptation environnementale à l'échelle moléculaire

les aptitudes d'un organisme à survivre dans son environnement sont conditionnées par des propriétés à des échelles multiples :

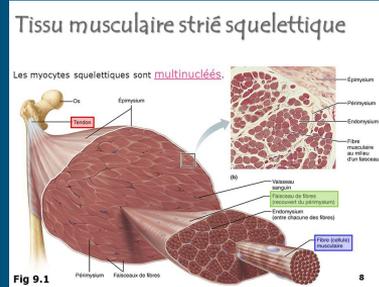
ces échelles vont de l'organisation de l'anatomie , des organes , des tissus des cellules qui les composent mais aussi de la composition et des propriétés des molécules qui les composent.

dans cet exposé nous allons nous intéresser principalement à l'échelle moléculaire. échelle ultime de l'évolution pour optimiser l'adaptation physico chimique du monde qui l'entoure lui permettant ainsi de survivre dans son biotope .

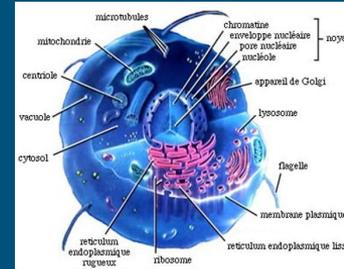
Pourquoi des Nouvelles Protéines ?

dans la nature on observe des animaux des plantes avec des propriétés différentes :
capacité propriété sont conditionnées à différentes échelles:

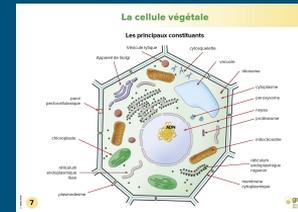
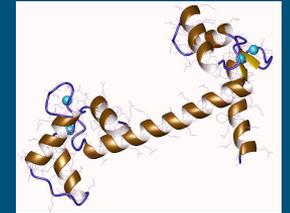
Animal



Cellule



Molécule :

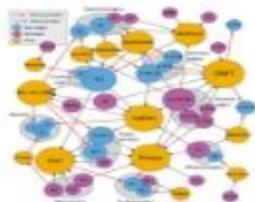


Multiscale Modeling of Drug Actions

Understanding of dynamics and kinetics of protein-ligand interactions

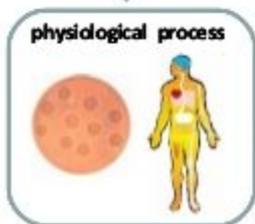


Prediction of molecular interaction network on a genome scale



Traditional Approach

Knowledge representation and discovery & model integration



Systems-based Approach

Reconstruction, analysis and simulation of biological networks

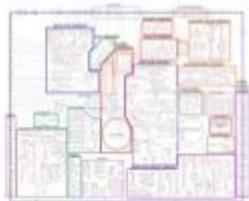


illustration : adaptation en fonction de manque d'oxygène

quand on est un animal un oiseau un reptile , et que l'on vit dans l'eau ,
en fonction de milieux ou l'on vit : riviere , lac , mer / océans

de la manière dont on se nourrit , de sa position dans la chaîne alimentaire , on peut avoir besoin de passer plus ou moins longtemps sous l'eau sans oxygène.

Au fils des générations les espèces se sont adaptés par pression de sélection pour survivre les traces de cette évolutions se retrouvent à toutes les échelles

y compris à l'échelle moléculaire

quizz :

quelles sont les espèces du regnes animal , qui selon vous sont capable de rester le plus longtemps sous l'eau ?

Les 10 plus longues apnées du règne animal

N°10 L'hippopotame et le castor ⇒ (15 min)



N°9 Le Manchot (empereur) ⇒ (18 min)



N°8 L'homme (22 min 32 secondes)



N°7 La baleine à bosse

la baleine bleue et l'orque (30 min)



N°6 Le cafard ⇒ (45 min)



N°5 L'iguane marin ⇒ 1h



N°4 Phoque de Weddel et l'éléphant de mer
(1h20)



N°3 Cachalot - Crocodile australien des eaux salées
(2 heures)



N°2 Serpent de mer

(4 heures)



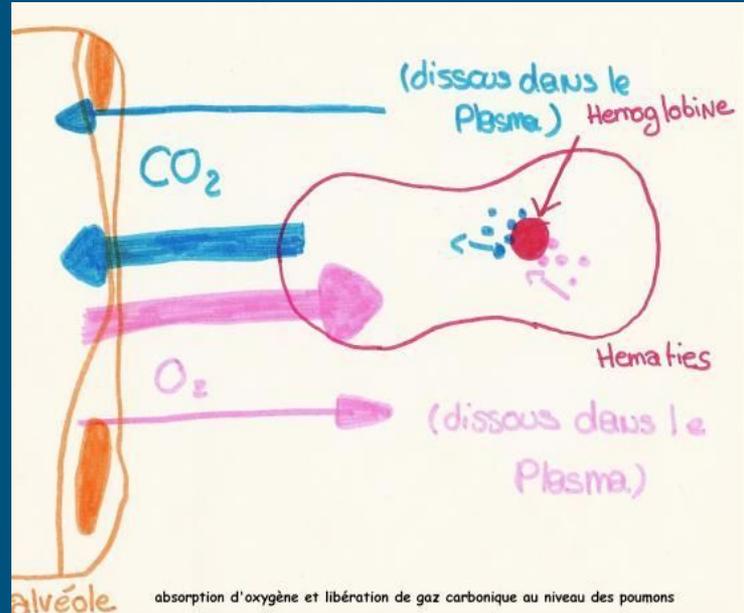
N°1 Tortue marine

(6 heures)

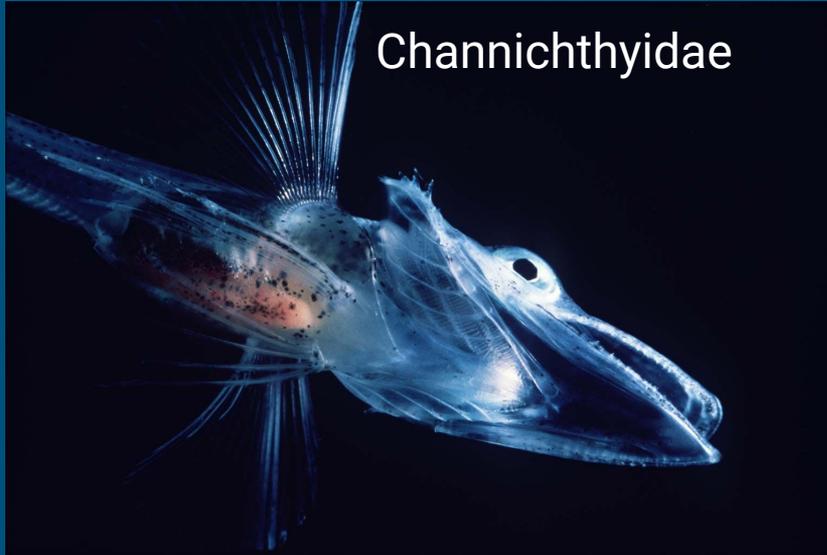
La championne de ce classement est donc un reptile. La tortue marine peut rester 6 heures sous l'eau, nager rapidement à 35 km/h et peut vivre plus 100 ans.



transport de l'oxygene dans le sang



Poissons des glaces pas de sang rouge !)



Leur sang est transparent car il ne contient pas d'hémoglobine ni de globule rouge. L'oxygène passe par simple diffusion, des branchies au sang, puis du sang aux organes. .

Les eaux australes qui entourent l'Antarctique constituent un environnement extrême pour la faune qui y vivent. La température y est constamment proche de 0°C, avec durant l'hiver austral des températures pouvant aller jusqu'à près -2°C.

ressources

les plus longues apnées:

<https://www.spotmydive.com/fr/news/les-10-plus-longues-apnees-du-regne-animal-plongee-sous-marine> :

video : de ADN a la protein

<https://www.youtube.com/watch?v=gG7uCskUOrA>

l'hémoglobine

<https://www.youtube.com/watch?v=MW0avyLWVDM>

exemple de protéine: hémoglobine:

<https://www.youtube.com/watch?v=qs3xONv548I>

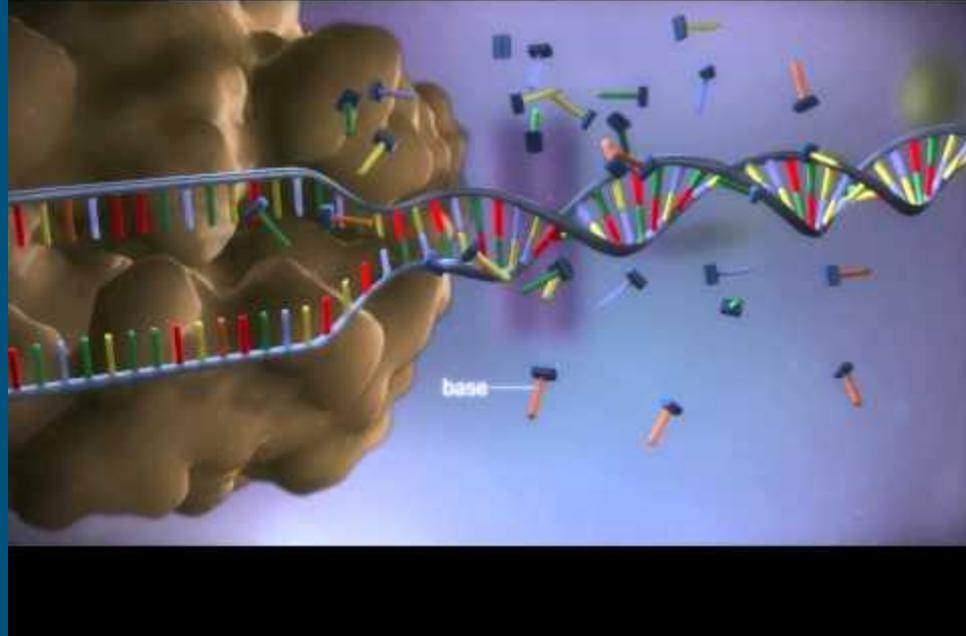
les structures: la PDB (base de données des structures experimentale de proteines)

<http://pdb101.rcsb.org/>

<https://libmol.org/>

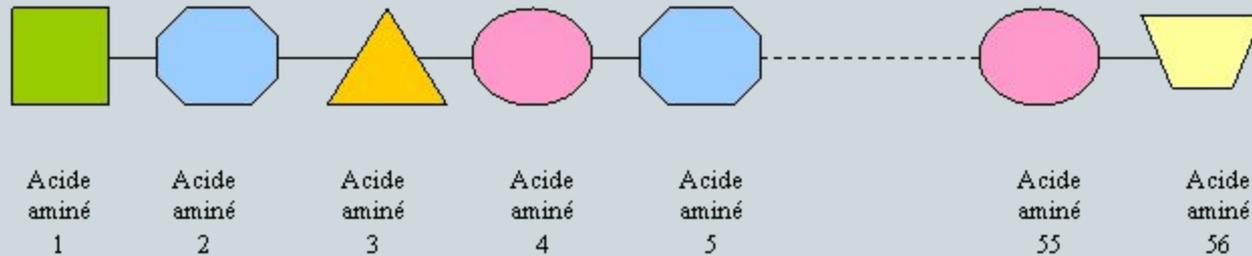
<https://fr.wikipedia.org/wiki/Channichthyidae> (poissons des glaces sans hemoglobines)

production du gene à la proteins



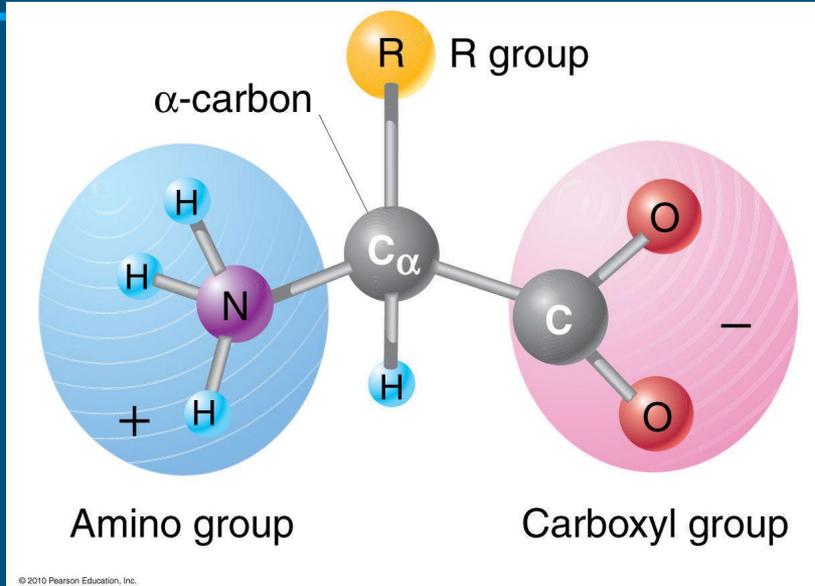
une protéine c'est quoi donc ?

un collier de perles dans un espace 3D .

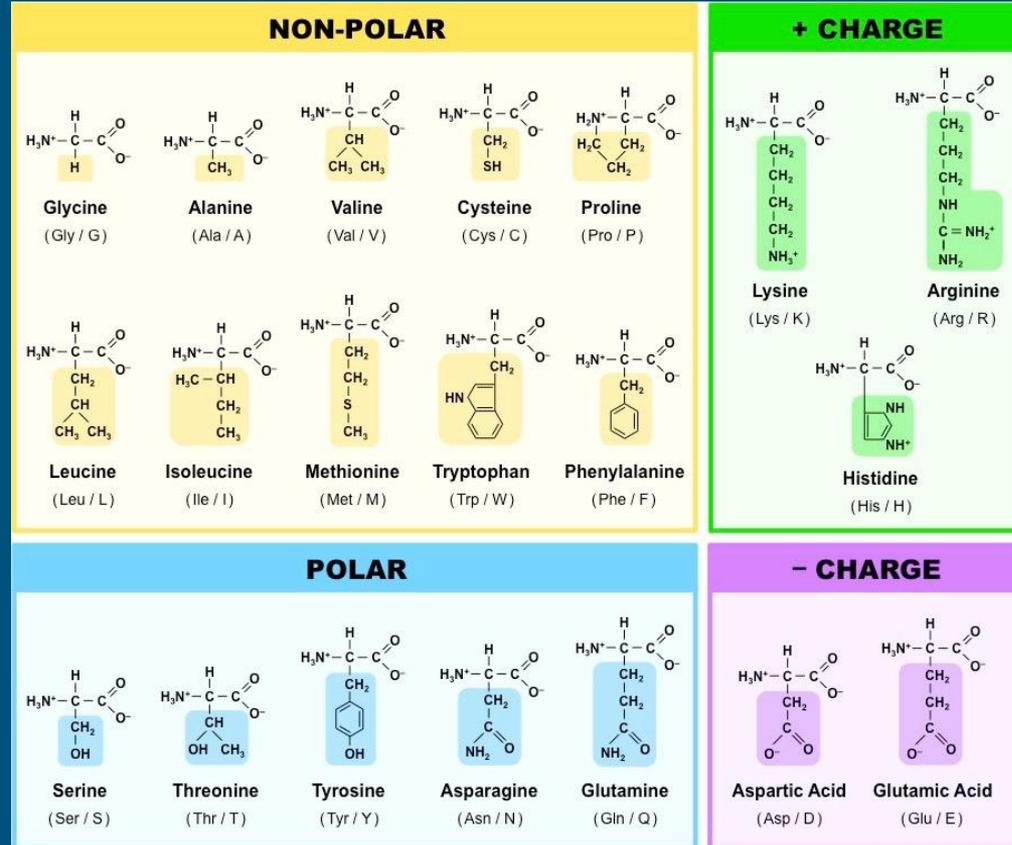


Protéine constituée de 56 acides aminés

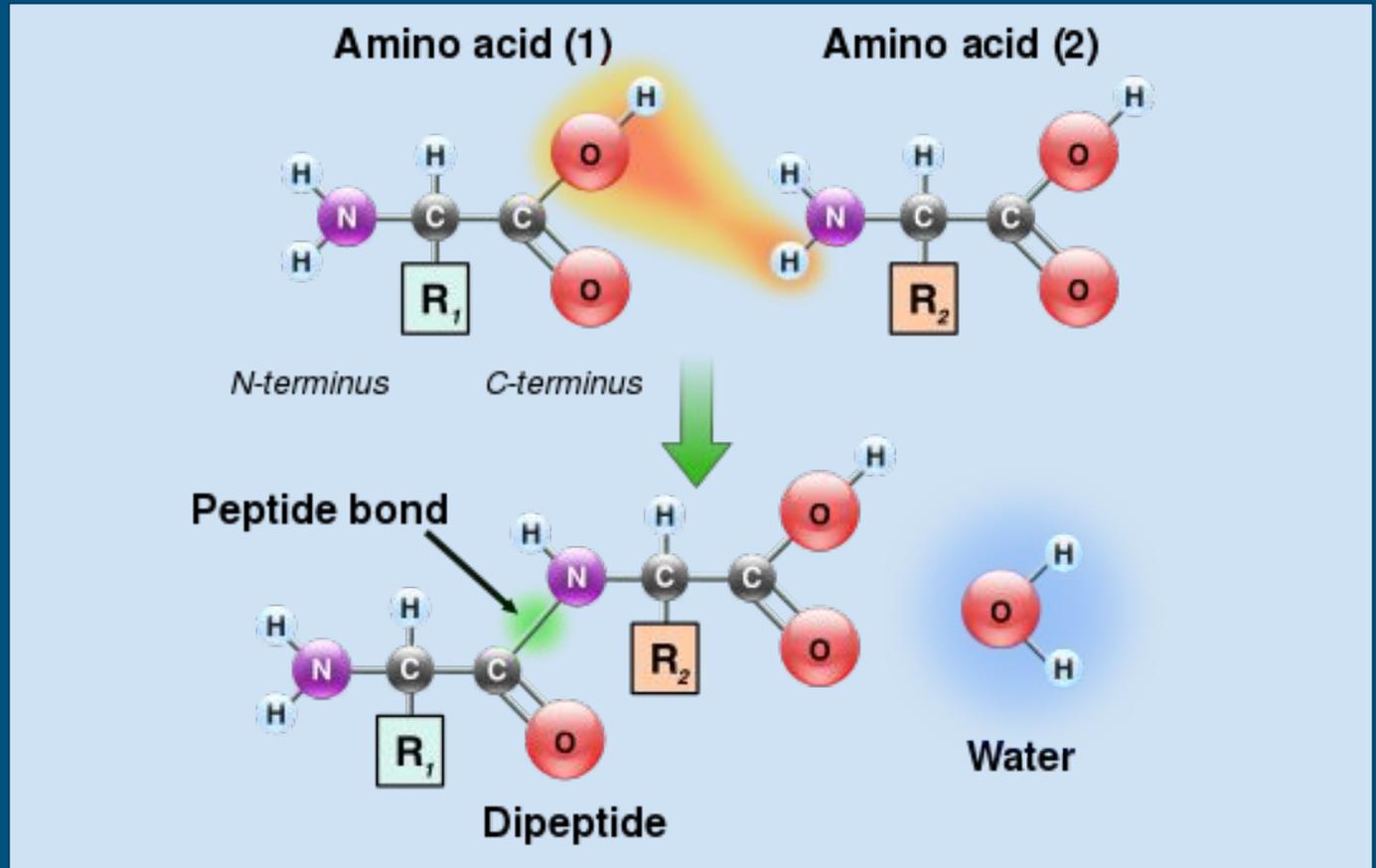
De la perle a l'AA



nombre de perle :



assemblage des perles \Rightarrow liaison peptidique



```

TRHB0_MYCLE/4-122 SFYDAI GGAET.FKAI VSRFYAQVPEDEI.LRELYP.ADDLAGAEERLRMFLEQYWGGRPTYSQRGHPRLRMRHAPFR
Q7CX73_AGRFC/7-127 TLYEAI GGDAT.VRAL TRRFYELMDTLPEAARCRATH.PADLSGSEAKFYDYLTGYLGGPPVYVEKHGHPMLRRRHFAVAP
Q7CX73_AGRFC/7-127 (SS) -HHHHHTCHHH.HHHHHHHHHHHHHH--GGGHHHHCS.-SSTHHHHHHHHHHHHHTTS--HHHHHS--HHHHHTTS-
Q7APC3_RHIME/8-128 TLYEAI GGDAT.VRAL TQRFYELMDSLPEAARCRAVH.PPDLTGSEEKFEYELTGWLGGPPIYVQKRGHMPLRRRHFIAG
Q8XXM3_RALS0/11-128 LAFDLLGGEAR.VRELADRFDYLDMLPEAFELRALHP.PSLDGSRDKLFWFLCGWLGGPSHYIERFGHPRLRARHLPFE
Q63VR4_BURPS/25-145 TAFEVLVGGEAR.VRELVDRFYDLDLEPEFAGIRALH.PPTLEGSRDKLFWFLCGWLGGPDHYIERFGHPRLRARHLPFP
F2CUR6_HORVD/26-145 NVFEALGGTPOPFDVLDSTNFFYTRVYEDDEEWFREIFS.GSKKEDAIQNQYEFVQRMGGPPLESQRGRHPALIGRHRFPF
GLB3_ARATH/26-142 NLFDKLGL.QT.FINLSTNFFYTRVYDDEEWFQSFIS.NSNKEDAIQNQYEFVQRMGGPPLESQRGRHPALIGRHRFPF
GLB3_ARATH/26-142 (SS) -HHHHHH.HH.HHHHHHHHHHHHT-S-CCCCGGGT.TS-HHHHHHHHHHHHTTS--HHHHHS--HHHHHTTS-
B2J6Y7_NOSP7/3-117 TLYDNI GGPAA.IEQVVDELHKRIATDSL.LSPIFA.GTDMAKQRNHLVAFLGQIFEGPKQY...GGRP.MDKTHAGLN
TRHBN_MYCTU/14-127 SIYDKI GGHGA.IEVVVEDFYVRVLAADDQ.LSAFFS.GTNMSRLKGGQVEFFAAALGGPEPYT...GAP.MKQVHQRG
TRHBN_MYCTU/14-127 (SS) -HHHHHTHHH.HHHHHHHHHHHHT-TT..TGGGGT.TS-HHHHHHHHHHHHTT-SS-.....S--.HHHHHTTS-

```

```

TRHB0_MYCLE/4-122 ITAIERDAWLRCMHTAVASIDSHTLDNEHRRELDYLEMAAHS
Q7CX73_AGRFC/7-127 IGP AERDEWLLCFRRAMDETIENAKLREIIVAPVERLAFHMQN
Q7CX73_AGRFC/7-127 (SS) --HHHHHHHHHHHHHHHHHH--HHHHHHHHHHHHHHHHHH--
Q7APC3_RHIME/8-128 IGP AERDEWLLCFTRALEETVSHPKLREIILEPITRLAHMQN
Q8XXM3_RALS0/11-128 IGTREDDQWRCMALAMODLG...IDDALQMLMQAFWQADW
Q63VR4_BURPS/25-145 IASSERDOWLRCIAWAMODVGLDEPLRERLMHSEHDTADWMRN
F2CUR6_HORVD/26-145 VTHRAAERWLHHMQQALETT..QSINPDTKTKMMNFRHTAYF
GLB3_ARATH/26-142 VTHQAAERWLEHMQNALDDSV..DIDQDSKIMMKFRHTAF.
GLB3_ARATH/26-142 (SS) --HHHHHHHHHHHHHHHHHH--S.-S--HHHHHHHHHHHHHHHH--
B2J6Y7_NOSP7/3-117 LQQPHFDAIAKHLGEAMAVRVS AENTKAALDRVTNMGALIN
TRHBN_MYCTU/14-127 ITMHFSLVAGHLADALTAAGVPS ETEITEILGVIAPLAVDVT.
TRHBN_MYCTU/14-127 (SS) -BHHHHHHHHHHHHHHHHHTT--HHHHHHHHHHGGGHHHS.

```

This alignment is coloured according to the ClustalX colouring scheme:

- Glycine (G)
- Proline (P)
- Small or hydrophobic (C,A,V,L,I,M,F,W)
- Hydroxyl or amine amino acids (S,T,N,Q)
- Charged amino-acids (D,E,R,K)
- Histidine or tyrosine (H,Y)

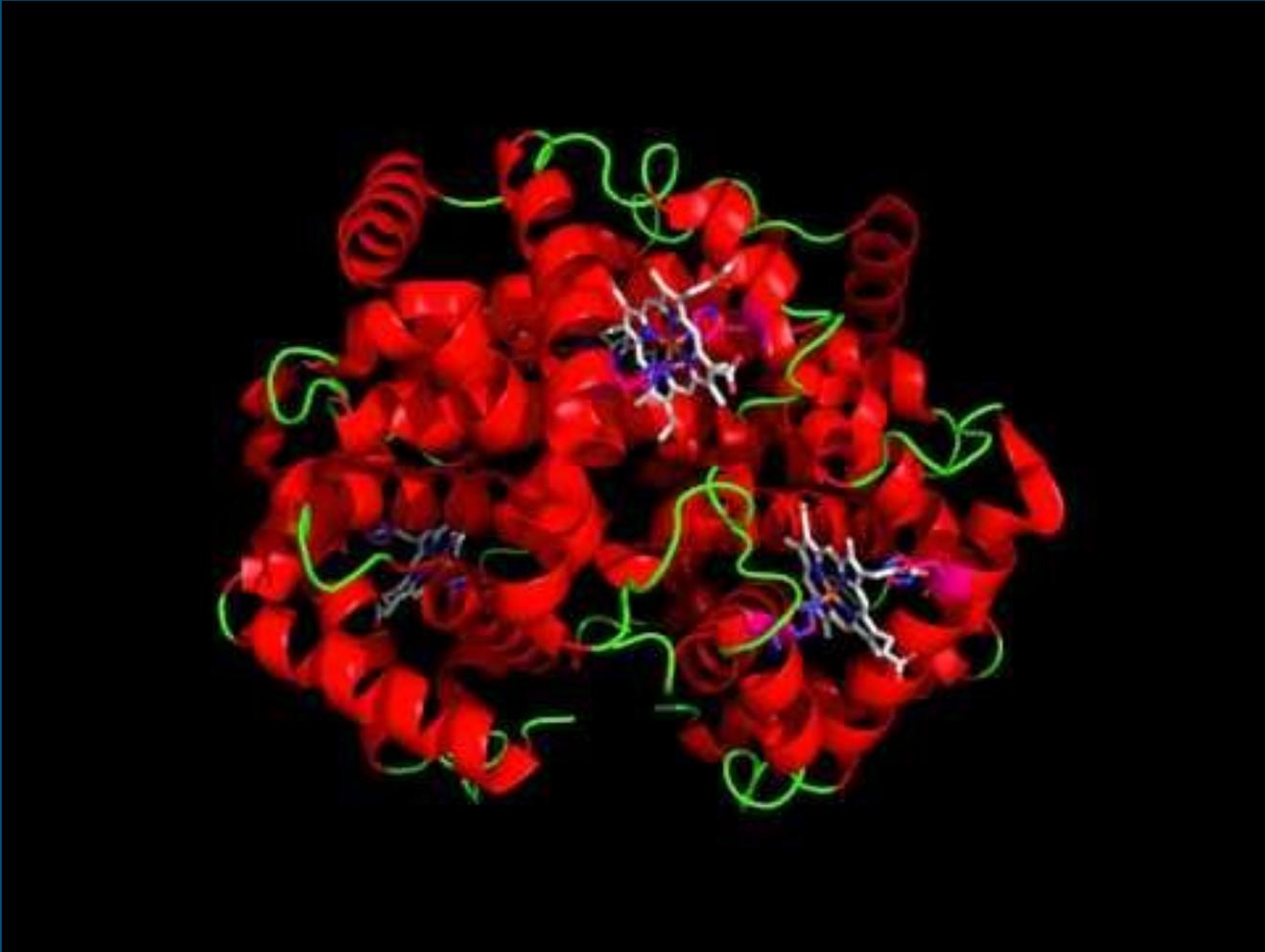
For UniProt-based alignments, we also add some additional mark-up to the alignments where appropriate. Active site information is shown as follows:

- Active site (residue annotated in SwissProt as an active site)
- Predicted active site (residue aligns in a Pfam alignment with a SwissProt active site)
- Predicted active site (residue annotated in SwissProt as a potential active site)

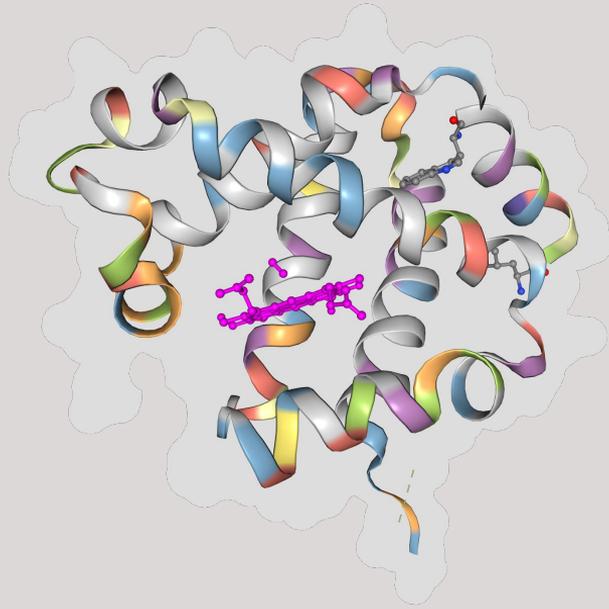
Some UniProt sequences can be mapped to protein structures, in which case we also show the secondary structure definition. These lines are shown below the sequence to which they apply and are marked (ss). The meaning of each of the symbols is as follows:

- C Random coil
- H Alpha-helix
- G 3(10) helix
- I Pi-helix
- E Hydrogen bonded beta-strand (extended strand)

—



<https://libmol.org/>



qu est ce que l'on cherche ?

la séquence qui est la plus stable énergétiquement

qui conserve la même fonction

et donc la même forme...3D \Rightarrow modulation de la fonction via la séquence

Pourquoi faire ???? (<http://pdb101.rcsb.org/>)

faire des nouveau médicament (cancer ⇒ anticorps catalytique, antibiotique , prion , Alzheimer)

- faire des nouvelles réactions chimique (enzym)

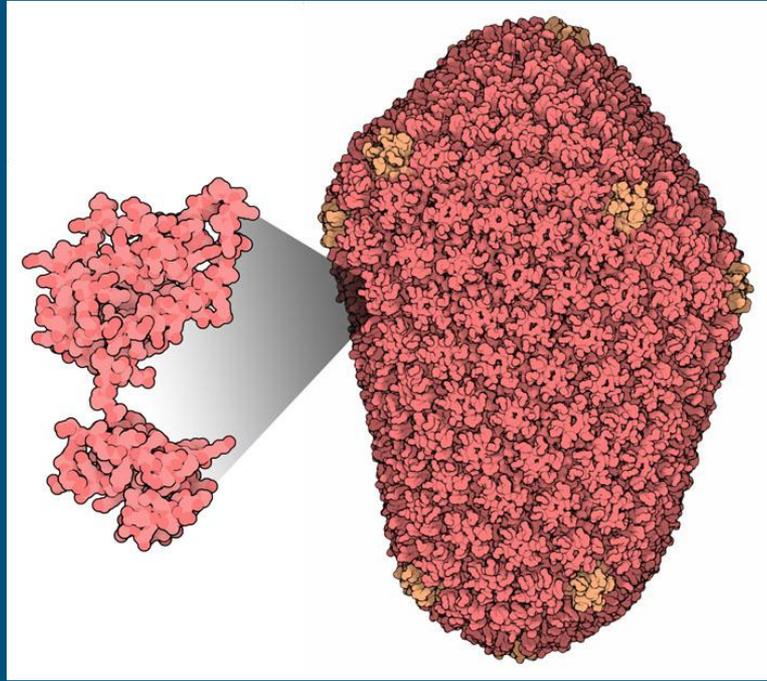
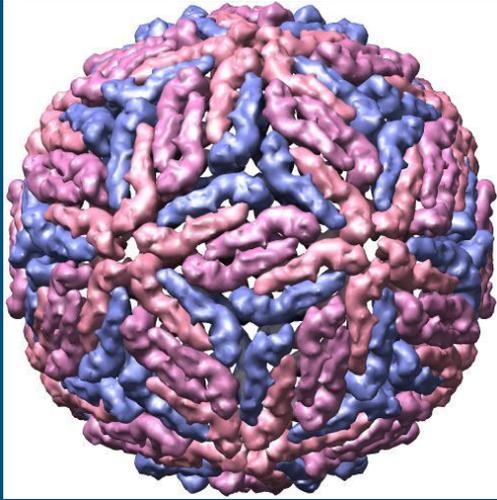
⇒ faire ce que fait la nature (produire du sucre a partir d'herbe)

du “pétrole” (des acides gras) a partir de membrane de micro organisme

⇒ energie renouvelable

⇒ comprendre comment fait la nature pour faire des auto assemblage parfait
virus

capsid



HIV Capsid

comment on aborde le problème



<http://pdb101.rcsb.org/learn/flyers-posters-and-other-resources/calendar/2018-calendar-mechanisms-of-antimicrobial-resistance>

De la protéine au modèle

problème:

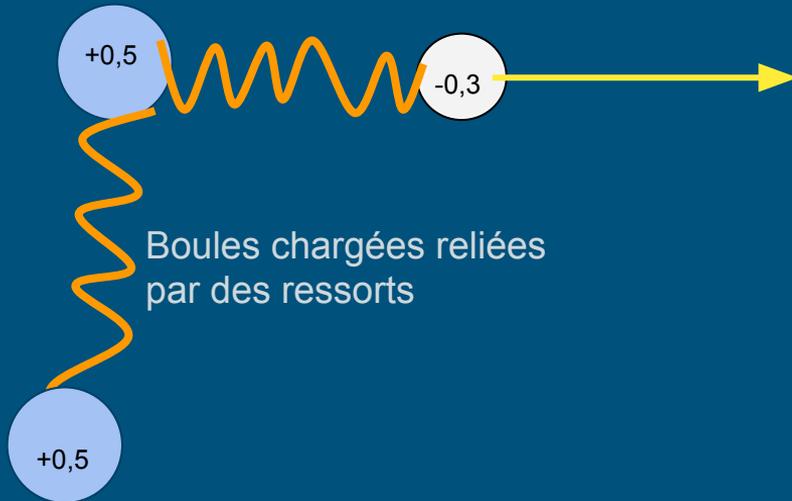
si une protéine est un collier de perle :

(en pratique) 20 perles différentes :

chaque perle peut avoir des géométrie différentes : ==> cf modele moleculaire:

de la molécule au programme : informatique

molécules \Rightarrow equation \Rightarrow programme informatique = model virtuel de molécule



$$V = \sum_{\text{bonds}} \epsilon_r (r - r_0)^2 + \sum_{\text{angles}} \epsilon_\theta (\theta - \theta_0)^2 + \sum_{\text{planar/improper dithedrals}} \epsilon_\xi (\xi - \xi_0)^2$$
$$+ \sum_{\text{backbone dithedrals}} \epsilon_{\phi,bb} F_d(\phi) + \sum_{\text{sidechain dithedrals}} \epsilon_{\phi,sc} F_d(\phi) + \sum_{\text{contacts}} \epsilon_c \left[\left(\frac{\sigma_{ij}}{r} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r} \right)^6 \right] + \sum_{\text{non-contacts}} \epsilon_{nc} \left(\frac{\sigma_{nc}}{r} \right)^{12}$$

programme informatique =
model virtuel de molécule

modele

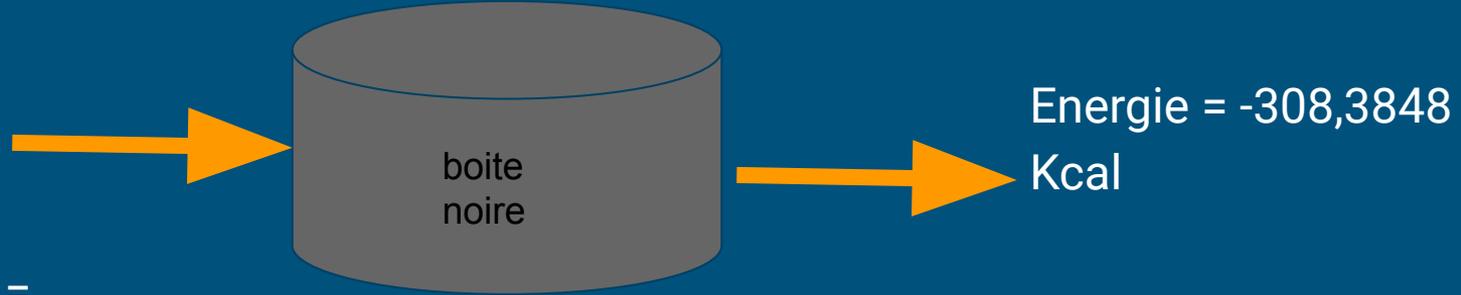
- modèle de molécule

- sequence

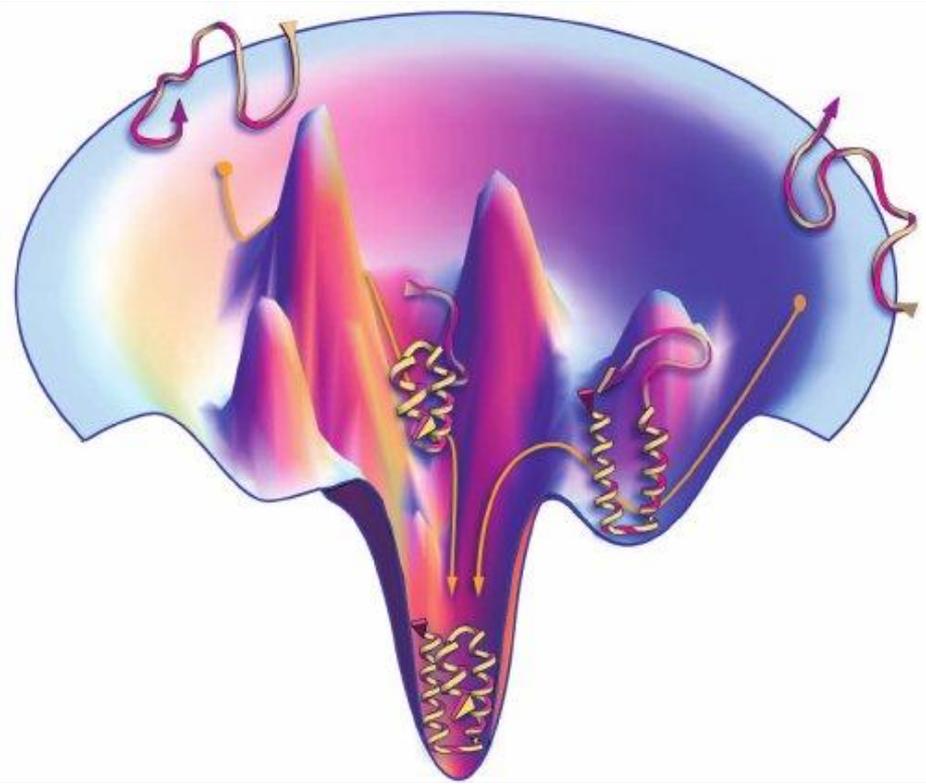
+

- coordonnée =

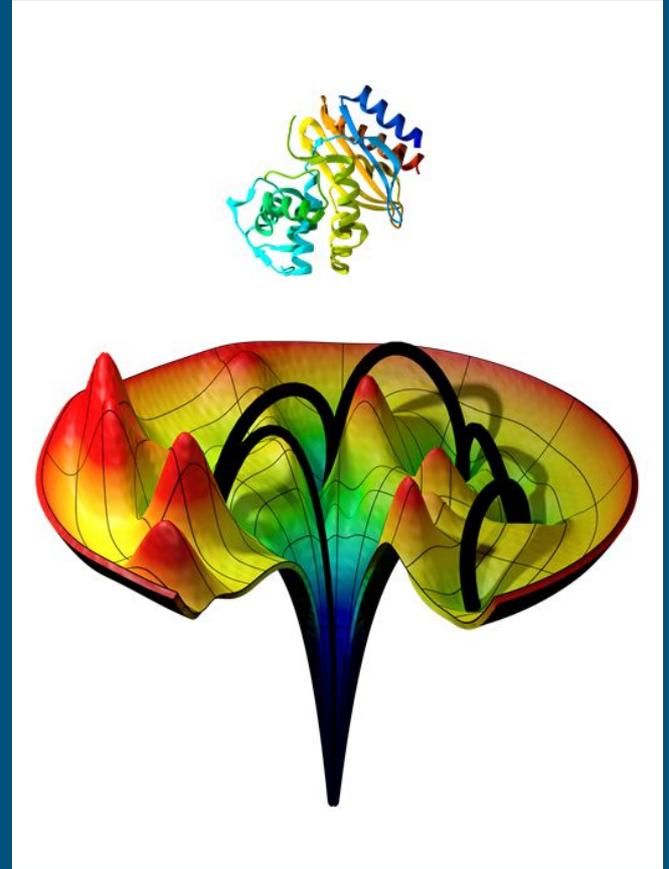
(X , Y , Z) des atomes



paysage de repliement



paysage de
mutation



explosion des possibilités:= explosion
combinatoire

quizz

des protéines a la combinatoire :

combien de collier différent : si on prend deux perles (rose et blue) .

- 1) rose ou bleu
- 2) rose ou bleu
- 3) rose ou bleu

$$2 \times 2 \times 2 = 2^{**3} = 8 \text{ colliers differents}$$

combien de collier avec 3 perles de 3 couleurs differentes (reponses)

$$n^{**C}$$

n = la longueur collier , $3 \times 3 \times 3 = 27$

si pour calculer l'énergie d'une protéine : on prends 0.1 seconde ==> 10^{99} seconde == $3.71 \cdot 10^{91}$ >> l'âge de l'univers (paradoxe de levintal).

en vraie il y a 20 perles et des géométries différentes (soit au total plusieurs 100 de possibilités par position)

pour une protéine de taille 10 et 100 possibilités , combien de possibilités au total ?

10^{100}

The Levinthal Paradox

- There are vastly too many different possible conformations for a protein to fold by a random search.
- Consider just for the peptide backbone, there are 3 conformations per amino acid in the unfolded state, For a 100 a.a. protein we have 3^{100} conformations.
- If the chain can sample 10^{12} conformations/sec, it takes 5×10^{35} sec (2×10^{28} year)

demo:

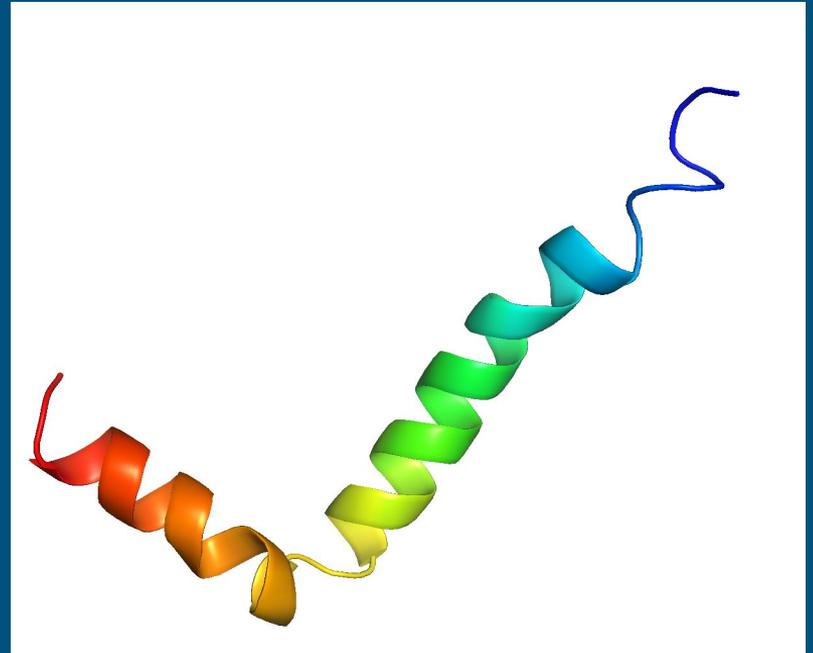
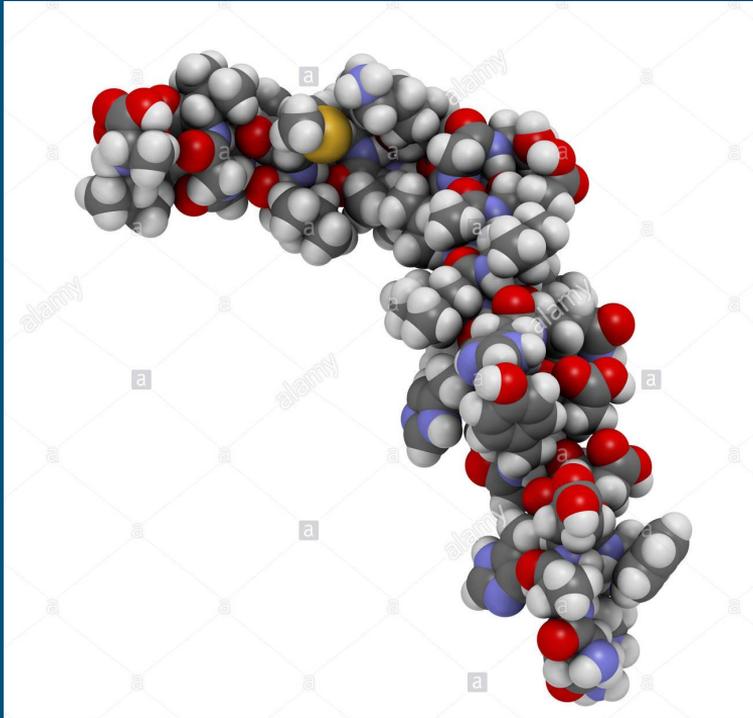


example peptide:

toulbar2:

**Bêta-amyloïde (Abêta) peptide.
Composante majeure de plaques dans
la maladie d'Alzheimer. Modèle de
structure secondaire de Cartoon,
coloration (hélices alpha bleu)**

42 Acides aminés



Merci d'être venu :

on utilise :

notre tête :

des connaissances dans différents domaines scientifiques:

intelligence artificielle

math / statistique

biologie (biochimie / biologie moléculaire)

chimie / physique moléculaire /

des ordinateurs / beaucoup de lecture / des présentations en anglais.

la rédaction de documents en français en anglais.