

TD : MOTIFS ET DOMAINES PROTÉIQUES

EXERCICE N°1 : Exemples de motifs protéiques de liaison à l'ADN

Q1.

Numéro d'accès UniProt du TFIIIA humain : Q92664.

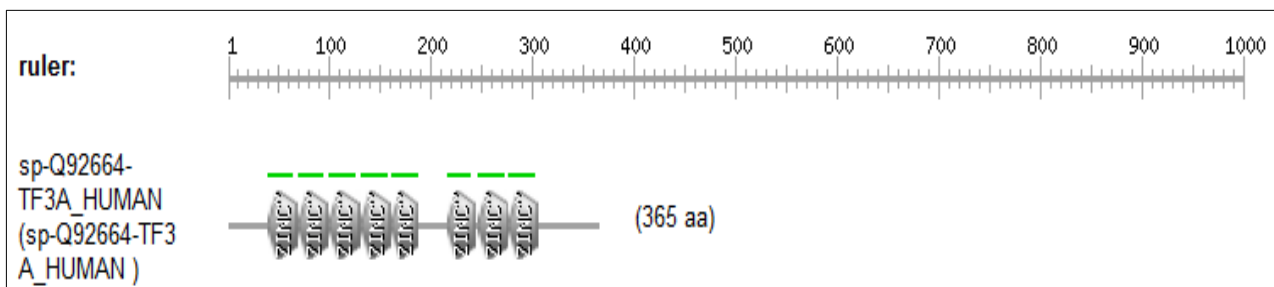
La fiche UniProt de cette protéine comprend les rubriques :

- 1) Fonction.
- 2) Noms et taxonomie.
- 3) Localisation subcellulaire.
- 4) Pathologie / Biotechnologies.
- 5) Données PTM / Traitement.
- 6) Expression.
- 7) Interaction protéine-protéine.
- 8) Structure.
- 9) Famille et domaine.
- 10) Séquence
- 11) protéines similaires.
- 12) Références.

La structure primaire (ou séquence) du TFIIIA humain au format FASTA est donnée ci après :

```
>sp|Q92664|TF3A_HUMAN Transcription factor IIIA OS=Homo
sapiens GN=GTF3A PE=1 SV=3
MDPPAVVAESVSSLTIAADAFIAAGESSAPTPPRPALPRRFICSFDPDCSANYSKAWKLD
LCKHTGERPFVCDYEGCGKAFIRDYHLSRHLTHTGKPFVCAANGCDQKFNTKSNLKKH
FERKHENQQKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSK
LKRHAKAHEGYVCQKGC SFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHM
KTHAPERDVCRCPREGCGRTYTTVFNLQSHILSFHEESRPFVCEHAGCGKTFAMQSLTR
HAVVHDPDKKKMMLKLVKKSREKRSLASHLSGYIPPKRKQGGLSLQNGESPNCVEDKML
STVALVTLG
```

Après le scan Prosite à partir de la séquence au format FASTA, on obtient le résultat suivant :



Le TFIIIA humain contient 8 motifs en doigt de zinc (numéro d'accès Prosite : PS50157), impliqués dans la liaison de TFIIIA à l'ADN.

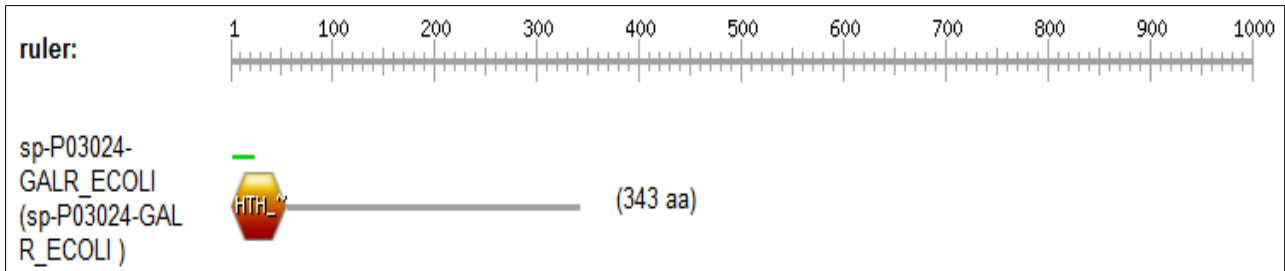
Q2.

Numéro d'accès UniProt du répresseur GalR de l'opéron galactose d'*Escherichia coli* : P03024.

La structure primaire (ou séquence) de GalR d'*E. Coli* au format FASTA est donnée ci après :

```
>sp|P03024|GALR_ECOLI HTH-type transcriptional regulator
GalR OS=Escherichia coli (strain K12) GN=galR PE=1 SV=1
MATIKDVARLAGVSVATVSRVINNSPKASEASRLAVHSAMESLSYHPNANARALAQQTTE
TVGLVVGDVSDPFFGAMVKAVEQVAYHTGNFLLIGNGYHNEQKERQAIEQLIRHRCAALV
VHAKMIPDADLASLMKQMPGMVLINRILPGFENRCIALDDRYGAWLATRHLIQQGHTRIG
YLCSNHSISDAEDRLQGYDALAESGIAANDRLVTFGEPEDESQGEQAMTELLGRGRNFTA
VACYNDSMAAGAMGVLNDNGIDVPGEISLIGFDDVLSRYVRPRLTTVRYPIVTMATQAA
ELALALADNRPLPEITNVFSPTLVRRHSVSTPSLEASHHATSD
```

Après le scan Prosite à partir de la séquence au format FASTA, on obtient le résultat suivant :



La protéine répresseur GalR contient un motif Hélice-Boucle-Hélice (HTH pour *Helix-Turn-Helix*) (numéro d'accèsion Prosite : PS50932), impliqué dans la liaison de GalR à l'ADN.

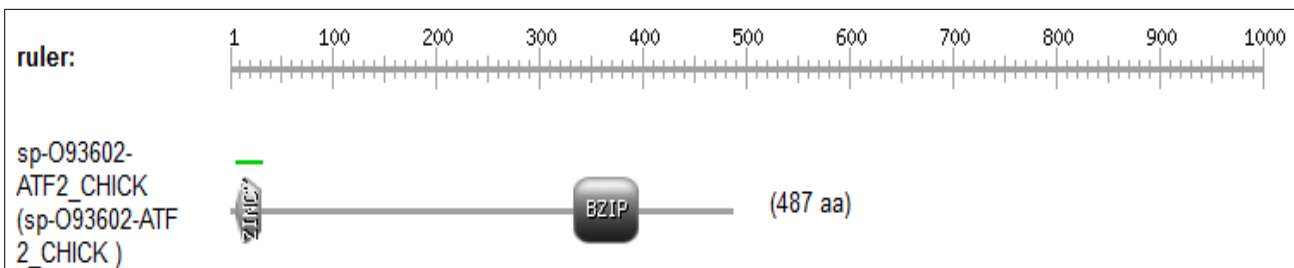
Q3.

Numéro d'accèsion UniProt du facteur de transcription II du poulet : O93602.

La structure primaire (ou séquence) du facteur de transcription 2 du poulet au format FASTA est donnée ci après :

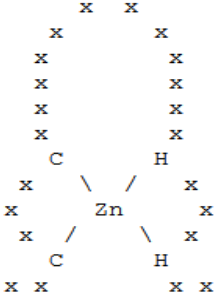
```
>sp|O93602|ATF2_CHICK Cyclic AMP-dependent transcription
factor ATF-2 OS=Gallus gallus GN=ATF2 PE=2 SV=1
MSDDKPFLLCTAPGCGQRFTNEDHLAVHKHKHEMTLKFPGPARNDVIVADQTPTRFLKN
CEEVGLFNELASPFENEFKKASEDDIKMPLDLSPLATPIIRNKIEEPSVETTHQDSPL
PHPESTTNDKEVSLQQTAPTSTIVRPASLQVNPVLLTSSDSSVVIQQAIPTSSTVI
TQAPSSNRPIVPVPGFPLLLHLPNGQTMPVAIPASITNSNVHVPAAVPLVRPVTMVPSI
PGIPGPSSQPQVQSEAKLRLKAALTQQHPQVTNGDTAKGHPSGLVRTQSEEPQSLQQP
ATSTTETPASPAQPTQQTPTNGRRRAANEDPDEKRRKFLERNRAAASRCRQKRKVWVQ
SLEKKAEDLSSLNGQLQNEVTLLRNEVAQLKQLLLAHKDCPVTAMQKKSQYHTADKDDSS
EDISVPSSPHTEAIQHSSVSTSNVGSSTSKAEAVATSVLTQLADQSSEPLPQVGVVPPS
QAQPSGS
```

Après le scan Prosite à partir de la séquence au format FASTA, on obtient le résultat suivant :

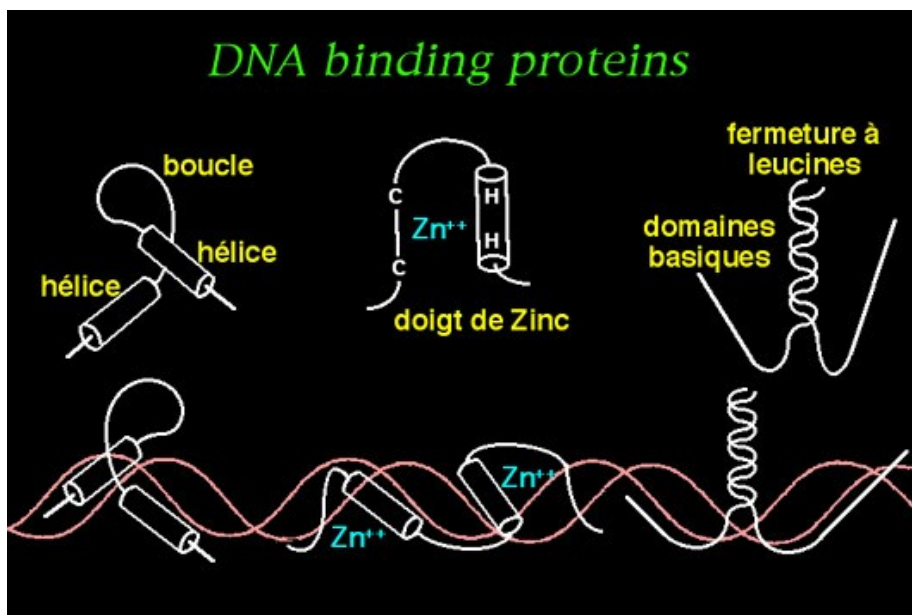


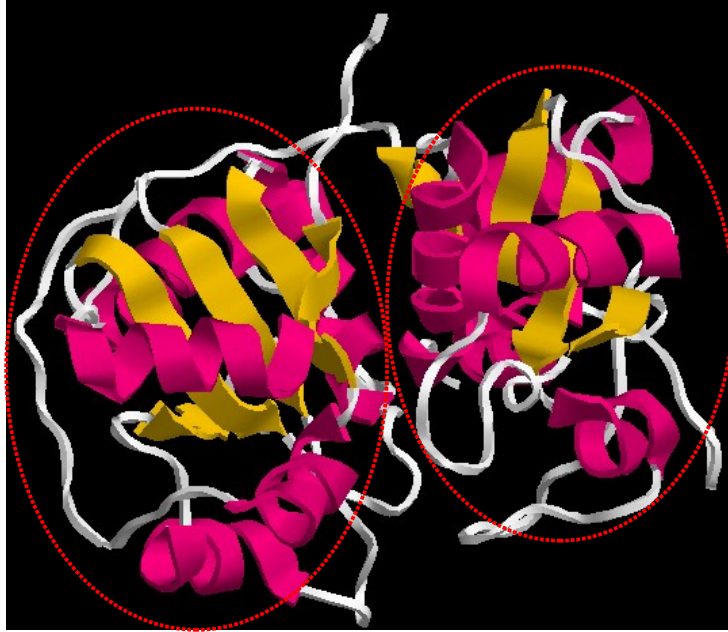
Le facteur de transcription 2 du poulet (numéro d'accèsion Prosite : PS50217) contient un motif en doigt de zinc et un motif Leucine Zipper basique, tous les deux impliqués dans la liaison du facteur de transcription 2 du poulet à l'ADN.

Tableau récapitulatif des motifs protéiques permettant la liaison à l'ADN

Nature / Nom	Description	Propriété fonctionnelle
Motif en doigt de zinc	 <ul style="list-style-type: none"> - Composé de 25 à 30 résidus d'acides aminés. - Il y a 2 résidus de cystéines à une extrémité du motif, et 2 résidus d'histidines à l'autre extrémité (doigt de zinc de type C2H2). - Les 2 cystéines et les 2 histidines sont impliquées dans des liaisons de coordination avec le zinc. 	<ul style="list-style-type: none"> - Présence dans de nombreuses protéines se liant à l'ADN. - Interaction avec environ 5 nucléotides.
Motif en Hélice-Boucle-Hélice	<ul style="list-style-type: none"> - Constitué d'une première hélice α, d'une boucle et d'une seconde hélice α. - Composé d'environ 50-60 résidus d'acides aminés. - Localisé à l'extrémité N-Terminale de la protéine. 	Impliqué dans la liaison de la protéine au niveau du grand sillon de l'ADN.
Motif « zipper » basique à leucines	<ul style="list-style-type: none"> - Contient une région basique suivie par une « fermeture éclair » composée de résidus de leucine. - Dimérisation grâce au zipper de leucines. 	Liaison avec des séquences palindromiques spécifiques composées de 6 nucléotides.

Les trois motifs protéiques de liaison à l'ADN



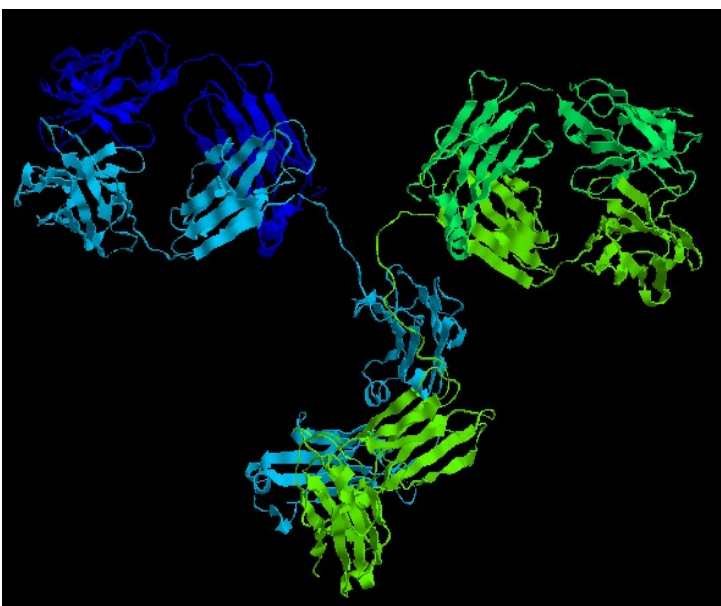
EXERCICE N°3 : Étude de la structure tridimensionnelle de la rhodanase**Structure tridimensionnelle de la rhodanase de foie de boeuf**

Numéro d'accèsion RCBS-PDB de la rhodanase de foie de bœuf : 1RHD.

La rhodanase du foie de bœuf est une protéine globulaire constituée d'une seule chaîne polypeptidique.

Elle est composée de 10 hélices α (en rose) et de 2 feuillets β parallèles formés de 5 brins β chacun (en jaune), organisés en deux domaines protéiques identiques (cerclés en rouge).

Chaque domaine de la rhodanase comprend 5 hélices α et 1 feuillet β parallèle (constitué de 5 brins β).

EXERCICE N°4 : Étude de la structure 3D d'une immunoglobuline (= anticorps) anti-lysozyme**Structure quaternaire d'une immunoglobuline G (IgG)****Q1.**

L'IgG est une protéine tétramérique composée de :

- deux chaînes lourdes (en bleu clair et en vert clair), appelées chaînes H (H pour *Heavy*).
- deux chaînes légères (en bleu foncé et en vert foncé), appelées chaînes L (L pour *Light*).

La séquence des deux chaînes H (chaînes B et D) est identique et contient 444 résidus d'acides aminés :

```

EVKLQESGGG LVQPGGSLKL SCATSGFTFS DYYMYWVRQT PEKRLEWVAY
ISNGGGSTYY PDTVKGRFTI SRDNAKNTLY LQMSRLKSED TAMYYCARHG
GYYAMDYWGQ GTTVTVSSAK TTAPSVYPLA PVCDDTTGSS VTLGCLVKGY
FPEPVTLTWN SGSLSSGVHT FPAVLQSDLY TLSSSVTVTS STWPSQSITC
NVAHPASSTK VDKKIEPRGP TIKPCPPCKC PAPNLLGGPS VFIFPPKIKD
VLMISLSPIV TCVVVDVSED DPDVQISWFV NNVEVHTAQT QTHREDYNST
LRVVSALPIQ HQDWMSGKEF KCKVNNKDLF APIERTISKP KGSVRAPQVY
VLPPPEEEMT KKQVTLTCMV TDFMPEDIYV EWTNNGKTEL NYKNTEPVLD
SDGSYFMYSK LRVEKKNWVE RNSYSCSVVH EGLHNNHHTTK SFSR

```

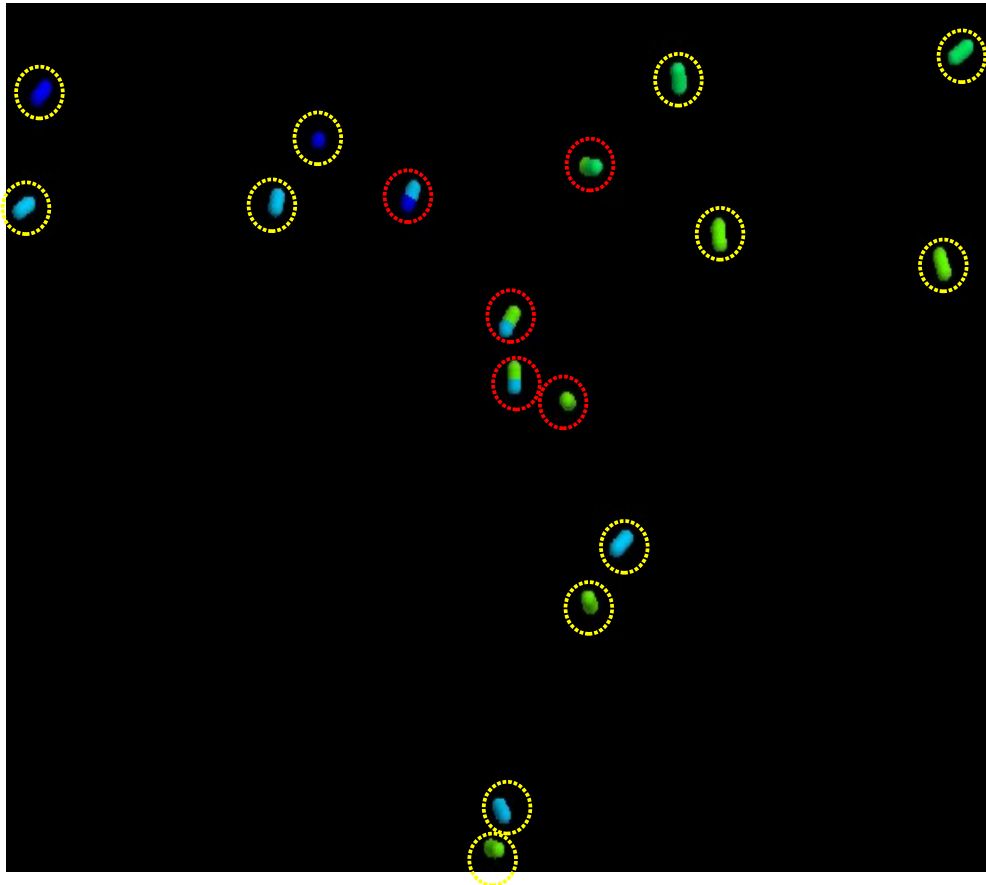
La séquence des deux chaînes L (chaînes A et C) est identique et contient 214 résidus d'acides aminés :

```

DIVLTQSPSS LSASLGDTIT ITCHASQNIN VWLSWYQQKP GNIPKLLIYK
ASNLTGVPS RFSGSGSGTG FTLTISSLQP EDIATYYCQQ GQSYPLTFGG
GTKLEIKRAD AAPTVSIFPP SSEQLTSGGA SVVCFLNNFY PKDINVKWKI
DGSERQNGVL NSWTDQDSKD STYSMSSTLT LTKDEYERHN SYTCEATHKT
STSPIVKSFN RNEC

```

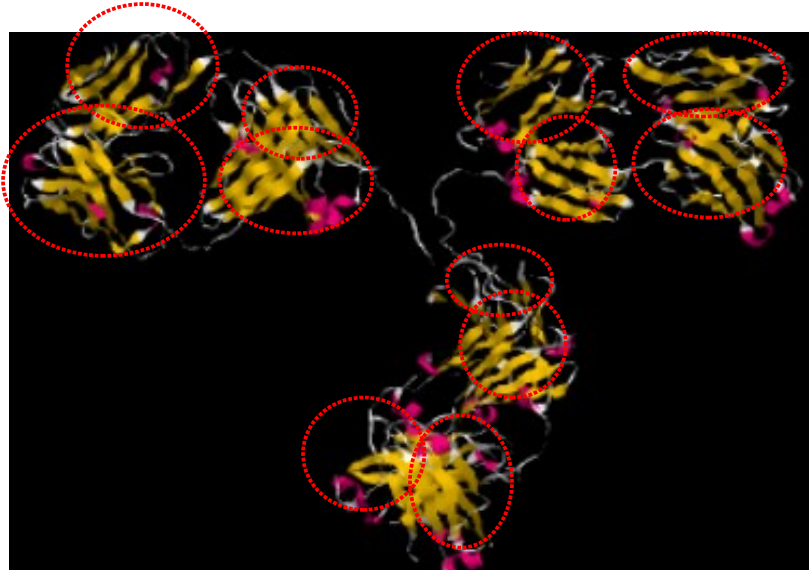
Localisation des ponts disulfures dans une IgG



On observe :

- des ponts disulfures intrachaînes (cerclés en **jaune**) :
 - 4 par chaîne lourde H.
 - 2 par chaîne légère L.
- des ponts disulfures interchaînes (cerclés en **rouge**) :
 - 3 entre les deux chaînes lourdes H.
 - 1 entre chaque chaîne H et chaîne L.

Q2.

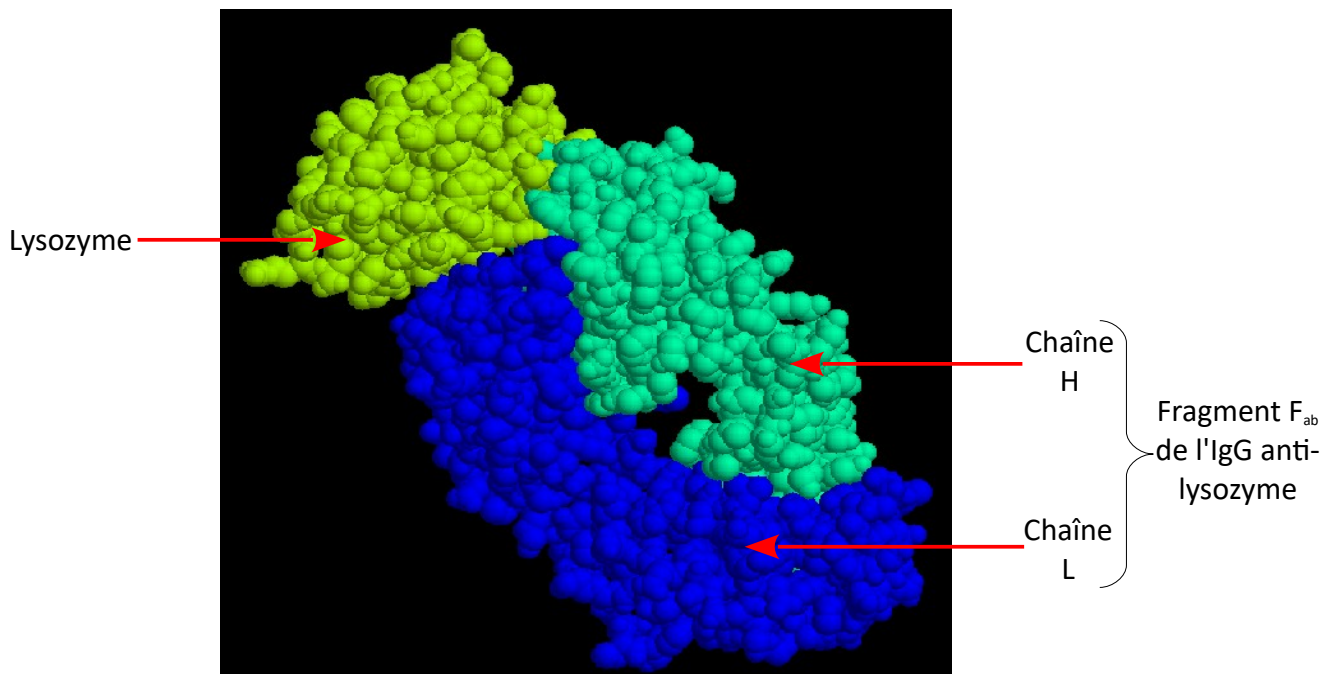
Domaines présents dans une IgG

On observe plusieurs domaines au sein d'une IgG (cerclés en rouge):

- 2 domaines pour chaque chaîne L.
- 4 domaines pour chaque chaîne H.

Ces domaines sont semblables et sont constitués de 7 brins β organisés en 2 feuillets β antiparallèles. Ces domaines sont appelés domaines « immunoglobuline ».

Q3.

Structure du fragment F_{ab} d'une IgG anti-lysozyme associé au lysozyme

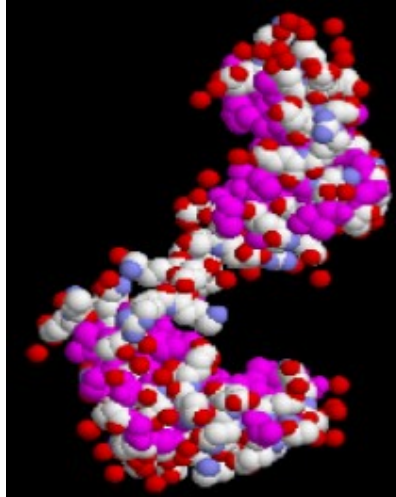
La reconnaissance et la fixation du lysozyme par l'IgG anti-lysozyme s'explique par la complémentarité de conformation (stéréospécificité) entre le fragment F_{ab} de l'IgG (paratope) et une zone du lysozyme (épitope).

EXERCICE N°5 : Différents types de protéines

1. Étude d'une protéine globulaire : la calmoduline

Q1.

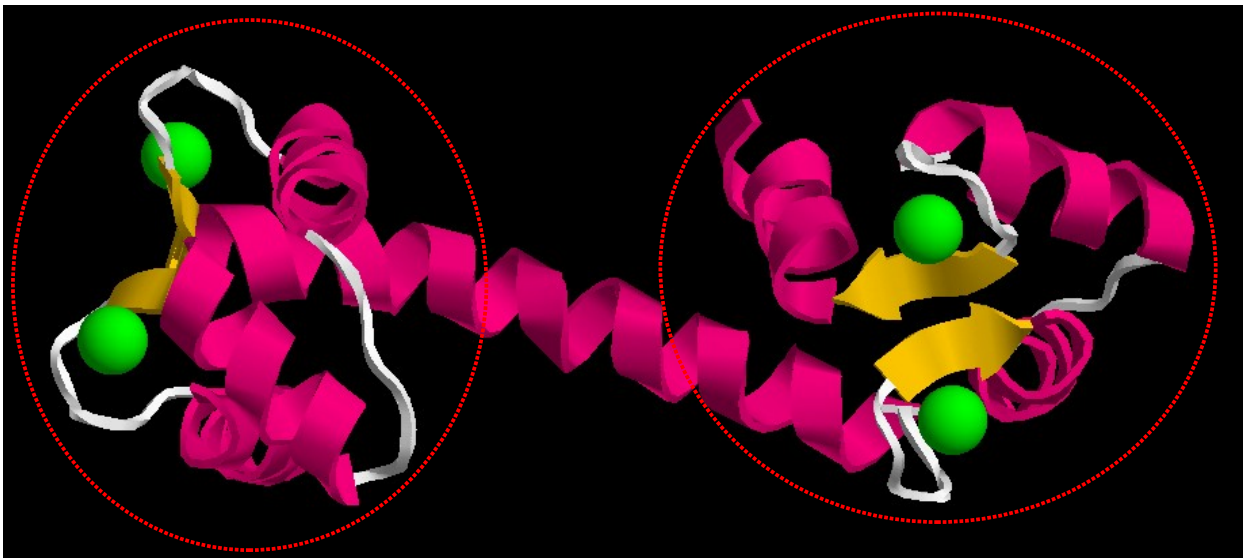
Structure tridimensionnelle de la calmoduline et localisation de résidus hydrophobes



La calmoduline est une protéine monomérique (une seule chaîne polypeptidique) globulaire. Les résidus d'acides aminés hydrophobes (en **rose**) sont enfouis à l'intérieur des deux « poches » (grâce aux interactions hydrophobes), pour ne pas être en contact avec l'environnement aqueux.

Q2.

Structure tridimensionnelle de la calmoduline liée au calcium



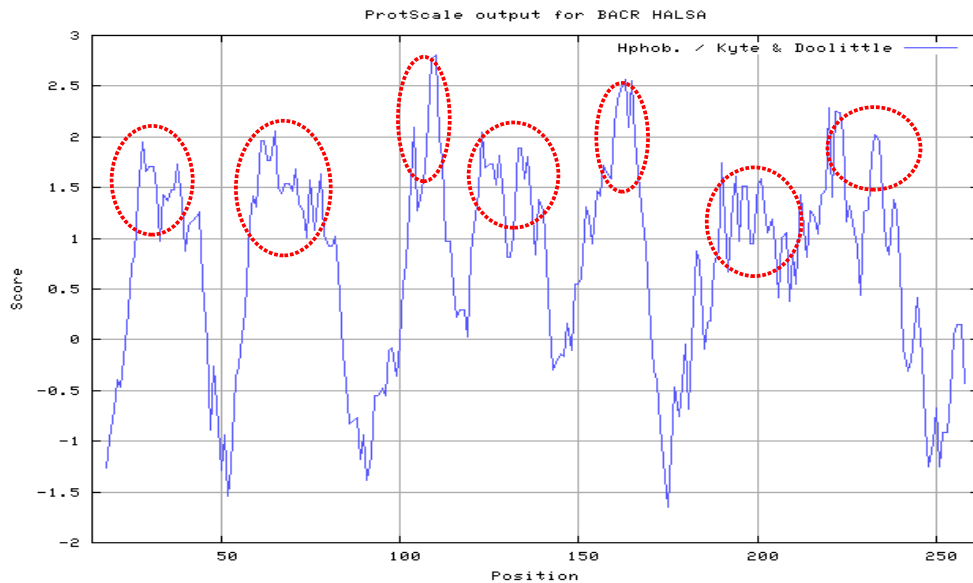
La calmoduline est une protéine monomérique (1 chaîne) composée de 7 hélices α (en **rose**) et de 2 feuillets β antiparallèles composés de 2 brins β chacun (en **jaune**).

Q3. La calmoduline peut fixer 4 atomes de calcium (sphères vertes) grâce à deux domaines identiques (cerclés en **rouge**) composés de deux motifs de liaison au calcium (hélice α - brin β - hélice α).

2. Étude de protéines transmembranaires

Q4. La bactériorhodopsine d'*Halobacterium salinarum* est une protéine transmembranaire.

Q5. Le profil d'hydrophobicité de la bactériorhodopsine d'*Halobacterium salinarum* est présenté ci-dessous :



Q6. D'après le profil d'hydrophobicité obtenu, on distingue 7 grandes régions hydrophobes (cercées en rouge).

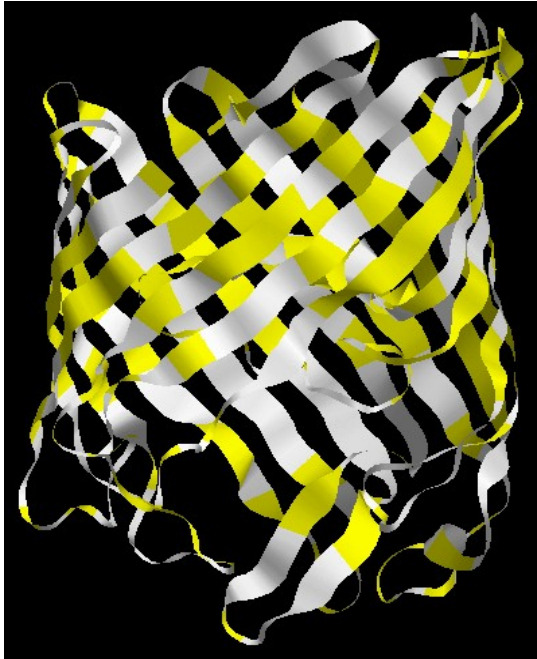
Ces régions correspondent aux structures en hélice α données par l'outil de prédiction PsiPred.

Structure tridimensionnelle de la bactériorhodopsine d'*Halobacterium salinarum*

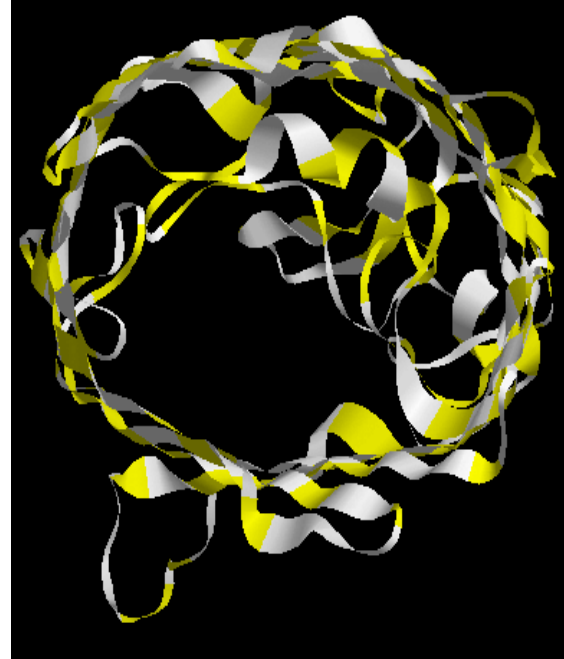


Q7. La structure de la bactériorhodopsine d'*Halobacterium salinarum* montre 7 hélices α constituées de résidus d'acides aminés hydrophobes (en jaune). Cette structure confirme les résultats obtenus avec le profil d'hydrophobicité ProtScale et l'outil de prédiction PsiPred. Les 7 hélices α traversent la membrane.

Q8.

Structure tridimensionnelle de la porine OmpF d'*Escherichia coli*

vue latérale



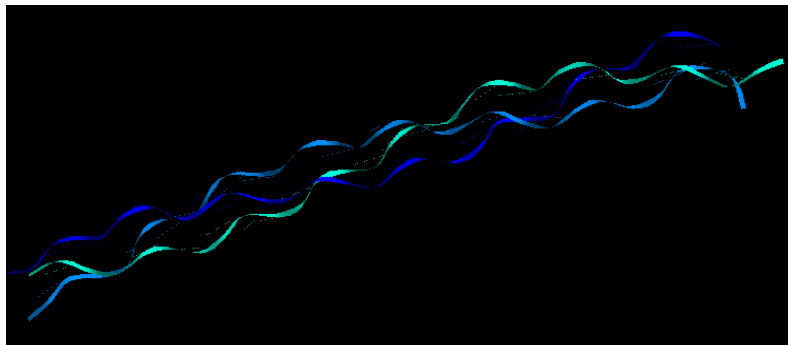
Vue polaire

La porine OmpF d'*Escherichia coli* est composée de 21 brins β , eux-mêmes constitués de résidus d'acides aminés hydrophobes (en **jaune**) permettant à cette protéine de s'insérer dans la membrane cellulaire.

La vue polaire indique que la porine, en s'insérant dans la membrane cellulaire, crée un pore permettant le passage de certaines substances.

3. Étude d'une protéine fibreuse : le collagène

Q9.

Structure tridimensionnelle d'une protéine fibreuse : le collagène humain

Le collagène humain est composé de 3 hélices (ATTENTION : ce ne sont pas des hélices α). La structure en triple hélice est stabilisée par la présence de liaison hydrogène (en pointillés).

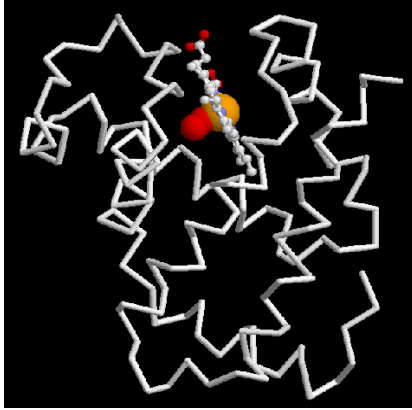
Q10. En rompant la liaisons peptidique avant chaque résidu glycine d'un motif de séquence X-Pro-Gly-Y, l'enzyme de *Clostridium perfringens* destructure le collagène. Le tissu est alors fragilisé et facilite l'envahissement de cette bactérie (pouvoir invasif).

EXERCICE N°6 : Relation structure-fonction d'une globine

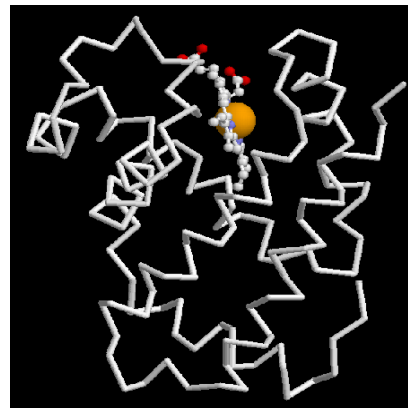
L'hémoglobine est une hétéroprotéine tétramérique (4 sous-unités protéiques), présente dans les hématies, et intervenant dans le transport du dioxygène. Elle est formée de 4 globines de nature protéique (2 α -globines et 2 β -globines) contenant chacune un groupement prosthétique (= non protéique) appelé hème, au niveau duquel se fixe le dioxygène.

1. Étude du positionnement de l'hème

Q1. Structure de la β -globine de l'hémoglobine en présence ou en absence de O_2



En présence de O_2



En absence de O_2

Le dioxygène (en rouge) se fixe au niveau de l'atome de fer (en orange) de l'hème.

La fixation de O_2 entraîne un changement du positionnement de l'hème au sein de la β -globine de l'hémoglobine.

2. Repérage de la structure impliquée dans la fixation de l'hème

L'hème est situé dans une « poche » au sein de la β -globine de l'hémoglobine.

En étudiant la structure impliquée dans la fixation de l'hème au sein de la β -globine de l'hémoglobine on remarque notamment la présence de deux résidus d'histidine, l'His63 et l'His92, qui sont très conservés dans toutes les globines connues.

3. Étude de l'évolution des globines

Les résultats d'alignement des séquences protéiques des différentes β -globines des hémoglobines de différentes espèces sont présentés ci-après :

Alignement multiple de séquences de la β -globine de l'hémoglobine de différentes espèces

```

T-COFFEE, Version_11.00.8cbe486 (2014-08-12 22:05:29 - Revision 8cbe486 - Build 47
Cedric Notredame
SCORE=982
*
  BAD  AVG  GOOD
*
sp|P02024|HBB_G   : 97
sp|P68871|HBB_H   : 97
sp|P02062|HBB_H   : 98
sp|P02140|HBB_C   : 97
sp|P02131|HBB_C   : 97
cons              : 98

sp|P02024|HBB_G   MVHLTPEEKSAVTALWGKVNVDVVGGEALGRLLVYYPWTQRFVESFGDLSTPDAVMGNPKVKA
sp|P68871|HBB_H   MVHLTPEEKSAVTALWGKVNVDVVGGEALGRLLVYYPWTQRFVESFGDLSTPDAVMGNPKVKA
sp|P02062|HBB_H   -VQLSGEEKAAVLALWDKVNVEEVGGEALGRLLVYYPWTQRFVDSFGDLSNPGAVMGNPKVKA
sp|P02140|HBB_C   -VEWTDAERSAIIIGLWGKLNDELGPQALARCLIVYPWTQRYFATFGNLSPPAAIMGNPKVAA
sp|P02131|HBB_C   -SPFSAHEEKLIVDLWAKVDVASC GG DALSRMLIIYPWKRRYFEHFGKLS TDQDVLHNEKIRE

cons              : * . : ** * : . * . * . * . * . * . * . * . * . * . * . * . * .

sp|P02024|HBB_G   HGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLHVDPENFKLLGNVLCVLAH HFG - KEFTP
sp|P68871|HBB_H   HGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLHVDPENFRLLGNVLCVLAH HFG - KEFTP
sp|P02062|HBB_H   HGKKVLHSFGEGVHHL DNLKGT FAALSELHCDKLHVDPENFRLLGNVLCVLAH HFG - KDFTP
sp|P02140|HBB_C   HGRTVMGGLERAIKNMDNIKATYAPLSVMHSEKLHVD PNFRL LADCITVCAAMKFGPSGFNA
sp|P02131|HBB_C   HGKKVLASFG EAVKHL DNIKGF AHL SKLHF EKFH VDCENFKLLGDIIIVVLGMHHP - KDFTL

cons              : * . * . : . : : * * : * . * * . * . * . * . * . * . * . * .

sp|P02024|HBB_G   PVOAAYQKVVAGVANALAHKYH
sp|P68871|HBB_H   PVOAAYQKVVAGVANALAHKYH
sp|P02062|HBB_H   ELQASYQKVVAGVANALAHKYH
sp|P02140|HBB_C   DVQEA WQKFLS VVVSALCRQYH
sp|P02131|HBB_C   QTHAAFQKLV RHVAAALSAEYH

cons              : : * * : * . * * . * *

```

- Légende : - P02024 : séquence de la β -globine de l'hémoglobine chez le gorille.
 - P68871 : séquence de la β -globine de l'hémoglobine chez l'être humain.
 - P02062 : séquence de la β -globine de l'hémoglobine chez le cheval.
 - P02140 : séquence de la β -globine de l'hémoglobine chez le poisson rouge.
 - P02131 : séquence de la β -globine de l'hémoglobine chez le caïman.

Q3. Toutes les β -globines comparées ont une longueur presque identique.

De plus, en alignant les séquences, on observe que la composition en résidus d'acides aminés est très semblable (surlignés en rose, et avec un *), surtout dans les zones de fixation de l'hème.

La β -globine est donc très conservée entre ces espèces.

EXERCICE N°7 : Carboxypeptidase et ajustement induit

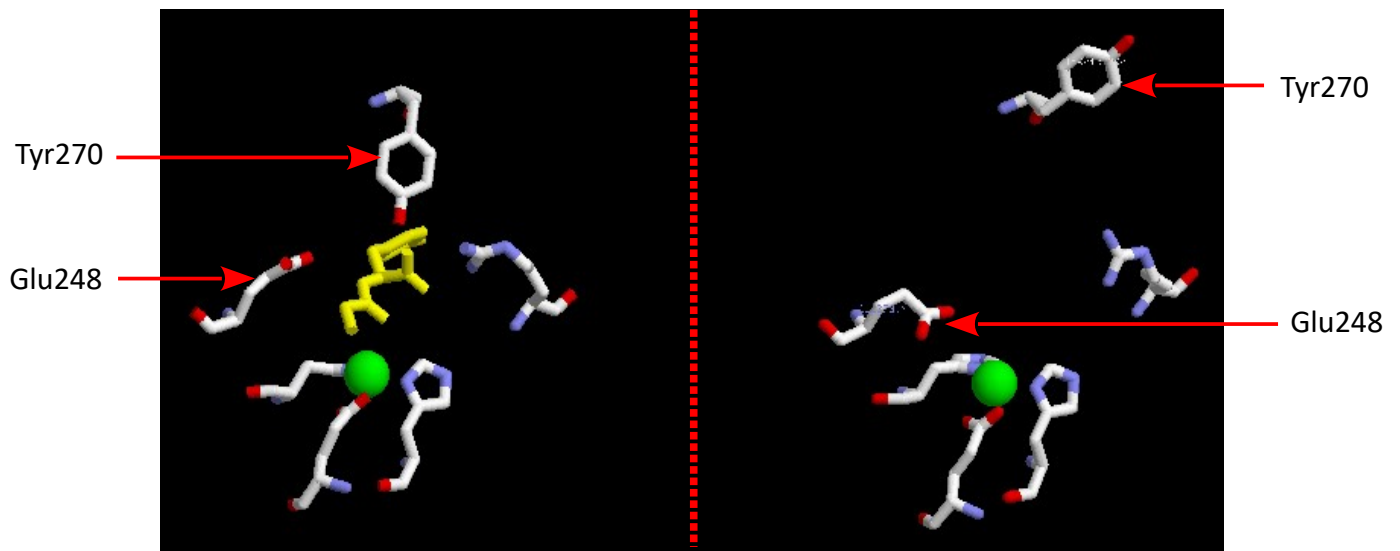
On cherche à visualiser le modèle de l'ajustement induit et ses conséquences sur les structures moléculaires lors de la liaison entre une enzyme, la carboxypeptidase, et son substrat.

Les acides aminés présents dans le site catalytique sont His69, Glu72, Arg145, His196, Tyr248 et Glu270.

Conformation du site catalytique de la carboxypeptidase en présence et en absence de son substrat

En présence du substrat :

En absence de substrat :



Légende : - en vert : atome de zinc.
- en jaune : substrat de la carboxypeptidase.

La position des résidus Glu248 et Tyr270 du site catalytique de la carboxypeptidase change selon la présence ou non de son substrat.
Cela montre que la conformation tridimensionnelle du site catalytique de la carboxypeptidase s'ajuste à la conformation du substrat.
Cela correspond au modèle de l'ajustement induit entre le site actif d'une enzyme et son substrat.