

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE RECHERCHE
SCIENTIFIQUE

UNIVERSITE Dr. MOULAY TAHAR.SAIDA



FACULTE DES SCIENCES
DEPARTEMENT DE BIOLOGIE

Mémoire

Présenté en vue de l'obtention du diplôme de Master

En : Biologie

Spécialité : biochimie et physiologie cellulaire

THEME

Motifs Structurels et Fonctionnels qui se lient les Glucides : une étude dans la compréhension des bases fondamentales derrière la fonction biologique des protéines.

URL: www.bioinformaticstools.org/prjs/glsfms/

Présenté par :

M^r Zéin Mehdi Mohamed Cheikh

Devant la commission de jury, composée par :

Président
Examineur
Examineur
Encadreur

Dr. Hamani houari
Prof. Benali Omar
Dr. Adli Djallal Eddine Houari
Dr. RACHEDI Abdelkrim

Année universitaire
2014/2015

Sommaire :

Résumé	
Liste des figures	
Liste des tableaux	
Liste des abréviations	
Introduction Générale	01

Première partie : Etude Bibliographique**Chapitre I : Recherche Bibliographique**

A. les protéines	03
I. Généralité sur les protéines.....	03
II. Structure des protéines	03
II.1.Structure primaire	04
II.2.Structure secondaire	04
II.2.1.L'hélice α	05
II.2.2.Brin et feuillet β	06
II.2.3.Coudes et boucles.....	06
II.3.Structure tertiaire	06
II.4.Structure quaternaire	07
III. Les motifs protéiques	08
IV .Les méthodes expérimentales de détermination des structures protéiques	09
IV.1 La cristallographie rayon X	09
IV.2 La spectroscopie RMN.....	10
IV.3 Cryo-microscopie électronique (cryo-EM)	10
B. Les Glucides	11
C. Les glycoprotéines	11

D. La banque de données sur les protéines (PDB)	12
1. Qu'es que la PDB	12
2. Le world wide PDB (wwPDB)	12
2.1. Membres de wwPDB	12
2.1.1. Le PDBe	13
2.1.2. Le PDBj	13
2.1.3. RCSB PDB (les États-Unis)	13
3 .Formats des fichiers	13
3.1. Le fichiers PDB	14
3.2. Les champs dans les fichiers PDB.....	15
3.3. Le champ « ATOM ».....	16
3.4. Limites du format PDB.....	18
4. Le format mmCIF.....	19
5. Les autres banques de structures protéiques	20
5.1. SCOP (Structural Classification of Proteins).....	20
5.2. CATH (Class Architecture Topology and Homology)	20

Deuxième partie : étude expérimentale

Chapitre II : Préparation des Données

1. Les structures protéiques (codes PDB)	22
2. List des Ligand	31
3. Les images des ligands.....	36

Chapitre III : matériels et méthodes

1. Sélection de la liste des codes PDB et la liste des ligands.....	48
2. Collection (Datamining) des détails de calcule de l'environnement de liaison des ligands.....	49
3. Création de la base de données	49
I. Plateforme Développement (local)	49
I.1 Langage de programmation :	49

I.1.1. PHP (Hypertext Preprocessor).....	49
I.1.2. Structured Query Language – SQL.....	49
I.1.3.HTML et CSS.....	50
I.2. Bases de données MySQL.....	50
I.3. Apache serveur.....	50
II Création de la base de données « GlcSFMs»	50
II.1.Les scripts PHP et SQL utilisée dans ce projet	50
II.1.1. Les scripts de la création de la base de données.....	50
II.1.1.i Le premier script est « DB_connection.php »	51
II.1.1.ii Le script de création « createDB.inc.php».....	51
II.1.2.Les scripts de l’insertion des données dans la base de données ..	51
II.1.2.i le script « insert_pdb_list.php »	51
II.1.2.ii le script « insert_ligands_list.php »	52
II.1.2.iii Le script « insert_details.php ».....	53
II.1.2.iv script « gnr_rsw_n_script .php»	53
II.1.5.v Le script « insertion_des_images.php»	53
II.1.3.Les scripts de l’affichage des données local sur un navigateur	54
II.1.3.i Le script « menu.php ».....	54
II.1.3.ii Le script « query.php »	54
II.2. La création des fichiers et les tables	55
II.2.1. Création des fichiers	55
II.2.1.1. Le fichier « LISTEPDB.txt »	55
II.2.1.2. Le fichier « LISTELIGANDS.txt».....	55
II.2.1.3. Le fichier RasWin	56
II.2.2. Création des tableaux.....	56

II.2.2.1. Structure de tableau «file_image».....	57
II.2.2.2 .Structure de tableau «pdb_data».....	57
II.2.2.3. Structure de tableau «ligands».....	58
II.2.2.4. Structure du tableau «prot_lig_contact».....	59
II.2.2.5. Le lien entre les tableaux.....	60
II.3 voie Générale sur le Processus de la création de la base de données GlcSFMs	61
III. Plateforme Production (web)	62
Chapitre VI : Résultat et discussion	
1. Présentation des résultats	
1.1. Présentation des résultats sur le serveur locale	63
1.2. Cas de sélection par PDB	64
1.3. Cas de sélectionne par Ligand.....	65
2. Discussion et interprétation des résultats	66
Conclusion Générale	81
Annexe 1.....	83
Annexe 2.....	94
Référence Bibliographique	

Summary:

This project was carried out to try and investigate some of the fundamental rules behind the structure-function relationship in proteins.

The database Protein Data Bank was used as the primary source of protein structures in complex with carbohydrate or glucide ligands. In this work, a population of a 100 protein structures have been selected for the investigation.

The study resulted in identifying, defining and characterizing a number of structural elements deemed by the study to play major contributions in binding the Glucides ligands which are important in helping the carrying out of the biological function exhibited by the selected protein structures.

In order to share the data and results with the scientific community both local and international, an online and searchable database named Glucides Binding Structural and Functional Motifs, GLcSFMs, has been created and made available at the web address: <http://www.bioinformaticstools.org/prjs/glsfms/>

Keywords: Protein,, Glucides, Ligands, Structural Motifs, Sequence, PDB, Database.

الملخص:

تم القيام بمشروع البحث هذا في محاولة لاستكشاف بعض من الأساسيات التي تحكم العلاقة بين "التركيب الفراغي والوظيفة البيولوجية" لدى البروتينات.

للقيام بالمشروع كان لزاما استخدام قاعدة بيانات التراكيب الفراغية للبروتينات أو Protein Data Bank للحصول على التركيب الفراغي لمجموعة من البروتينات وهي في حالة ارتباط مع ليجندات (ربيطات) سكرية الطبيعة حيث تم استخدام مجموعة من 100 تركيب فراغي بروتيني مرتبطة بمجموعة من السكريات.

نتائج البحث تمخضت عن الكشف، التحديد، والتوصيف لمجموعة من العناصر التركيبية التي اعتبرت في هذه الدراسة على أنها المسؤول الأساسي في عملية التعرف والارتباط بالليجندات السكرية والتي تلعب دورا حاسما في القيام بالوظيفة البيولوجية لدى البروتينات المختارة في هذا المشروع.

وحتى تتم الاستفادة من بيانات ونتائج هذا البحث ومشاركته مع البعثة محليا ودوليا تم إنشاء قاعدة بيانات سميت بـ قاعدة بيانات النماذج التركيبية والوظيفية الرابطة للسكريات في البروتينات وهي قابلة للبحث والتصفح على الشبكة الدولية للإنترنت على العنوان التالي: <http://bioinformaticstools.org/prjs/glsfms/>

الكلمات الرئيسية:

، قواعد البيانات.PDB، بروتين، سكريات، الليجندات أو الربيطات، الأنماط الهيكلية، سلسلة،

Résumé:

Ce projet a été réalisé pour essayer et d'enquêter sur certaines des règles fondamentales derrière la relation structure-fonction des protéines.

La base de données, Protein Data Bank, a été utilisée comme source primaire de structures de protéines qui sont dans un complexe forme avec des ligands Glucidiques. Dans ce travail, une population de 100 structures de protéines ont été sélectionnés pour la réalisation de l'enquête.

L'étude a abouti à l'identification, la définition et la caractérisation d'un certain nombre d'éléments structurels jugés par l'étude à jouer des contributions majeures dans la liaison des ligands Glucidiques qui sont importants dans la réalisation de la fonction biologique présenté par les protéines sélectionnées.

Afin de partager les données et les résultats avec la communauté scientifique locales et internationales, une base de données en ligne et consultable nommé Glucides Binding Structural and Functional Motifs, GLcSFMs, a été créé et mis à disposition à l'adresse web: <http://www.bioinformaticstools.org/prjs/glsfms/>

Les mots clés: Protéine, Glucides, Ligands, Motifs Structurel, Séquence, PDB, Base de données.

Liste des figures

Figure 01 : Image présente Les niveaux des Structures des protéines (primaire, Secondaire, tertiaire, et quaternaire).

Figure 02 : structure primaire des protéines

Figure 03 : structure secondaire des protéines

Figure 04 : structure de l'hélice α

Figure 05 : structure de feuillet et brins β

Figure 06 : La structure tertiaire des protéines.

Figure 07: La structure quaternaire de la protéine.

Figure 08 : Doigt de zinc

Figure 09 : le motif hélice-Toure-hélice

Figure 10 : structure de glycoprotéines

Figure 11 : Extrait des données brutes du début d'un fichier PDB. Il est visible que chaque ligne contient exactement 80 colonnes, soit 80 caractères.

Figure 12 : Les 6 premières colonnes définissent le champ de la ligne.

Figure 13 : Comparaison des données brutes de fichiers PDB de molécule (1HHO)

Figure 14 : Détail du champ « HEADER ». Ici ce champ indique que la molécule est Oxygène transport, que ses coordonnées ont été déposées sur la PDB le 10 juin 1983 et que son identifiant est 1HHO.

Figure 15 : Exemple de champs « ATOM » qu'il est possible de trouver dans un fichier PDB

Figure 16 : Ici le champ « TER » marque la fin de la chaîne A.

Figure 17 : Illustration de ce que permet le logiciel de visualisation moléculaire.

Figure 18: limitations de format PDB

Figure 19 : Exemple du début d'un fichier au format mmCIF.

Figure 20 : La structure de la BGC.BETA-D-GLUCOSE

Figure 21 : La structure de la G6P. ALPHA-D-GLUCOSE-6-PHOSPHATE

Figure 22 : La structure de la NAG. N-ACETYL-D-GLUCOSAMINE

Figure 23 : La structure de la 2FP. 1,6-FRUCTOSE DIPHOSPHATE (LINEAR FORM)

Figure 24 : La structure de la FBP. BETA-FRUCTOSE-1,6-DIPHOSPHATE

Figure 25 : La structure de la 0MK. L-RIBOPYRANOSE.

Figure 26 : La structure de la 3BU. L-ALLOSE.

Figure 27 : La structure de la 5RP ,RIBULOSE-1,5-DIPHOSPHATE.

Figure 28 : La structure de la A6P,6-O-PHOSPHONO-ALPHA-DALLOPYRANOSE

Figure 29 : La structure de la ALL,D-ALLOPYRANOSE.

Figure 30 : La structure de la AOS,D-ALLOSE.

Figure 31 : La structure de la ARB, BETA-L-ARABINOSE.

Figure 32 : La structure de la CAP, 2-CARBOXYARABINITOL-1,5-DIPHOSPHATE.

Figure 33 : La structure de la F6P, FRUCTOSE-6-PHOSPHATE.

Figure 34 : La structure de la GAL, BETA-D-GALACTOSE.

Figure 35 : La structure de la IUG, 1-(2-NITROPHENYL)ETHYL UDP-GALACTOSE

Figure 36 : La structure de la LAT, BETA-LACTOSE.

Figure 37 : La structure de la LBT, ALPHA-LACTOSE.

Figure 38 : La structure de la MAN, ALPHA-D-MANNOSE.

Figure 39 : La structure de la MBG, METHYL-BETA-GALACTOSE.

Figure 40 : La structure de la MMA, O1-METHYL-MANNOSE.

Figure 41 : La structure de la MDM, METHYL-O3-(ALPHA-D-MANNOSE)-ALPHA-D-MANNOSE.

Figure 42 : La structure de la NA1, METHYL N-ACETYL ALLOSAMINE.

Figure 43 : La structure de la NGA, BN-ACETYL-D-GALACTOSAMINEE.

Figure 44 : La structure de la PAN, 5-PHOSPHO-D-ARABINOHYDROXAMIC ACID.

Figure 45 : La structure de la R5P, RIBOSE-5-PHOSPHATE

Figure 46: La structure de la RIP, RIBOSE (PYRANOSE FORM)

- Figure 47 : La structure de la RUB, RIBULOSE-1,5-DIPHOSPHATE
- Figure 48 : La structure de la TDG, THIODIGALACTOSIDE.
- Figure 49 : La structure de la XLS, D-XYLOSE (LINEAR FORM).
- Figure 50 : La structure de la XUL, D-XYLULOSE.
- Figure 51 : La structure de la XYL, D-XYLITOL.
- Figure 52: La structure de la XY5, XYLOPYRANOSE.
- Figure 53 : La structure de la ARA, ALPHA-L-ARABINOSE.
- Figure 54 : capture d'écran représente le script « DB_connection.php »
- Figure 55 : capture d'écran de script « createDB.inc .php»
- Figure 56 : capture d'écran de script « insert_pdb_list.php
- Figure 57 : capture d'écran représente le script « insert_ligands_list.php »
- Figure 58 : capture d'écran représentatif de script « fetch_insert_lgct_dtls.php
- Figure 59 : capture d'écran représente les détails de script « gnr_rswn_script .php»
- Figure 60 : capture d'écran représente le script « insertion_des_images.php»
- Figure 61 : capture d'écran représente le script « menu.php »
- Figure 62 : capture d'écran représente le script « query.php »
- Figure 63 : capture d'écran représentatif du fichier « LISTEPDB.txt »
- Figure 64 : capture d'écran représentatif du fichier « LISTELIGANDS.txt »
- Figure 65 : Capture d'écran représentatif des tableaux de la base de données sous phpMyAdmi
- Figure 66 : Capture d'écran représentatif de la structure de tableau file_imege sous phpMyAdmin
- Figure 67 : Capture d'écran représentatif de la structure de tableau pdb_data sous phpMyAdmin
- Figure 68 : Capture d'écran représentatif du tableau de ligands sous phpMyAdmin
- Figure 69 : Capture d'écran représentatif de la structure de tableau prot_lig_contacts sous phpMyAdmin
- Figure 70 : Capture d'écran représentatif le lien entre les tableaux

Figure 71 : Schéma représentatif des processus de la création et exploitation de la base de données

Figure 72 : capture d'écrans représente la page menu.

Figure 73 : capture d'écran représente la page résultat après la sélection du PDB 3ACI.

Figure 74 : capture d'écran représente la page résultat après la sélection du ligand 2FP.

Figure 75 : capture d'écran représente la partie d'affichage des données sur l'entrée PDB

Figure 76 : Images 3-D des motifs présentés dans le tableau 07.

Figure 77 : Représentation graphique à secteur des taux de fréquence des motifs.

Figure 78 : La fréquence des motifs par apport aux ligands.

Figure 79 : Le taux de présence des acides aminés dans la séquence en contact avec les ligands par apport aux ligands

Figure 80 : Le taux de présence des acides aminés en contact avec les ligands par apport aux Motifs.

Figure 81 : Interactions de Van Der Waals

Figure 82 : ponts hydrogène

Figure 83 : Interactions électrostatiques

Figure 84 : Formation d'un pont disulfure.

Figure 85 : exemple d'un hexose (glucose)

Figure 86 : exemple d'un pentose (xylose)

Figure 87 : structure de saccharose

Figure 88 : image représente la structure de l'amidon

Figure 89 : structure de glycogène

Figure 90 : structure de cellulose

Figure 91 : structure de Mannotriochitobiose

Figure 92 : structure de Transferrine

Figure 93 : la liste de PDB « LISTEPDB.txt »

Figure 94 : Liste des ligands « LISTELIGANDS.txt. »

Figure 95 : Capture d'écran représentatif de XAMPP

Figure 96 : Capture d'écran représentatif de RasMol

Liste des Tableaux

Tableau 01 : Exemple de champs avec leur description.

Tableau 02 : Les différentes sections qui regroupent les différents champs

Tableau 03 : Description du champ « ATOM »

Tableau 04: Liste des structures de protéines avec des ligands liés utilisés dans l'étude. Résolution et R-factor reflètent la qualité de la structure; des valeurs plus petites structures signifient plus fiables.

Tableau 05: Liste des ligands en complexe avec les structures PDB

Tableau 06 : tableau représente le nombre de contacts des ligands à les PDB IDs ainsi que les motifs liées a chacun de ces ligand

Tableau 07 : Les principaux types de motifs en liaison avec les ligands du (tableau 06).

Tableau 08 : les acides aminés les plus abondants par rapport aux ligands.

Tableau 09 : interaction et liaisons impliquée dans la structure protéique (cheftel et al,1985),

Tableau 10 : Liste des structures de protéines avec des ligands liés utilisés dans l'étude

Tableau 11 : La liste des ligands en complexe avec les structures PDB

Liste des abréviations

BGC : BETA-D-GLUCOSE

G6P : ALPHA-D-GLUCOSE-6-PHOSPHATE

NAG : N-ACETYL-D-GLUCOSAMINE

2FP : 1,6-FRUCTOSE DIPHOSPHATE (LINEAR FORM)

FBP : BETA-FRUCTOSE-1,6-DIPHOSPHATE

F6P : FRUCTOSE-6-PHOSPHATE

MDM : METHYL-O3-(ALPHA-D-MANNOSE)-ALPHA-D-MANNOSE

MMA : O1-METHYL-MANNOSE

MAN : ALPHA-D-MANNOSE

LBT : ALPHA-LACTOSE

TDG : THIODIGALACTOSIDE

LAT : BETA-LACTOSE

R5P : RIBOSE-5-PHOSPHATE

RIP : RIBOSE(PYRANOSE FORM)

0MK : L-RIBOPYRANOSE

GAL : BETA-D-GALACTOSE

NGA : BN-ACETYL-D-GALACTOSAMINEE

IUG : 1-(2-NITROPHENYL)ETHYL UDP-GALACTOSE

ARB : BETA-L-ARABINOSE

ARA : ALPHA-L-ARABINOSE

ALL : D-ALLOPYRANOSE

AOS : D-ALLOSE

NA1 : METHYL N-ACETYL ALLOSAMINE

3BU : L-ALLOSE

A6P : 6-O-PHOSPHONO-ALPHA-D-ALLOPYRANOSE

5RP : RIBULOSE-5-PHOSPHATE

CAP : 2-CARBOXYARABINITOL-1,5-DIPHOSPHATE

RUB : RIBULOSE-1,5-DIPHOSPHATE

XLS : D-XYLOSE (LINEAR FORM)

XYS : XYLOPYRANOSE

XUL : D-XYLULOSE

XYL : D-XYLITOL

MBG : METHYL-BETA-GALACTOSE

PAN : 5-PHOSPHO-D-ARABINOHYDROXAMIC ACID

XML : extensible markup language

PDB: La banque de données des protéines.

PDBe: Protein Data Bank Europe.

PDBj : Protein Data Bank Japon.

RCSB PDB : Protein Data Bank USA.

RMN: La spectroscopie par résonance magnétique nucléaire.

Cryo-EM :Cryo-microscopie électronique

SCOP: Structural Classification of Protéines.

CATH: Class Architecture Topology and Homology.

PHP: Hypertext Preprocessor .

HTTP: hyper text transfer protocol.

SGBD : le système de gestion de bases de données.

SGBDR: le système de gestion de bases de données relationnelles.

mmCIF: macromolecular Crystallographic Information File.

Ph : Potentiel Hydrogène.

Datamining : C'est une technique utilisée pour le calcul automatique et la collecte des données.

HTML: Hypertext Markup Language.

CSS: Cascade Style Sheet.

SSFS: Séquence, Structure and Function Server; un système de bioinformatique pour l'étude structurale des macromolécules telles que les protéines et les acides nucléiques

trouvé dans la PDB et fournir des analyse orientée vers la tâche de comprendre les bases de la relation structure-fonction dans ces molécules. SSFS est accessible à l'adresse url: <http://www.bioinformaticstools.org/ssfs>

Xampp: X (cross) Apache MySQL Perl PHP.

MySQL: My Structured Query Language.

GlcSFMs: Glucides Binding Sturctural & Functional Motifs; Les motifs structurelle et fonctionnels qui se lient des Glucides.

URL ou url: Uniform Resource Locator; Localisation des ressources Uniform.

Introduction Générale

Les protéines sont des constituants très importants des cellules vivantes, dont la diversité et la complexité structurale sont les plus élevées, aussi bien d'un point de vue quantitatif (elles représentent en général plus de la moitié du poids sec des cellules), que du point de vue qualitatif, puisqu' on trouve des protéines ayant un rôle biologique principal à coté des protéines dites Structurales.

Au niveau du site actif, le processus d'interaction se fait par liaison complémentaire du ligand avec l'architecture protéique, Cette association intermoléculaire est stabilisée par plusieurs effets et des liaisons non covalentes.

Les ligands glucidiques jouent un rôle important dans diverse processus biologiques et qui font des interactions avec les protéines pour former des complexes glycoprotéiniques.

Les structures de ces complexes glycoprotéiniques sont stockes dans une base de donnée mondiale (PDB) et qui concerne les structure 3-dimensionnel des macromolécules pour but de partager les informations sur ces structures avec la communauté scientifique et internationale.

L'objectif de cet' étude est d'analyser le site de liaison entre les ligand glucidiques et des protéines disponibles dans la banque de donnée sur les protéines pour **identifier, définir, et caractériser** les différents **motifs structurelle** qui peuvent existe dans les structures protéiniques ainsi que la fréquence des acides aminée dans chaque ligand glucidique étudié.

Ce projet, ainsi, tente de projeter la lumière sur les règles fondamentales derrière la relation structure-fonction des protéines qui leur permettent d'exercer leur fonction biologique. Une telle compréhension permettrait d'agir sur de nombreuses préoccupations en biologie et en médecine comme la lutte plus efficace contre les maladies et dans la conception des générations nouvelles de médicaments parmi beaucoup d'autres applications.

Ce projet est partagé en quatre grands chapitre le premier chapitre est une recherche bibliographique pour comprendre des notions sur les protéines et leur niveaux des structures et même pour savoir des informations sur les glycoprotéines, les différentes motifs protéiniques ainsi que le PDB qui est le lieu de stockage des informations sur les structures 3D des macromolécules et en fin les banques de classification des protéines (scop) et (cath) .

Le deuxième chapitre est concerné la préparation de donnée et qui présente les données importantes utilisées dans ce projet telle la classe de chaque code PDB utilisé et le ligand en contact avec ce dernier.

Dans le chapitre de matérielles et méthodes et qui est le troisième chapitre les outils et les matérielles utilisés sont présentés.

Le quatrième chapitre résultat et discussion présente les résultats obtenus dans cette étude,

En fin le travail est terminé par une conclusion générale.

A. Les protéines

I. Généralité sur les protéines

Les protéines sont des polymères biologiques constitués à partir de 20 acides aminés reliés entre eux par des liaisons peptidiques, dont la séquence est déterminée par l'information génétique. Ces macromolécules présentent une masse moléculaire supérieure à 10^4 . Une protéine comporte au moins 100 acide aminés

Les protéines sont des biomolécules de première importance

- Par leur présence universelle dans le monde vivant .seuls les viroïdes en sont dépourvus.
- Par leurs abondances cellulaires : c'est le premier constituant après l'eau (10 fois plus que les glucides).
- Par leurs extrême diversité : elles assurent des fonctions vitales tant structurales que dynamique et elles sont le support de la spécificité des « espèces » (benseguni et chikhi.2006 ; sultanem, 2004)

II. Structure des protéines

Les acides aminés constituant les protéines, diffèrent par la nature chimique de résidus .ils se divisent en trois familles : polaire, non chargé, chargé et hydrophobe; ces acides aminés peuvent s'associer en un grand nombre de séquences différentes et former des domaines spécifiques, dont la structure tridimensionnelle détermine la fonction biologique de la protéine (sultanem,2004).

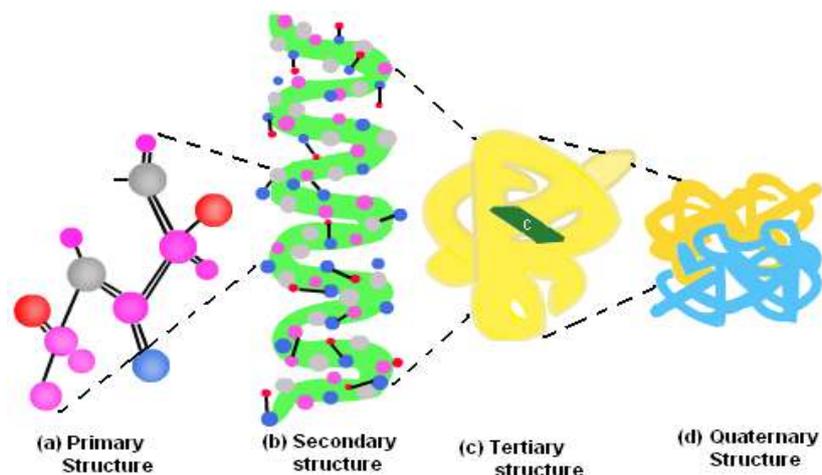


Figure 1. Image présente Les niveaux des Structures des protéines (primaire, Secondaire, tertiaire, et quaternaire).

La structure primaire ne définit que partiellement l'état de structure d'une protéine .la chaîne synthétisée par les ribosomes est linéaire ,elle subit ultérieurement une

évolution tridimensionnelle sous l'influence de forces variées (attractives ou répulsives), qui se manifestent le long de la chaîne et qui dépendent de la nature des chaînes latérales, de leur écartement et du milieu, elle aboutit à une conformation définie dont l'état thermodynamique est caractérisé par la stabilité et le minimum d'énergie libre.

La conformation de la protéine native, dans son milieu biologique, est assurée par des liaisons de faible énergie et éventuellement par le pont disulfure.

II.1. Structure primaire :

La structure primaire désigne la séquence des acides aminés le long de la chaîne principale. Elle englobe toutes les liaisons covalentes y compris les ponts disulfures. Cette structure est généralement contrôlée (figure 02) pour plus d'information voir (tableau 09 ANNEXE 1) (Frénot et Vierling, 2006 ; Alias et Lindes, 1994).



Figure 2. Structure primaire des protéines

II.2. Structure secondaire :

C'est la conformation prise par la chaîne polypeptidique stabilisée par des liaisons hydrogènes intra et extra moléculaires (tableau 09 (ANNEXE 1), et non des chaînes latérales). Et par ces liaisons, certaines régions de la séquence s'organisent localement en hélices α , en feuilles β ou coudes β ; ce sont les motifs qui caractérisent la structure secondaire de la protéine (figure 07) (Frénot et Vierling, 2006 ; Sultanem, 2004).

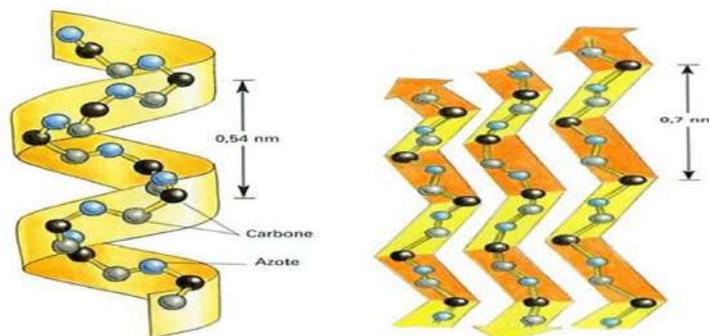


Figure 3. Structure secondaire des protéines

II.2.1. L'hélice α :

La plus part des protéines possédant des régions ont une structure en hélice α , que l'on peut imaginer couramment comme un enroulement autour d'un cylindre imaginaire, composant une spirale régulière le pas de vis est de $5,4 \text{ \AA}$, avec 3,6 restes d'acides aminés.

Or, cette structure est stable grâce à de plusieurs liaison d'hydrogènes entre les groupements CO et NH des liaisons peptidique (**Catherine, Alain**).

En réalité, l'hélice comporte 3,7 monomères d'acide amine par tour de spire.

Ainsi que, les plans des liaisons peptidiques fabriquent entre eux un angle de 80° environ, note que l'intersection entre les deux plans se retrouve toujours au niveau du carbone α .

Par contre, les chaînes latérales sont dirigées vers l'extérieur et peuvent réagie entre elles ou avec le milieu (**Signal Transduction through Prion Protéine Science, 15 septembre (2000)**).

Généralement l'hélice α est toujours une hélice droite, ainsi que 3_{10} et π sont droites on va voir ces type dans le paragraphe suivant (**P .Kaiser &al**).

. Elle s'éloigne en sinueux dans le sens des aiguilles d'une montre lorsque on observe dans l'axe de la chaîne principale (**A. MART I &al**).

Vue la nature des chaînes latérales, sont souvent distribuées autour de l'hélice, on peut soutenir une qualité hydrophobe si la maturité des chaînes sont hydrophobes ou amphiphile c'est-à-dire d'un côté Hydrophobe et d'un autre hydrophile, si les monomères d'une face sont hydrophobes et ceux de l'autre face hydrophile.

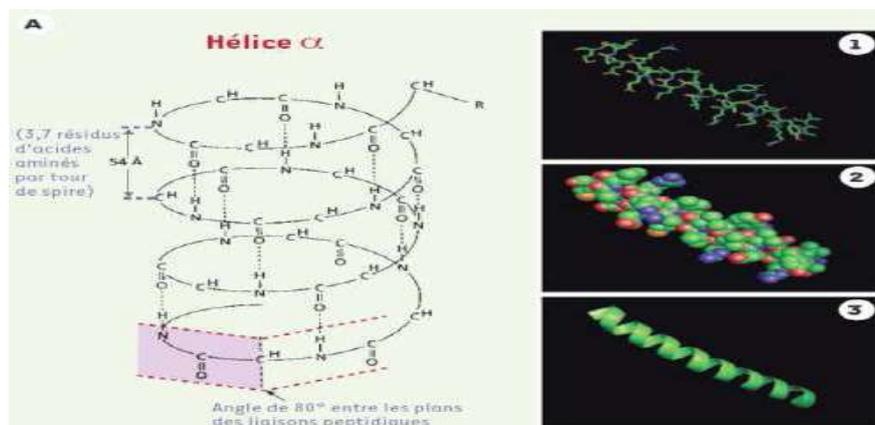


Figure 4. Structure de l'hélice α

II.2.2. Brin et feuillet β .

En effet, le feuillet β est une hélice tout à fait spécifique avec un axe de symétrie d'ordre deux peut être formée avec deux chaînes parallèles, ou deux chaînes antiparallèles.

Quoi que, dans les deux cas il ya constitution d'une structure β , plissée en forme d'accordéon, les chaînes latérales R étant périodiquement d'un côté et de l'autre de ruban.

Or, la longueur de la liaison hydrogène dans les feuillets β est en moyenne de 3 \AA , tandis que la participation de la liaison hydrogène à la stabilité de la structure secondaire est plus élevée (**DM. Koepf & al**).

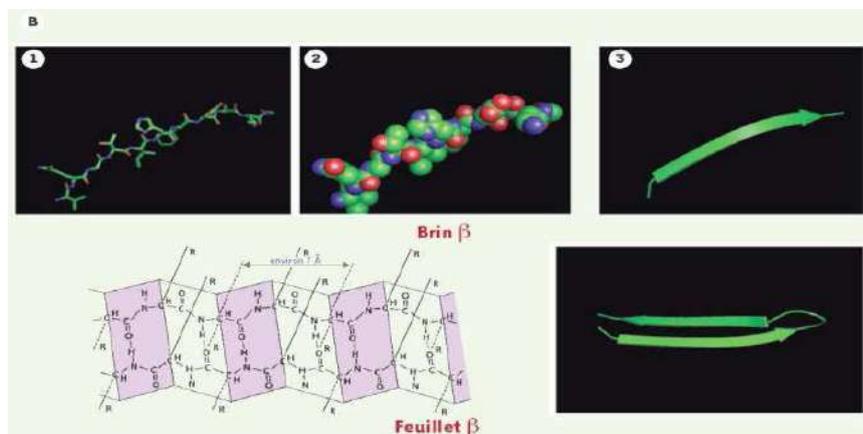


Figure 5. Structure de feuillet et brins β

II.2.3. Coudes et boucles :

Les coudes β sont des séquences de 4 acides aminés hydrophiles constituant des plis de 180° stabilisés par une liaison hydrogène, à la surface de protéines (**Catherine, Alain**).

Or, on parle généralement d'épingles à cheveux β (β hairpin) parce que les deux extrémités sont parallèles entre elles.

Alors, pour les boucles Ω peuvent tenir plusieurs coudes β et possède l'aspect de la lettre grecque majuscule.

En la retrouve compactes parce que leurs chaînes latérales ont l'aptitude à remplir l'intérieur de leurs creux (**A. MART I &al**).

II.3. Structure tertiaire :

La structure tertiaire est réellement l'état tridimensionnelle qui rassemble les formes précédentes. Elle résulte de repliement de la chaîne principale, conséquence de l'établissement de liaisons entre les résidus R, et fait intervenir des liaisons

hydrogènes électrostatiques et hydrophobe Donc elle permet le rapprochement de zones très éloignées dans la séquence .Certains acides aminés se placent au cœur de la protéine et échappent au solvant ,d'autres se disposent en surface et facilitent les interaction avec l'environnement externe (figure 06)(Frénot et Vierling, 2006).

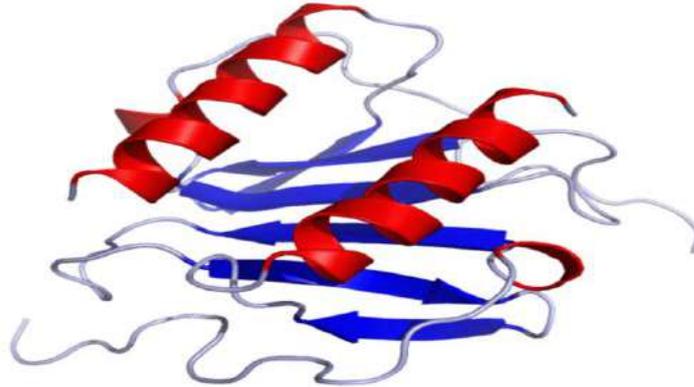


Figure 6 : La Structure tertiaire des protéines.

II.4. Structure quaternaire :

La structure quaternaire des protéines résulte d'association des sous-unités protéiques (figure 07) .les liaisons sont des liaisons hydrogènes pour les protéines globulaire et des liaisons covalentes pour les protéines fibrillaire .plusieurs organisations sont possible (exemple, forme fibreuse de l'actine, forme globulaire d'actine, de l'hémoglobine) (Frénot et Vierling,2006 ; Alias et Linden, 1994).

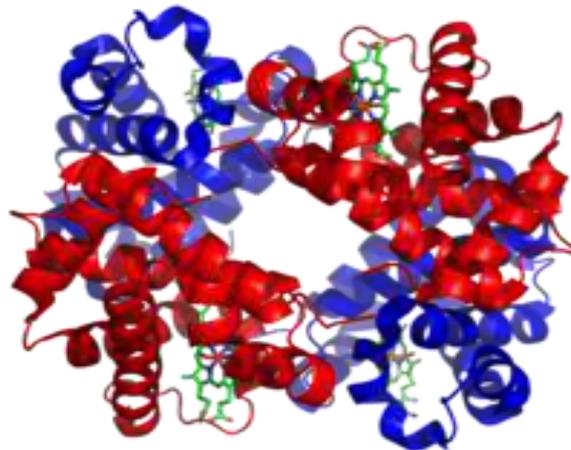


Figure 7: La structure quaternaire de la protéine.

Pour plus d'information sur la « classification, Propriétés, et Liaisons qu'interviennent dans la stabilité des protéines» voire (ANNEXE 1 section : 1)

III. Les motifs protéiques

Les motifs protéiques peuvent être définis d'après leur séquence primaire ou d'après l'arrangement des éléments de structure secondaire

Le terme motif est utilisé de deux façons différentes en biologie structurale, le premier fait référence à une séquence particulière d'acides aminés, caractéristique d'une fonction biochimique spécifique.

Le motif doigt à zinc, CXX(XX) CXXXXXXXXXXXXHXXXH, qui est présent dans une famille très variée de protéines de liaison à l'ADN (figure 8) est un exemple.

Les résidus cystéine et histidine conserve dans ce motif de séquence forment des ligands pour un ion zinc dont la coordination est essentielle pour stabiliser la structure tertiaire.

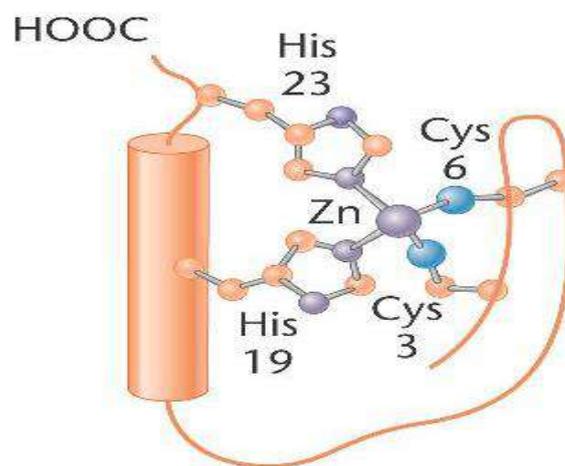


Figure 8 : Doigt de zinc

La conservation correspondant parfois à une classe de résidus plutôt qu'à un résidu spécifique par exemple, dans la boucle de 12 résidus entre les ligands du zinc, une position est préférentiellement hydrophobe et contient généralement une leucine ou une phénylalanine.

Des motifs de séquence peuvent souvent être identifiés grâce au simple examen de la séquence d'acides aminés d'une protéine, la protéase issue de virus de l'immunodéficience humaine a été identifiée comme étant une aspartyl protéase en raison d'un motif de séquence caractéristique de ces protéases, reconnu dans sa structure primaire.

Le deuxième usage de terme, tout aussi courant, fait référence à un ensemble d'éléments contigus de structure secondaire qui présente une importance fonctionnelle particulière ou définissent une portion d'un domaine replié indépendant. Comme les

motifs fonctionnelles de séquence, Ces dernier sont généralement appelée motifs fonctionnelles.

Le motif hélice-tour-hélice présent des nombreuses protéines de liaison à l'ADN figure (9)

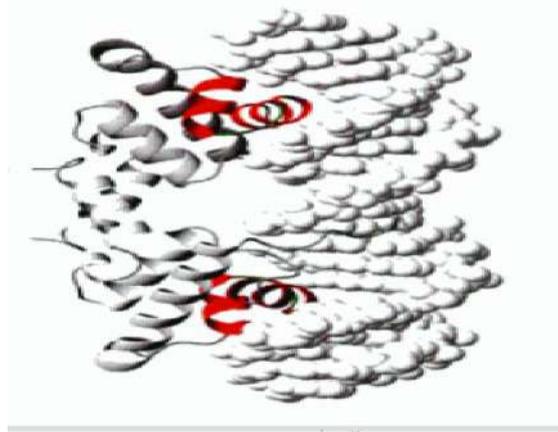


Figure 9. Le Motif hélice-Touere-hélice

Ce motif structural simple n'existera pas sous la forme d'un domaine replié de façon stable s'il est exprimé séparément du reste de son contexte protéique, mais lorsqu'on peut le détecter dans une protéine déjà soupçonnée de se fixer à des acides nucléiques, il est un candidat potentiel au rôle d'élément de reconnaissance.

Les exemples des motifs structuraux qui représentent une part importante d'un domaine replié de façon stable comprennent le faisceau de quatre hélice, un groupe de quatre hélices alpha antiparallèles que l'on trouve dans des nombreuses hormones ainsi que d'autre types de protéines

Comme l'indiquent ces exemples, ces motifs structuraux donnent parfois un indice de la fonction des protéines, mais le plus souvent ce n'est pas le cas : la seule situation dans laquelle les implications fonctionnelles sont évidentes (Altken, A., 1999)

IV .Les méthodes expérimentales de détermination des structures protéiques

IV.1La cristallographie rayon X

Le principe de la d' détermination d'une protéine par cristallographie au rayon X est de déterminer la position de ses atomes afin d'en déduire un modèle structural. La position précise de chaque atome d'une molécule peut être déterminée seulement si la molécule est cristallisée (l'organisation des atomes forme des cristaux).

Lorsque les rayons X frappent une molécule cristallisée, les électrons entourant chaque atome courbent ou diffractent le faisceau rayon X, ce phénomène permet de déduire un modèle de diffraction par rayon X par analyse de la densité des électrons. Un ordinateur interprété alors mathématiquement ce modèle et reconstruit la position

des atomes pour finalement donner un modèle quasi exact de la structure de la molécule

IV.2 La spectroscopie RMN

La spectroscopie par résonance magnétique nucléaire (RMN) repose sur l'analyse des noyaux atomiques de la molécule. Chaque noyau porte une charge. Lorsque les spins des protons et des neutrons d'un noyau sont impairs, le spin total du noyau crée un dipôle magnétique le long de l'axe spinal, la magnitude intrinsèque de ce dipôle est une propriété nucléaire fondamentale appelée le moment magnétique nucléaire.

Depuis cette propriété est établie une corrélation entre la symétrie de la distribution de charge du noyau et sa structure interne. Par la suite, les distances entre certains atomes de la molécule sont connues. La connaissance de ces distances permet alors de déterminer un modèle de la molécule satisfaisant toutes les contraintes stéréochimiques liées à la structure. La structure de la molécule est finalement déduite (Guillaume Chakroun.,2004)

IV.3 Cryo-microscopie électronique (cryo-EM) :

Cryo-microscopie électronique (cryo-EM) est de plus en plus un grand public pour l'étude de l'architecture des cellules, les virus et assemblages de protéines à une résolution moléculaire.

Au cours de la dernière décennie, le "cryo-microscopie électronique" phrase, souvent abrégé en «cryo-EM», a évolué pour englober un large éventail de méthodes expérimentales. À la base, chacune d'entre elles est basée sur le principe de l'imagerie d'échantillons sensibles au rayonnement dans un microscope électronique à transmission, dans des conditions cryogéniques. En biologie, les applications de cryo-EM maintenant couvrent un large spectre allant de l'imagerie de coupes de tissus et des cellules intactes en plongée congelé à des bactéries, des virus et des particulières molécules de protéines. Tomographie cryo-électronique, la microscopie à une particule cryo-électronique, et cristallographie électronique sont tous les sous-disciplines de la cryo-EM qui ont été utilisées avec succès pour analyser les structures biologiques dans différents contextes (**Hoenger A**).

B. Les Glucides :

Les sucres ou saccharides sont les biomolécules les plus abondantes sur la planète. Ils sont importants dans bon nombre de rôles biologiques. Ce sont les constituants majoritaires de l'alimentation. Les sucres insolubles servent également de matériel structurel dans la paroi cellulaire des végétaux et des bactéries ainsi que dans les tissus conjonctifs et membranes cellulaires des animaux. Les polymères de sucres servent également à lubrifier les articulations du squelette, ont une fonction de reconnaissance cellulaire, de sites antigéniques, etc. (S. Gillet).

Le terme de « glucides » est synonyme de « hydrates de carbone » ou de « saccharides » (Lehninger, 1981). Il s'agit de polyalcools comportant une fonction aldéhyde (CHO) ou cétone (R-CO-R'). La plupart des glucides répondent à la formule brute $(CH_2O)_n$ avec $n > 3$. Ils sont subdivisés principalement en OSES et OSIDES

Pour plus d'informations sur (classification des glucides, oses, osides, polyosides) Voir (Annexe1, Section : 2)

C .Les glycoprotéines :

Les glycoprotéines sont des hétéroprotéines dont une partie est constituée d'une ou plusieurs chaînes d'acides aminés, mais sur laquelle ou lesquelles viennent se greffer par des liaisons covalentes des chaînons constitués de plusieurs oses ou dérivés d'oses (oligosaccharides ou glycanes).

L'attachement de certains oligosaccharides se fait par une liaison N-osidique sur l'amide d'une asparagine de la structure primaire du polypeptide

(Protéine N-glycosylée). Le deuxième acide aminé en aval (côté COOH-terminal) de cette Asn est toujours un acide aminé alcool : sérine ou thréonine, formant un site de glycosylation : Asn-X-Ser ou Asn-X-Thr (**Pr A. Raisonnier**).

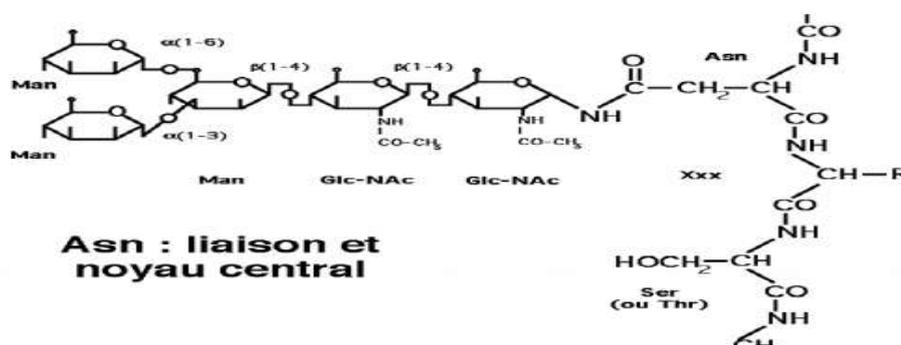


Figure 10 : Structure de glycoprotéines

Pour voir des exemples sur les glycoprotéines voir Annexe 1(Section 3).

Les glucides jouent un rôle majoritaire dans la survie des êtres vivants ils se lient avec des protéines et des enzymes pour assurer son rôle biologique (métabolisme, pouvoir antigénique etc.), des études sont effectuées sur la structure 3D des glucides qui se lient à des protéines et des enzymes et sont stockées dans une base de données mondiale qui concerne les structures de macromolécules

D .La banque de données sur les protéines (PDB) :

La modélisation moléculaire est importante dans la biologie il aide à la compréhension des structures et fonctions des macromolécules et voir les contacts qui font entre ces dernières, ces structures 3D sont stockées dans une banque de données pour les protéines nommée Protéine Data Bank (PDB)

1. Qu'est-ce que la PDB :

La banque de données des protéines (PDB) au Brookhaven National Laboratory (BNL), est une base de données contenant expérimentalement des structures tridimensionnelles déterminées des protéines, des acides nucléiques et d'autres macromolécules biologiques (Abola et al., 1987, 1997; Bernstein et al., 1977).

Protéine Data Bank est la banque de référence des structures protéiques obtenues expérimentalement par cristallographie rayon X ou spectroscopie RMN. Jusqu'en mai 2003, 21000 structures protéiques ont été entrées dans la banque ; de 50 à 100 nouvelles structures sont déposées chaque semaine. Les coordonnées des atomes formant la structure d'une protéine, le détail de la séquence, les conditions de cristallisation sont les principales informations disponibles pour chaque structure de la banque PDB (Guillaume Chakroun.,2004).

2. Le world wide PDB (wwPDB) :

Depuis 1971, la protéine archives Data Bank (PDB) a servi de référentiel unique d'information sur les structures 3D des protéines, des acides nucléiques et des assemblages complexes.

Le Worldwide PDB (wwPDB) organisation gère l'archive PDB et assure que le PDB est librement et publiquement disponible à la communauté mondiale (wwpdb.org).

2.1. Membres de wwPDB :

Le wwPDB a trois contributeurs internationaux le PDBj(Protein Data Bank Japan),pdbe(Protein Data Bank in Europe), RCSB PDB (les États-Unis).

2.1.1. Le PDBe :

Le PDBe est la ressource européenne pour la collecte, l'organisation et la diffusion de données sur les structures macromoléculaires biologiques

Objectifs

Afin de fournir une ressource intégrée des structures macromoléculaires de haute qualité et des données connexes et les rendre accessibles à la communauté biomédicale via des interfaces utilisateur intuitives.

Pour maintenir l'expertise interne dans toutes les grandes techniques structure détermination (rayons X, RMN et EM) afin de rester au courant des développements techniques et méthodologiques dans ces domaines, et de travailler avec la communauté sur les questions d'intérêt commun (comme que la représentation des données, la récolte, les formats et les normes, ou la validation des données structurales).

Pour fournir des dépôts et d'annotation des installations de haute qualité pour les données structurales comme un site de dépôt wwPDB et EMDataBank.

2.1.2. Le PDBj :

PDBj signifie la Protéine Data Bank Japon. Du site PDBj, vous pouvez accéder à des structures macromoléculaires, dépôt de données expérimentales, et également faire usage d'une variété de services de bioinformatique web structurales

Les services de PDBj sont principalement basées sur les fichiers XML (extensible markup language) appelés PDBML et PDBML plus

2.1.3. RCSB PDB (les États-Unis) :

Le Collaboratoire recherche de bioinformatique structurale (RCSB) est devenu responsable de la gestion de le PDB. En 2003, le wwPDB a été formé afin de maintenir une archive unique PDB de données structurales macro moléculaire qui est librement accessible au public et à la communauté mondiale. Il se compose d'organismes qui agissent en tant que centres de dépôt, de traitement et de distribution de données pour les données PDB.

En outre, le PDB RCSB soutient un site Web où les visiteurs peuvent effectuer des requêtes simples et complexes sur les données, analyser et visualiser les résultats. Détails sur l'histoire, la fonction.

3 .Formats des fichiers :

Plusieurs formats existent pour les fichiers de la PDB :

3.1. Le fichiers PDB :

Les fichiers PDB contiennent les coordonnées cartésiennes des atomes qui constituent la molécule ainsi que des métadonnées. Ces métadonnées peuvent par exemple être la structure primaire de la molécule, ses éventuelles structures secondaires, la méthode expérimentale qui a permis d’obtenir les coordonnées des atomes, etc. (Stanislas Drey, 2010).

1 – Le fichier PDB est composé de lignes de 80 colonnes où chaque colonne a une signification propre

A l’origine, le format PDB a été dicté par la largeur de cartes perforées IBM pour ordinateur.

En conséquence, chaque ligne contient exactement 80 colonnes, soit 80 caractères.

```

      80 colonnes = 80 caractères
←-----→
HEADER   OXYGEN STORAGE                               19-DEC-08  2W6X
TITLE    CRYSTAL STRUCTURE OF SPERM WHALE MYOGLOBIN MUTANT YQRF IN
TITLE    2 COMPLEX WITH XENON
COMPND   MOL_ID: 1;
COMPND   2 MOLECULE: MYOGLOBIN;
COMPND   3 CHAIN: A;
COMPND   4 ENGINEERED: YES;
COMPND   5 MUTATION: YES
SOURCE   MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: PHYSETER CATODON;
SOURCE   3 ORGANISM_COMMON: SPERM WHALE;
    
```

Figure 11 : Extrait des données brutes du début d’un fichier PDB. Il est visible que chaque ligne contient exactement 80 colonnes, soit 80 caractères.

Chaque colonne possède sa signification, ainsi les 6 premières colonnes, c’est-à-dire les 6

Premiers caractères pour une ligne donnée, déterminent le champ. Certains champs, qu’il est possible de trouver, sont présentés dans le tableau ci-dessous à titre d’exemple.

Nom du champ	Description du champ
TITLE	Titre de la macromolécule étudiée
KEYWDS	Les mots-clé de l’entrée
EXPDTA	Donne des informations sur la méthode expérimentale employée
SEQRES	La séquence de la protéine étudiée
ATOM	Coordonnées des atomes des résidus standard
HETATM	Coordonnées des atomes des résidus non standard (solvant, substrat, ion, détergent...)

Tableau 1 : Exemple de champs avec leur description.

Les 6 premières colonnes définissent le champ de la ligne

Colonne n°	1	2	3	4	5	6	7	8
Ligne 1	HEADER	OXYGEN STORAGE			19-DEC-08	2W6X		
Ligne 2	TITLE	CRYSTAL STRUCTURE OF SPERM WHALE MYOGLOBIN MUTANT YQRF IN						
Ligne 3	TITLE	2 COMPLEX WITH XENON						
Ligne 4	COMPND	MOL_ID: 1;						

Figure 12 : Les 6 premières colonnes définissent le champ de la ligne.

Le champ de la ligne 1 est donc HEADER, celui des lignes 2 et 3 est TITLE et celui de la ligne 4 est COMPND.

HEADER	OXYGEN TRANSPORT				10-JUN-83	1HHO	1HHO	3		
COMPND	HEMOGLOBIN A (OXY)							1HHO	4	
SOURCE	HUMAN (HOMO SAPIENS)							1HHO	5	
AUTHOR	B.SHAANAN							1HHO	6	
REVDAT	2	31-JAN-84	1HHOA	1	JRNL			1HHOA	1	
REVDAT	1	27-OCT-83	1HHO	0				1HHO	7	
JRNL	AUTH	B.SHAANAN						1HHO	8	
JRNL	TITL	STRUCTURE OF HUMAN OXYHAEMOGLOBIN AT 2.1 ANGSTROMS						1HHOA	2	
JRNL	TITL 2	RESOLUTION						1HHOA	3	
JRNL	REF	J.MOL.BIOL.			V. 171	31 1983		1HHOA	4	
JRNL	REFN	ASTM JMOBAC UK ISSN 0022-2836					070	1HHOA	5	
REMARK	1								1HHO	13
REMARK	1	REFERENCE 1							1HHO	14
REMARK	1	AUTH	B.SHAANAN						1HHO	15
REMARK	1	TITL	THE IRON-OXYGEN BOND IN HUMAN OXYHAEMOGLOBIN						1HHO	16
REMARK	1	REF	NATURE			V. 296	683 1982	1HHO	17	
REMARK	1	REFN	ASTM NATUAS UK ISSN 0028-0836					006	1HHO	18

Figure 13 : Comparaison des données brutes de fichiers PDB de molécule (1HHO)

3.2. Les champs dans les fichiers PDB

Les 6 premiers caractères pour une ligne donnée déterminent le champ. l'ordre des champs est toujours le même, cependant ils ne sont pas tous obligatoires.

Le premier champ de tout fichier PDB est le champ « HEADER » et il est obligatoire. Cette première ligne va permettre d'identifier le fichier PDB – elle indique la molécule concernée par le fichier, la date à laquelle les coordonnées de la molécule ont été reçues par la PDB et l'identifiant de la molécule (unique au sein de la PDB).

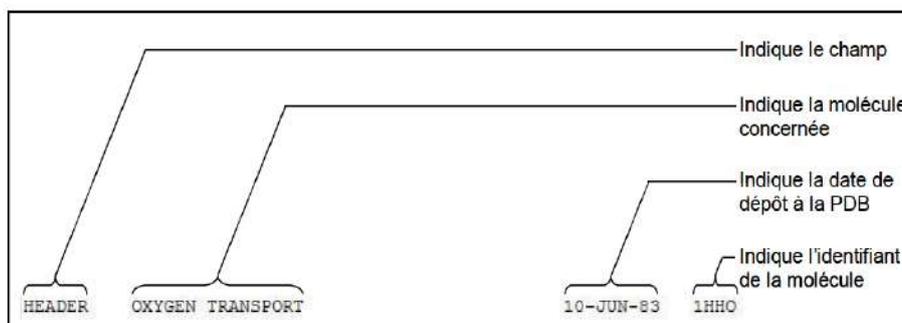


Figure 14 : Détail du champ « HEADER ». Ici ce champ indique que la molécule est Oxygène transport, que ses coordonnées ont été déposées sur la PDB le 10 juin 1983 et que son identifiant est 1HHO.

Les différents champs peuvent être regroupés en sections, comme montré dans le tableau ci-dessous.

Section	Description	Types de champs
Titre	Remarques descriptives résumées	HEADER, OBSLTE, TITLE, CAVEAT, COMPND, SOURCE, KEYWDS, EXPDTA, AUTHOR, REVDAT, SPRSDE, JRNL
Remarque	Bibliographie, divers détails	REMARKS 1, 2, 3 & annotations
Structure primaire	Séquence peptidique et/ou nucléotidique et la relation entre la séquence PDB et la séquence trouvée dans la/les base(s) de données	DBREF, SEQADV, SEQRES, MODRES
Hétérogène	Description de groupes non standard	HET, HETNAM, HETSYN, FORMUL
Structure secondaire	Description des structures secondaires	HELIX, SHEET, TURN
Annotation sur les liaisons	Liaison chimique	SSBOND, LINK, HYDBND, SLTBG, CISPEP
Divers caractéristiques	Caractéristique dans la macromolécule	SITE
Crystallographie	Description de la cellule cristallographique	CRYST1
Transformation de coordonnées	Les opérateurs de transformation de coordonnées	ORIGXn, SCALEn, MTRIXn, TVECT
Coordonnées	Données sur les coordonnées des atomes	MODEL, ATOM, SIGATM, ANISOU, SIGUIJ, TER, HETATM, ENDMDL
Liaison	Liaison chimique	CONNECT
Comptabilité	Résumé des informations, marqueur de fin	MASTER, END

Tableau 2 : Les différentes sections qui regroupent les différents champs

3.3. Le champ « ATOM »

Le fichier PDB donne les coordonnées des atomes pour les résidus standards (acides aminés, acides nucléiques et polysaccharide). Pour les atomes des résidus non-standard c'est le champ « HETATM » qui donne leurs coordonnées atomiques.

La signification des colonnes pour le champ « ATOM » est indiquée dans le tableau ci-dessous.

Colonnes	Signification des colonnes
1 à 6	Indique que c'est le champ ATOM
7 à 11	Numéro de série de l'atome
13 à 16	Nom de l'atome
17	Indicateur d'une localisation alternative
18 à 20	Nom du résidu
22	Identifiant de la chaîne
23 à 26	Numéro du résidu
27	Code pour l'insertion des résidus
31 à 38	Coordonnées orthogonales pour X en Angstrom
39 à 46	Coordonnées orthogonales pour Y en Angstrom
47 à 54	Coordonnées orthogonales pour Z en Angstrom
55 à 60	Occupation
61 à 66	Facteur de température
77 à 78	Symbole de l'élément
79 à 80	Charge de l'atome

Tableau 3 : Description du champ « ATOM »

Colonne n°	1	2	3	4	5	6	7	8	
ATOM	80	CE	LYS A	11	16.865	23.437	0.587	1.00100.00	C
ATOM	81	NZ	LYS A	11	16.300	24.743	1.031	1.00 50.42	N
ATOM	82	N	ALA A	12	21.743	18.025	1.437	1.00 18.17	N
ATOM	83	CA	ALA A	12	23.107	17.671	1.637	1.00 10.44	C
ATOM	84	C	ALA A	12	23.555	16.291	1.157	1.00 15.53	C
ATOM	85	O	ALA A	12	24.697	16.040	0.726	1.00 16.26	O

Figure 15 : Exemple de champs « ATOM » qu'il est possible de trouver dans un fichier PDB

A noter que les atomes d'hydrogène sont rarement présents dans un fichier PDB.

Pour les protéines, les résidus sont listés de l'extrémité amine à l'extrémité carboxyle.

Pour les acides nucléiques, les résidus sont listés de l'extrémité 5' à l'extrémité 3'. Il n'y a pas d'ordre spécifique pour les polysaccharides.

La liste des champs « ATOM » d'une chaîne se termine par un champ « TER ».

ATOM	1067	NH1	ARG	A	141	-2.911	10.577	-1.995	1.00	52.81	N
ATOM	1068	NH2	ARG	A	141	-0.604	10.395	-2.146	1.00	65.36	N
ATOM	1069	OXT	ARG	A	141	-6.729	15.170	-5.560	1.00	82.14	O
TER	1070		ARG	A	141						
ATOM	1071	N	VAL	B	1	9.445	-18.730	-3.132	1.00	58.52	N

Figure 16 : Ici le champ « TER » marque la fin de la chaîne A.

Le fichier PDB permet le lien entre les données obtenues expérimentalement et les logiciels de visualisation moléculaire

Pouvoir afficher un modèle de la molécule. Le fichier PDB ne donne pas d'information sur la liaison entre les atomes (sauf liaisons spécifiques comme les ponts disulfures). En effet, le logiciel va, à partir de l'ordre des atomes et de leur position spatiale (obtenu grâce aux coordonnées tridimensionnelles), calculer les liaisons entre les atomes.

Les coordonnées 3D rentrées dans le fichier PDB sont obtenues de manière expérimentale

(Ex : diffraction aux rayons X). Et ce sont donc ces coordonnées qui vont conditionner la représentation de la molécule affichée par le logiciel.

Le champ « EXPDTA », qui est obligatoire, indique la méthode expérimentale utilisée pour obtenir les coordonnées des atomes (**Stanislas Drey, 2010**).

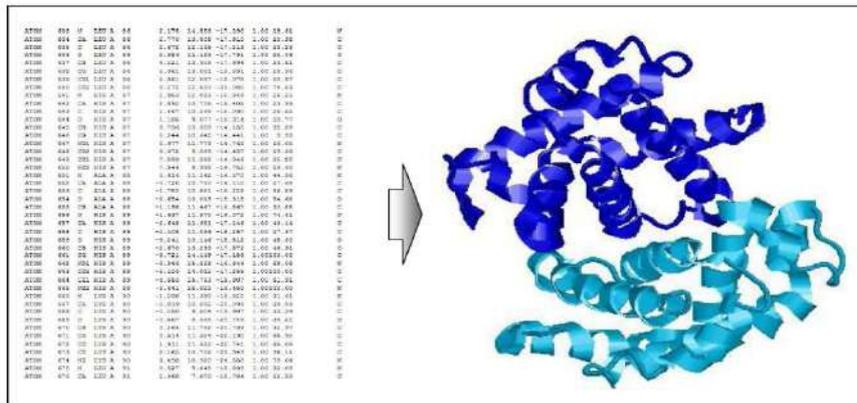


Figure 17 : Illustration de ce que permet le logiciel de visualisation moléculaire.

3.4. Limites du format PDB

Bien que très utilisé, le format PDB présente certains défauts. En effet, le format en 80 colonnes est relativement restrictif.

Cela limite le nombre maximum d'atomes d'un fichier pdb qui est de 99999, vu qu'il n'y a que 5 colonnes allouées pour les numéros des atomes (colonnes n°7 à n°11). De même, le nombre de résidus par chaîne est au maximum de 9999 : il n'y a que 4 colonnes autorisées pour ce nombre (colonnes n°23 à 26). Le nombre de chaînes, lui, est limité à 62 (une seule colonne est disponible (colonne 22), et les valeurs possibles sont parmi les 26 lettres de l'alphabet, en minuscule ou en majuscule, ou l'un des chiffres de 0 à 9) (Stanislas Drey, 2010).

Colonne n°	1	2	3	4	5	6	7	8		
ATOM	80	CE	LYS	A	11	16.865	23.437	0.587	1.00100.00	C
ATOM	81	NZ	LYS	A	11	16.300	24.743	1.031	1.00 50.42	N
ATOM	82	N	ALA	A	12	21.743	18.025	1.437	1.00 18.17	N
ATOM	83	CA	ALA	A	12	23.107	17.671	1.637	1.00 10.44	C
ATOM	84	C	ALA	A	12	23.555	16.291	1.157	1.00 15.53	C
ATOM	85	O	ALA	A	12	24.697	16.040	0.726	1.00 16.26	O

Figure 18: limitations de format PDB

4. Le format mmCIF

Etant donné les limitations du format PDB, l'Union internationale de la cristallographie (International Union of Crystallography, IUCr), en 1990 étendu aux macromolécules la représentation des données utilisées pour décrire les structures cristallographiques des molécules de faible poids moléculaire (appelée CIF, pour Crystallographic Information File).

A partir de celle-ci, le format mmCIF (macromolecular Crystallographic Information File) a donc été développé. La première version du format mmCIF a été publiée en 1996.

Ce format, mmCIF, permet une représentation plus structurée, uniformisée et non limitée. En mmCIF, chaque champ de chaque section d'un fichier PDB est représenté par une description d'une caractéristique d'un objet, qui comprend, d'une part, le nom de la caractéristique (par exemple : `_struct.entry_id`), et, d'autre part, le contenu de la description (pour `_struct.entry_id` ça sera le code pdb : 1cbn). On parle de paire « nom-valeur ».

Il est aisé de convertir, sans perte d'informations, un fichier mmCIF au format PDB, puisque toute l'information est directement analysable. Il n'est pas possible, en revanche, de complètement automatiser la conversion d'un fichier PDB au format mmCIF, puisque plusieurs descripteurs mmCIF sont, soit absents du fichier PDB, soit présents dans un champ « REMARK » qui ne peut pas toujours être analysé.

La documentation sur le format mmCIF peut être accessible à cette adresse : <http://mmcif.pdb.org>.

```
data_2W6X
#
_entry.id 2W6X
#
_audit_conform.dict_name      mmcif_pdbx.dic
_audit_conform.dict_version   1.0675
_audit_conform.dict_location  http://mmcif.pdb.org/dictionaries/ascii/mmcif_pdbx.dic
#
loop_
_database_2.database_id
```

Figure 19 : Exemple du début d'un fichier au format mmCIF.

Comme nous avons vu précédemment que la protéine il à trios niveaux de structures (primaire secondaire et tertiaire), et chaque niveau de structure il a leur détails et leur complexité.

Pour cette raison les chercheurs ont créé des nouvelles banques de structure.

5. Les autres banques de structures protéiques

5.1. SCOP (Structural Classification of Proteins) :

Banque de données regroupant les protéines de la PDB présentant une relation de similarité structurale et d'évolution. Le classement dans la banque se fait selon trois principaux niveaux hiérarchiques :

1. Famille : relation d'évolution commune clairement établie. Le degré de similarité de séquence entre paires de résidus est supérieur à 30%.
2. Superfamille : origine d'évolution commune probable. Le degré de similarité de séquence est faible, néanmoins, les dispositifs structuraux et fonctionnels des protéines suggèrent une évolution d'origine commune.
3. Repliement : similarité structurale majeure. Les protéines partagent ici une majorité de structures secondaires dans le même arrangement et avec les mêmes connexions topologiques. Les régions en coude et périphériques aux autres structures secondaires différentes souvent en taille et en conformation. Les protéines ayant des repliements communs ne semblent pas avoir une origine d'évolution commune, seules les propriétés physiques et chimiques leur donnent un arrangement similaire (Guillaume Chakroun .,2004).

5.2. CATH (Class Architecture Topology and Homology) :

Les protéines dont la structure a été déterminée par RMN et celles dont la résolution de détermination des structures par cristallographie rayon X est supérieure à 3 angströms sont répertoriées dans cette banque. La hiérarchie de classification de CATH est composée de quatre principaux niveaux :

Niveau C (Classe) : sont répertoriées les structures protéiques appartenant aux classes α et β . Une troisième classe $\alpha - \beta$ regroupe à la fois les structures des classes α/β et $\alpha + \beta$. Une quatrième classe contient les domaines protéiques avec peu de structures secondaires.

2. Niveau A (Architecture) : classe les protéines en fonction de l'architecture de leurs structures secondaires sans tenir compte de leurs connectivités. Ce niveau regroupe principalement les protéines adoptant les architectures en baril ou en sandwich trois couches.

3. Niveau T (Topologie) : les protéines sont classées par familles de repliement, tenant compte de l'architecture des structures secondaires mais également de leurs connectivités.

4. Niveau H (superfamilles homologues) : regroupe les domaines protéiques considérés comme partageant un ancêtre commun et pouvant ainsi être décrites comme homologues. Les similarités sont d'abord identifiées par comparaison de séquences puis par comparaison de structure. (Guillaume Chakroun.,2004).

Chapitre II : Préparation des Données

Les structures de protéines sélectionnées pour ce projet étaient principalement ceux en complexe avec des ligands qui sont de nature glucidique.

Généralement les structure des protéines dans le PDB sont stocker sous forme des entrée et ces entrée possèdent des codes appelée code d'identification ou PDB IDs et qui sont utilisée pour accéder a les donnée fournies.

1- Les structures protéiniques (codes PDB) :

L'ensemble des codes PDB utilisée dans cette projet et qui sont 100 code sont déterminée par la méthode cristallographie aux rayons X.

Ces codes ont des classes différentes comme vu ci-dessous (tableau 04)

Tableau 04: Liste des structures de protéines avec des ligands liés utilisés dans l'étude.

Résolution et R-factor reflètent la qualité de la structure; des valeurs plus petites structures signifient plus fiables.

Classe	Titre de l'Antée PDB	PDB ID	Résolution	R-Factor	Ligands en Contact
1.Oxydoréductase	Soluble Quinoprotein Glucose Dehydrogenase from Acinetobacter Calcoaceticus in Complex with PQQH2 and Glucose	1CQ1	1.90	0.190	BGC
	Sulfolobus Solfataricus Glucose Dehydrogenase 1 in complex with NADP and Glucose	2CDB	1.60	0.192	BGC
	Trypanosoma cruzi Glucose-6-P Dehydrogenase in complex with G6P	4EM5	3.35	0.218	
	Complex structure of Lactobacillus casei lactate dehydrogenase with fructose-1,6-bisphosphate	3VKV	2.70	0.200	FBP
	L-lactate dehydrogenase from Thermus caldophilus GK24 complexed with oxamate, NADH and FBP	3VPH	2.00	0.209	FBP

	Crystal structure of L-lactate dehydrogenase from <i>Bacillus subtilis</i> complexed with FBP and NAD ⁺	3PQD	2.38	0.215	FBP
	Glucose Oxidase from <i>Penicillium Amagasakiense</i>	1GPE	1.80	0.164	NAG
2. Binding protein	novel stereospecificity of the Arabinose-Binding protein	1ABE	1.70	0.137	ARA ARB
	Identical mutations at corresponding positions in two Homologous Proteins with NON-Identical effects	1DBP	2.20	0.160	RIP
	Crystal structure of the type 1 RIP bouganin	1DR K	1.80	0.176	RIP
	Sugar-Binding and Crystallographic studies of an Arabinose-Binding protein mutant (met108leu) which exhibits enhanced affinity and altered specificity	6ABP	1.67	0.160	ARB ARA
3. DNA Binding Protein	Structure of the effector-binding domain of arabinose repressor AraR from <i>Bacillus subtilis</i>	3TB6	2.21	0.188	ARB
4. Hydrolase	Crystal structure of Human chitinase in Complex with Glucoallosamidin B	1HKI	2.55	0.223	NAG
	Crystal structure of AAL mutant H59Q complex with lactose	2ZGO	2.00	0.200	LAT
	Crystal Structure of beta-L-Arabinopyranosidase complexed with L-arabinose	3A22	1.90	0.164	ARA
	mouse SMP30/GNL-xylitol complex	4GN A	1.85	0.167	XYL
	Crystal Structure of Carbohydrate-Binding Module Family 28 from <i>Clostridium josui</i> Cel5A in complex with cellopentaose	3ACI	1.60	0.144	BGC

	Structure Of The 6-Phospho-Beta Glucosidase From Thermotoga Maritima At 2.55 Angstrom Resolution In The Tetragonal Form With Manganese, Nad+ And Glucose-6-Phosphate	1UP6	2.55	0.202	G6P
5. Toxin	Vibrio Cholerae Cytolysin Beta-Prism Domain With Methyl-Alpha-Mannose Bound	4GX7	2.85	0.197	MMA
6. Plant Protein	Canavalia gladiata lectin complexed with Man1-3Man-OMe	2EF6	2.10	0.192	MDM
7. Apoptosis	Crystal Structure of Galectin-1 L11A mutant	4NO4	1.40	0.188	LAT
8. Isomerase	Crystal Structure Of Mycobacterium Tuberculosis Ribose-5-Phosphate Isomerase B In Complex With Alpha D-Allose 6-Phosphate	2VV O	1.85	0.162	A6P
	A Metal-Mediated Hydride Shift Mechanism For Xylose Isomerase Based On The 1.6 Angstroms Streptomyces Rubiginosus Structures With Xylitol And D-Xylose	3XIS	1.60	0.151	XLS XYL
	Modes Of Binding Substrates And Their Analogues To The Enzyme D-Xylose Isomerase	1XII	1.70	0.147	XUL
	Crystal Structure Of Mycobacterium Tuberculosis Ribose-5-Phosphate Isomerase B In Complex With Its Substrates Ribose 5-Phosphate And Ribulose 5-Phosphate	2VVP	1.65	0.168	5RP
	Phosphoglucose isomerase from Leishmania mexicana in complex with substrate D-fructose-6-phosphate	1T10	2.35	0.222	F6P
	Crystal structure of mouse AMF / F6P complex	2CXS	1.50	0.173	F6P
	Crystal Structure of 3,4-Dihydroxy-2-butanone 4-phosphate Synthase of	1TKU	1.66	0.184	5RP

Candida albicans in complex with Ribulose-5-phosphate				
D254E, D256E Mutant Of D-Xylose Isomerase Complexed With A13 And Xylitol	1XL M	2.40	0.179	XYL
Modes Of Binding Substrates And Their Analogues To The Enzyme D-Xylose Isomerase	1XIC	1.60	0.152	XLS
Modes Of Binding Substrates And Their Analogues To The Enzyme D-Xylose Isomerase	1XIG	1.70	0.162	XYL
Crytal Structure of hRPE and D-Ribulose-5-Phospate Complex	3OV Q	2.00	0.149	5RP
Clostridium thermocellum Ribose-5-Phosphate Isomerase B with d-allose	3PH4	2.07	0.186	AOS
A Metal-Mediated Hydride Shift Mechanism For Xylose Isomerase Based On The 1.6 Angstroms Streptomyces Rubiginosus Structures With Xylitol And D-Xylose	2XIS	1.71	0.149	XYL
Crystal Structure of L-Rhamnose Isomerase from Pseudomonas stutzeri in Complex with D-Allose	2I57	1.97	0.172	AOS
Crystal Structure Of Rabbit Phosphoglucose Isomerase Complexed With Fructose-6-Phosphate	1HO X	2.10	0.221	F6P
Crystal structure of rabbit phosphoglucose isomerase complexed with 5-phospho-D-arabinonohydroxamic acid	1KOJ	1.90	0.223	PAN
Crystal structure of galactose mutarotase from Lactococcus lactis complexed with L-arabinose	1MM Z	1.80	0.172	ARB
X-RAY Analysis of D-Xylose Isomerase At 1.9 Angstroms: Native Enzyme In Complex With Substrate And With A Mechanism-Designed Inactivator	8XIA	1.90	0.140	XLS

Crystal structure of Acinetobacter sp. DL28 L-ribose isomerase in complex with L-ribose	4Q0P	1.93	0.208	OMK
Structures Of D-Xylose Isomerase From Arthrobacter Strain B3728 Containing The Inhibitors Xylitol And D-Sorbitol At 2.5 Angstroms And 2.3 Angstroms Resolution, Respectively	5XIA	2.50	0.135	XYL
Crystal structure of Pseudomonas stutzeri L-rhamnose isomerase mutant H101N in complex with D-allopyranose	4GJJ	2.38	0.196	AOS
Crystal structure of Acinetobacter sp. DL28 L-ribose isomerase in complex with L-ribose	4Q0P	1.93	0.208	OMK
Joint X-ray and neutron structure of Streptomyces rubiginosus D-xylose isomerase in complex with two Cd ²⁺ ions and cyclic beta-L-arabinose	4QDP	2.00	0.231	ARB
Structure of a glucose-6-phosphate isomerase from Trypanosoma cruzi	4QFH	1.80	0.140	G6P
Crystal structure of Pseudomonas stutzeri L-rhamnose isomerase mutant S329F in complex with D-allose	3M0 M	1.45	0.177	AOS
A Metal-Mediated Hydride Shift Mechanism For Xylose Isomerase Based On The 1.6 Angstroms Streptomyces Rubiginosus Structures With Xylitol And D-Xylose	4XIS	1.60	0.135	XLS XYS
X-ray structure of Cellulomonas parahominis L-ribose isomerase with L-ribose	3WW 1	1.95	0.206	OMK
X-ray structures of Cellulomonas parahominis L-ribose isomerase with L-allose	3WW 4	1.95	0.217	3BU

	crystal structure of RpiA-R5P complex	4M8L	2.37	0.239	5RP
9. Hydrolase	The Acidic Mammalian Chitinase catalytic domain in complex with methylallosamidin	3FY1	1.85	0.162	NA1
	Crystal Structure Of Human Chitinase In Complex With Methylallosamidin	1HKJ	2.60	0.211	NA1
	X-RAY Crystallography of The Binding of The Bacterial Cell Wall Trisaccharide Nam-Nag-Nam To Lysozyme	9LYZ	2.50	0.230	NAG
	Crystal structure of fructose biphosphate aldolase from Bartonella henselae, bound to fructose biphosphate	3MM T	2.35	0.179	2FP
	E. coli (lacZ) beta-galactosidase (N460S) in complex with L-ribose	4DU X	2.30	0.170	OMK

10. Lyase	Fructose-1,6-bisphosphate Schiff base intermediate in FBP aldolase from Leishmania Mexicana	2QDG	2.20	0.185	2FP
	Fructose 1,6-Bisphosphate Aldolase From Rabbit Liver	1FDJ	2.10	0.174	2FP
	Crystal structures of complex F. succinogenes 1,3-1,4-beta-D-glucanase and beta-1,3-1,4-celotriose	1ZM1	2.30	0.180	BGC
	Structure of ribB complexed with substrate Ru5P	4P77	2.04	0.176	5RP
11. Transport	The Structure Of Bacillus Subtilis Rbsd Complexed With D-Ribose	1OGD	1.95	0.198	RIP
12. Transport Protein	Crystal structure of lactose permease with TDG	1PV7	3.60	0.271	TDG
	Crystal structure of carbohydrate transporter solute binding protein Arad_9553 from Agrobacterium Radiobacter, Target EFI-511541, in complex with D-arabinose	4RXT	1.35	0.157	ARA
	HA17-HA33-Gal	4LO1	2.25	0.200	GAL

13. Tranferase	A. aeolicus KDO8PS R106G mutant in complex with PEP and R5P	1ZHA	1.74	0.207	R5P
	Crystal structure of the Fucosylgalactoside alpha N-acetylgalactosaminyltransferase (GTA, cisAB mutant L266G, G268A) in complex with in complex with NPE caged UDP-Gal (P2(1)2(1)2(1) space group)	3ZGF	0.183	0.217	IUG
	Crystal structure of the Fucosylgalactoside alpha N-acetylgalactosaminyltransferase (GTA, cisAB mutant L266G, G268A) in complex with NPE caged UDP-Gal (C222(1) space group)	3ZGG	1.90	0.186	IUG
	Crystal structure of plasmodial PLP synthase with bound R5P intermediate	4ADU	2.44	0.196	R5P
14. Hexokinases	Rat Brain Hexokinase Type I Complex With Glucose And Inhibitor Glucose-6-Phosphate	1BG3	2.80	0.196	BGC
15. Lyase	Structure Determination And Refinement Of Ribulose 1,5 Bisphosphate Carboxylase(Slash)Oxygenase From Synechococcus Pcc6301	1RBL	2.20	0.200	CAP
	Structure of the PLP synthase Holoenzyme from Thermotoga maritima	2ISS	2.90	0.213	5RP
	Crystal Structure of E. coli LsrF in complex with Ribulose-5-phosphate	3GND	2.90	0.197	5RP
	Fructose-1,6-bisphosphate Schiff base intermediate in FBP aldolase from rabbit muscle	1ZAI	1.76	0.155	2FP
	Activated Spinach Rubisco In Complex With Its Substrate Ribulose-1,5-Bisphosphate And Calcium	1RXO	2.20	0.205	RUB
	Non-Activated Spinach Rubisco In Complex With	1RCX	2.40	0.224	RUB

	Its Substrate Ribulose-1,5-Bisphosphate				
16. Periplasmic Sugar Receptor	Crystal Structure Of D-Allose Binding Protein From Escherichia Coli	1RPJ	1.80	0.194	ALL
17. Immune System	Crystal structure of the complex of peptidoglycan recognition protein (PGRP-S) with N-acetylgalactosamine at 3.1 A resolution	3COR	3.10	0.208	NGA
18. Viral Protein	Crystal Structure of Putative Tailspike Protein (TSP1, orf210) from Escherichia coli O157:H7 Bacteriophage CBA120 in Complex with Lactose	4OJO	2.00	0.177	LBT
19. Lectin	Structural Basis Of Galactose Recognition In C-Type Animal Lectins	1AFA	2.00	0.221	MBG
	A Pro To Gly Mutation In The Hinge Of The Arabinose-Binding Protein Enhances Binding And Alters Specificity: Sugar-Binding And Crystallographic Studies	1BAP	1.75	0.201	ARB ,AR A
	X-Ray Structure Of A (Alpha-Man(1-3)Beta-Man(1-4)Glcnae)-Lectin Complex At 2.1 Angstroms Resolution	1LOG	2.10	0.175	NAG MAN
	Mannose-Specific Agglutinin (Lectin) From Snowdrop (Galanthus Nivalis) Bulbs In Complex With Mannose-Alpha 1,3-Methyl-D-Mannose	1NIV	3.00	0.223	MDM
20. Transcription Factor	Escherichia Coli Regulatory Protein Arac Complexed With L-Arabinose	2ARC	1.50	0.179	ARA
21. Sugar Transport	Probing Protein-Protein Interactions: The Ribose-Binding Protein In Bacterial Transport And Chemotaxis	1DRJ	2.50	0.176	RIP
22. Sugar Binding Protein	Peanut Lectin Complexed With N-Acetyllactosamine	1CIW	2.70	0.194	GAL
	Crystal structure of Jacalin complexed with 1-O-	1KUJ	2.00	0.191	MMA

	methyl-alpha-D-mannose				
	Crystal structure of the galectin-8 N-terminal carbohydrate recognition domain in complex with lactose	3AP4	2.33	0.194	LBT
	BEL beta-trefoil complex with N-acetylgalactosamine	4I4V	1.50	0.188	NGA
	Structure, interactions and evolutionary implications of a domain-swapped lectin dimer from Mycobacterium smegmatis	4OIZ	3.40	0.244	MMA
	Crystal structure of Canavalia boliviana lectin in complex with Man1-3Man-OMe	4K1Y	2.50	0.205	MDM
	Crystal Structure of Human Galectin-3 CRD in Complex with Type 1 N-acetyllactosamine	4XBN	2.21	0.168	NAG GAL
	Crystal Structure of Human Galectin-7 in Complex with Type 1 N-acetyllactosamine	4XBQ	2.23	0.197	NAG GAL
	Crystal structure of the human galectin-4 C-terminal carbohydrate recognition domain in complex with lactose	4YM3	1.89	0.169	LAT
	Crystal structure of the N terminal domain of human Galectin 8, F19Y mutant	4BME	2.00	0.197	LBT
	Crystal structure of AAL mutant H59Q complex with lactose	3VT0	2.00	0.200	LAT

	MAN(Aplha1-6)Man(Alpha1-O)Methyl Concanavalin A Complex	1QDC	2.00	0.175	MAN MMA
	Crystal structure of the Rhizoctonia solani agglutinin in complex with N'-acetyl-galactosamine	4G9N	2.20	0.154	NGA
	Crystal Structure of Rat Galectin-1 in Complex with Lactose	4GA9	1.88	0.218	LBT
	Structural basis of	2FMD	1.90	0.186	MAN

	carbohydrate recognition by Bowringia milbraedii seed agglutinin				
	Crystal structure of human galectin-3 carbohydrate-recognition domain with lactose bound, at 1.35 angstrom resolution	2NN8	1.35	0.165	LBT LAT
	MAN(Aplha1-3)Man(Alpha1-O)Methyl Concanavalin A Complex	1QDO	2.80	0.164	MMA MAN

2. List des Ligand

Le Tableau 05 représente l'ensemble des ligands glucidiques utilisée dans ce projet qui font des contacts avec des protéines de classes différentes.

Tableau 05: Liste des ligands en complexe avec les structures PDB.

Id (dans PDB)	Nom de ligand	Formule	Classes et Codes PDB
BGC	BETA-D-GLUCOSE	$C_6H_{12}O_6$ Voir image présenté dans (Figure 20)	Classe1 :1CQ1 ,2CDB,3VKV,3VPH 3PQD 1GPE Classe 4 : 3ACI Classe 10 : 1ZM1 Classe 14 : 1BG3
G6P	ALPHA-D-GLUCOSE-6-PHOSPHATE	$C_6H_{13}O_9P$ Voir image présenté dans (Figure 21)	Classe 4 : 1UP6 Classe 8 : 4QFH
NAG	N-ACETYL-D-GLUCOSAMINE	$C_8H_{15}NO_6$ Voir image présenté dans (Figure 22)	Classe : 1 1GPE Classe : 4 1HKI Classe : 9 9LYZ Classe : 19 1LOG Classe : 22 4XBN ,4XBQ
2FP	1,6-FRUCTOSE	$C_6H_{14}O_{12}P_2$	Classe : 9 3MMT Classe : 10 2QDG,1FDJ

	DIPHOSPHATE (LINEAR FORM)	Voir image présenté dans (Figure 23)	Classe : 15 1ZAI
FBP	BETA-FRUCTOSE-1,6-DIPHOSPHATE	$C_6 H_{14} O_{12} P_2$ Voir image présenté dans (Figure 24)	Classe 1 : 3VKV, 3VPH, 3PQD
F6P	FRUCTOSE-6-PHOSPHATE	$C_6 H_{13} O_9 P$ Voir image présenté dans (Figure 25)	Classe : 8 2CXS,1T10,1HOX
MDM	METHYL-O3-(ALPHA-D-MANNOSE)-ALPHA-D-MANNOSE	$C_{13} H_{24} O_{11}$ Voir image présenté dans (Figure 26)	Classe : 6 2EF6 Classe : 19 ANIV Classe : 22 4K1Y
MMA	O1-METHYL-MANNOSE	$C_7 H_{14} O_6$ Voir image présenté dans (Figure 27)	Classe : 5 4GX7 Classe : 22 1KUJ 4OIZ 1QDC 1QDO
MAN	ALPHA-D-MANNOSE	$C_6 H_{12} O_6$ Voir image présenté dans (Figure 28)	Classe : 19 1LOG Classe : 22 1QDC, 2FMD,1QDO
LBT	ALPHA-LACTOSE	$C_{12} H_{22} O_{11}$ Voir image présenté dans (Figure 29)	Classe : 18 4OJO Classe : 22 3AP4, 4BME ,4GA9, 2NN8
TDG	THIODIGALACTOSIDE	$C_{12} H_{22} O_{10} S$	Classe : 12 1PV7

		Voir image présenté dans (Figure 30)	
LAT	BETA-LACTOSE	$C_{12} H_{22} O_{11}$ Voir image présenté dans (Figure 31)	Classe : 4 1HKI Classe : 7 4NO4 Asse : 22 4YM3, 3VT0, 2NN8
R5P	RIBOSE-5-PHOSPHATE	$C_5 H_{11} O_8 P$ Voir image présenté dans (Figure 32)	Classe : 8 4M8L Classe : 13 1ZHA,4ADU
RIP	RIBOSE(PYRANOSE FORM)	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 33)	Classe : 2 1DRK Classe : 11 1OGD Classe : 21 1DRJ Classe : 8 4Q0P
OMK	L-RIBOPYRANOSE	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 34)	Classe : 8 4Q0P Classe : 9 3WW1,4DUX
GAL	BETA-D-GALACTOSE	$C_6 H_{12} O_6$ Voir image présenté dans (Figure 35)	Classe : 12 4LO1 Classe : 22 1CIW,4XBQ ,4XBN
NGA	BN-ACETYL-D- GALACTOSAMINEE	$C_8 H_{15} N O_6$ Voir image présenté dans (Figure 36)	Classe : 17 3COR Classe : 19 1LOG Classe : 22 4I4V, 4G9N
IUG	1-(2- NITROPHENYL)ETHYL UDP-GALACTOSE	$C_{22} H_{30} N_3 O_{19}$ P2(1+) Voir image présenté dans	Classe : 13 3ZGF 3ZGG

		(Figure 37)	
ARB	BETA-L-ARABINOSE	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 38)	Classe : 2 1ABE,6ABP Classe : 3 3TB6 Classe : 8 1MMZ,4QDP Classe : 19 1BAP
ARA	ALPHA-L-ARABINOSE	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 39)	Classe : 2 1ABE,6ABP Classe : 4 3A22 Classe : 12 4RXT Classe : 19 1BAP Classe : 20 2ARC
ALL	D-ALLOPYRANOSE	$C_6 H_{12} O_6$ Voir image présenté dans (Figure 40)	Classe : 16 1RPJ
AOS	D-ALLOSE	$C_6 H_{12} O_6$ Voir image présenté dans (Figure 41)	Classe : 8 3PH4, 2I57,4GJJ,3M0M
NA1	METHYL N-ACETYL ALLOSAMINE	$C_9 H_{17} NO_{16}$ Voir image présenté dans (Figure 42)	Classe : 9 3FY11HKJ
3BU	L-ALLOSE	$C_6 H_{12} O_6$ Voir image présenté dans (Figure 43)	Classe : 8 3WW4
A6P	6-O-PHOSPHONO-ALPHA- D-ALLOPYRANOSE	$C_6 H_{13} O_9$ Voir image présenté dans (Figure 44)	Classe : 8 2VVO
5RP	RIBULOSE-5-PHOSPHATE	$C_5 H_{11} O_8 P$	Classe : 8 2VVP

		Voir image présenté dans (Figure 45)	,1TKU 3OVQ,4M8L Classe : 10 4P77 Classe : 13 1ZHA Classe : 15 2ISS, 3GND
CAP	2- CARBOXYARABINITOL- 1,5-DIPHOSPHATE	$C_6 H_{14} O_{13} P_2$ Voir image présenté dans (Figure 46)	Classe : 15 1RBL
RUB	RIBULOSE-1,5- DIPHOSPHATE	$C_5 H_{12} O_{11} P_2$ Voir image présenté dans (Figure 47)	Classe : 15 1RXO ,1RCX
XLS	D-XYLOSE (LINEAR FORM)	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 48)	Classe : 8 3XIS,1XIC 8XIA,4XIS
XYS	XYLOPYRANOSE	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 49)	Classe 8 : 4XIS
XUL	D-XYLULOSE	$C_5 H_{10} O_5$ Voir image présenté dans (Figure 50)	Classe : 8 1XII
XYL	D-XYLITOL	$C_5 H_{12} O_5$ Voir image présenté dans (Figure 51)	Classe : 4 4GNA Classe : 8 3XIS, 1XLM 1XIG, 2XIS, 5XIA
MBG	METHYL-BETA- GALACTOSE	$C_7 H_{14} O_6$	Classe : 19 1AFA

		Voir image présenté dans (Figure 52)	
PAN	5-PHOSPHO-D- ARABINOHYDROXAMIC ACID	C ₅ H ₁₂ N O ₉ P Voir image présenté dans (Figure 53)	Classe : 8 1KOJ

3. Les images des ligands

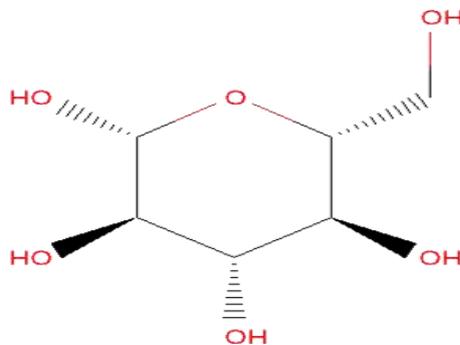


Figure 20 : La structure de la BGC.BETA-D-GLUCOSE

Source : <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/BGC>

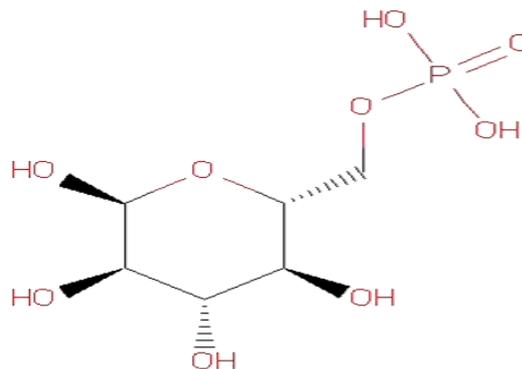


Figure 21 : La structure de la G6P. ALPHA-D-GLUCOSE-6-PHOSPHATE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/G6P>

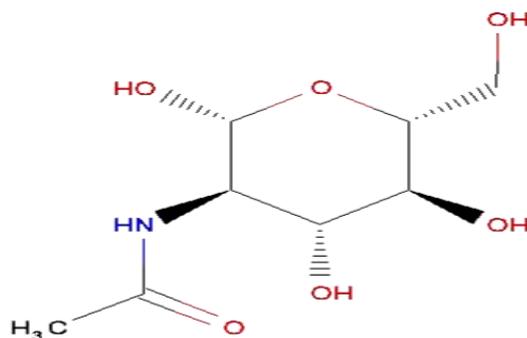


Figure 22 : La structure de la NAG. N-ACETYL-D-GLUCOSAMINE

Source : <http://www.ebi.ac.uk/pdbesrv/pdbechem/chemicalCompound/show/NAG>

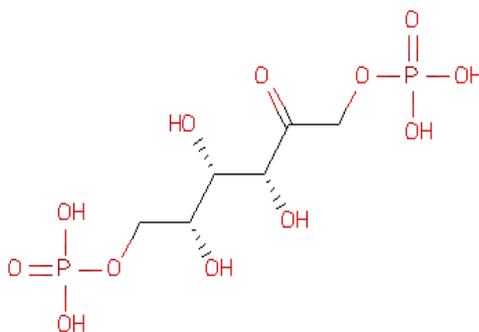


Figure 23 : La structure de la 2FP. 1,6-FRUCTOSE DIPHOSPHATE (LINEAR FORM)

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/2FP>

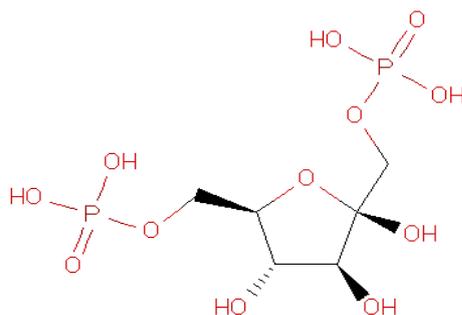


Figure 24 : La structure de la FBP. BETA-FRUCTOSE-1,6-DIPHOSPHATE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/FBP>

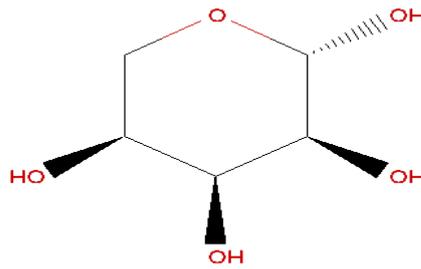


Figure 25 : La structure de la 0MK. L-RIBOPYRANOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/0MK>

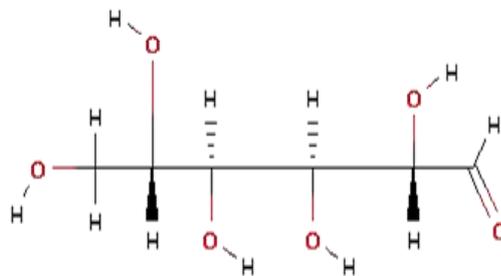


Figure 26 : La structure de la 3BU. L-ALLOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/3BU>

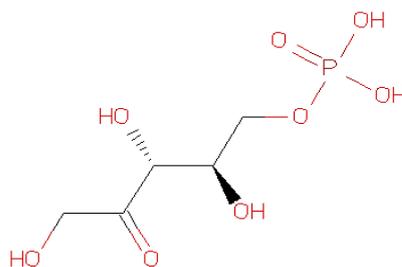


Figure 27 : La structure de la 5RP, RIBULOSE-1,5-DIPHOSPHATE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/5RP>

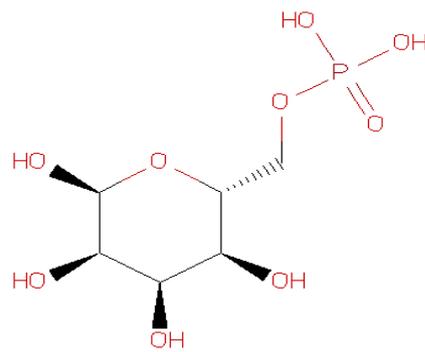


Figure 28 : La structure de la A6P,6-O-PHOSPHONO-ALPHA-D-ALLOPYRANOSE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/A6P>

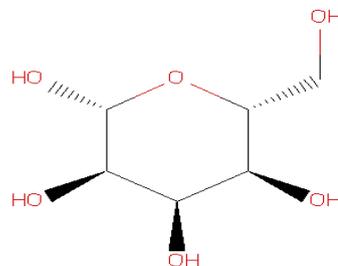


Figure 29 : La structure de la ALL,D-ALLOPYRANOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ALL>

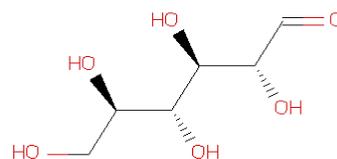


Figure 30 : La structure de la AOS,D-ALLOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/AOS>

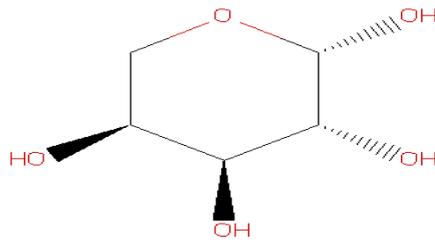


Figure 31 : La structure de la ARB, BETA-L-ARABINOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ARB>

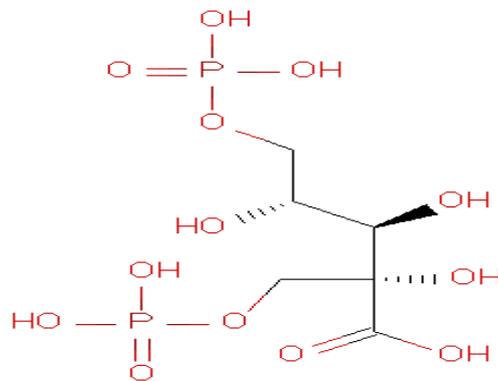


Figure 32 : La structure de la CAP, 2-CARBOXYARABINITOL-1,5-DIPHOSPHATE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/CAP>

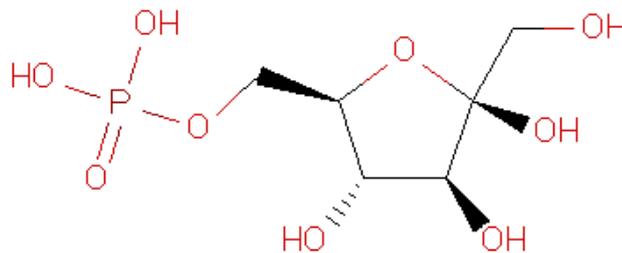


Figure 33 : La structure de la F6P, FRUCTOSE-6-PHOSPHATE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/F6P>

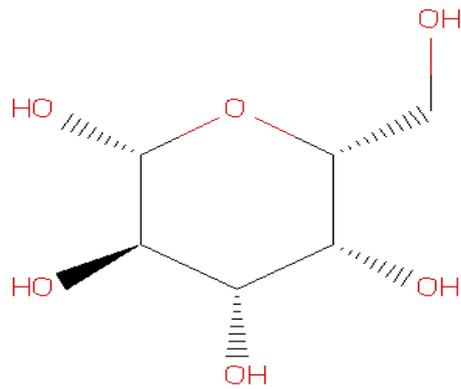


Figure 34 : La structure de la GAL, BETA-D-GALACTOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/GAL>

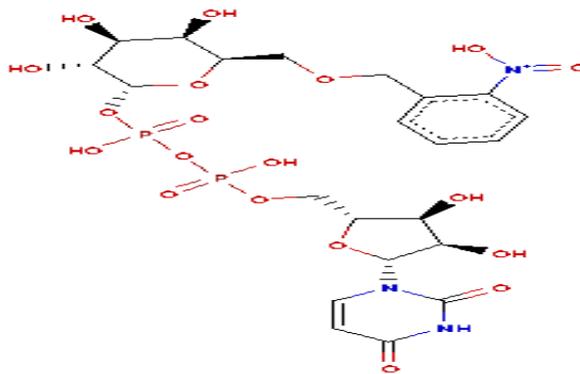


Figure 35 : La structure de la IUG, 1-(2-NITROPHENYL)ETHYL UDP-GALACTOSE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/IUG>

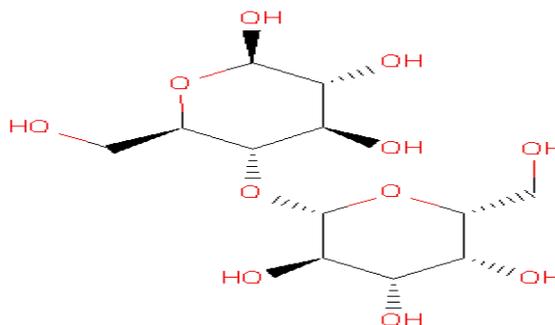


Figure 36 : La structure de la LAT, BETA-LACTOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LAT>

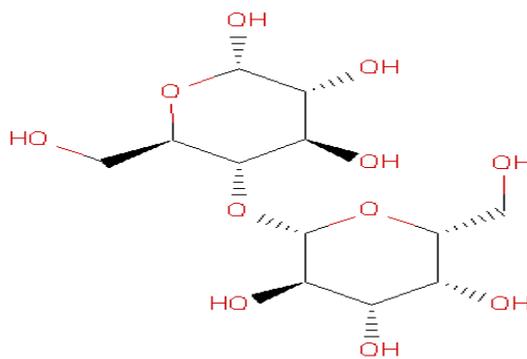


Figure 37 : La structure de la LBT, ALPHA-LACTOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LBT>

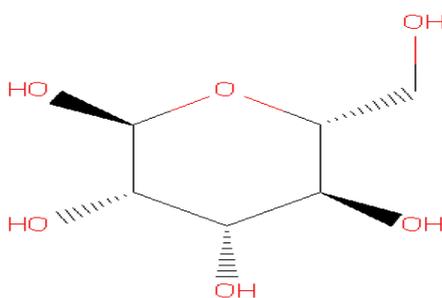


Figure 38 : La structure de la MAN, ALPHA-D-MANNOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MAN>

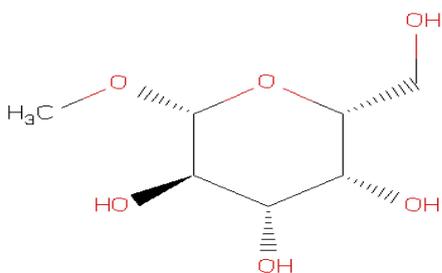


Figure 39 : La structure de la MBG, METHYL-BETA-GALACTOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MBG>

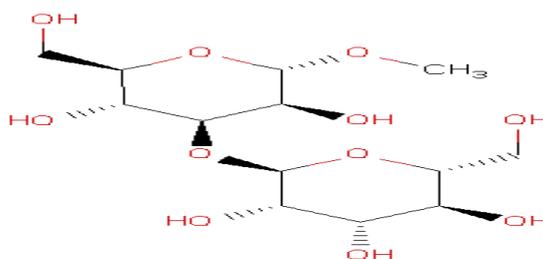


Figure 40 : La structure de la MDM, METHYL-O3-(ALPHA-D-MANNOSE)-ALPHA-D-MANNOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MDM>

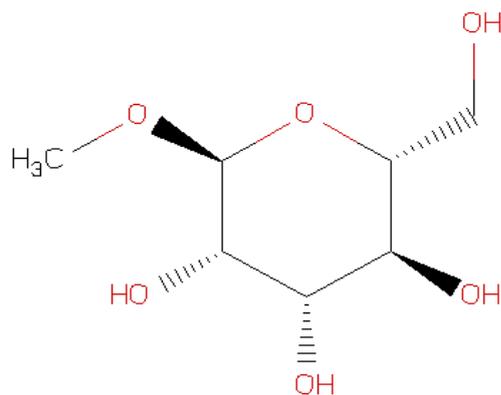


Figure 41 : La structure de la MMA, O1-METHYL-MANNOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MMA>

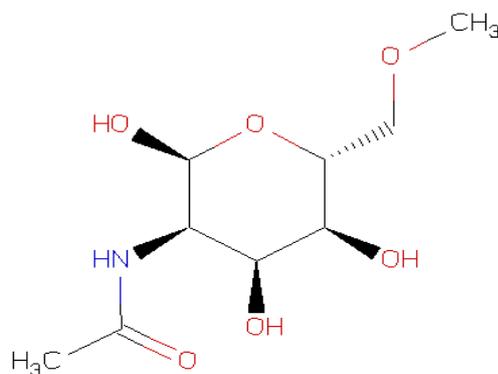


Figure 42 : La structure de la NA1, METHYL N-ACETYL ALLOSAMINE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/NA1>

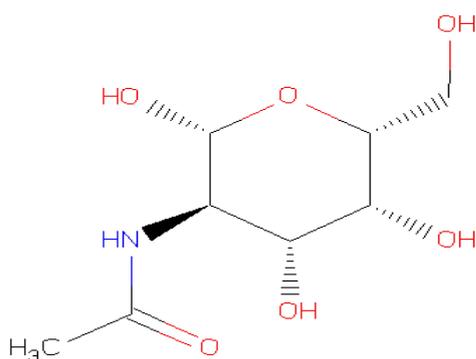


Figure 43 : La structure de la NGA, BN-ACETYL-D-GALACTOSAMINEE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/NGA>

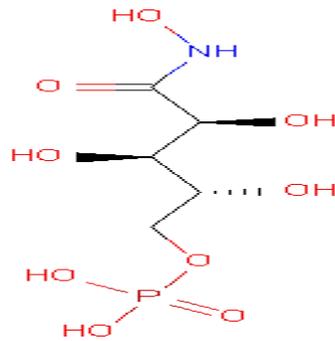


Figure 44 : La structure de la PAN, 5-PHOSPHO-D-ARABINOHYDROXAMIC ACID.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/PAN>

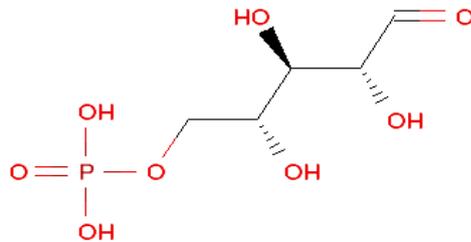


Figure 45 : La structure de la R5P, RIBOSE-5-PHOSPHATE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/R5P>

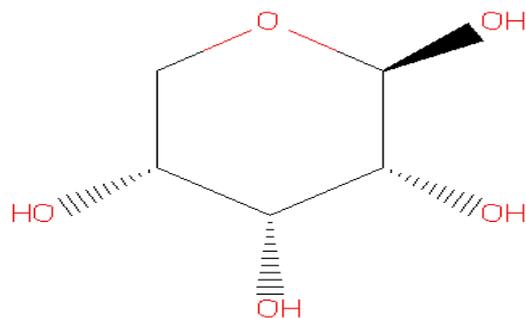


Figure 46 : La structure de la RIP, RIBOSE(PYRANOSE FORM)

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/RIP>

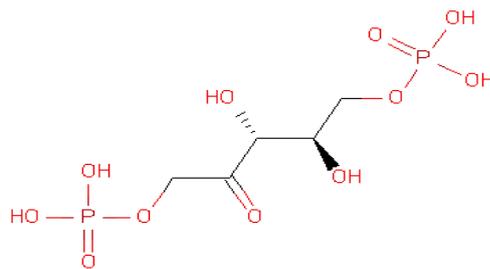


Figure 47 : La structure de la RUB, RIBULOSE-1,5-DIPHOSPHATE

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/RUB>

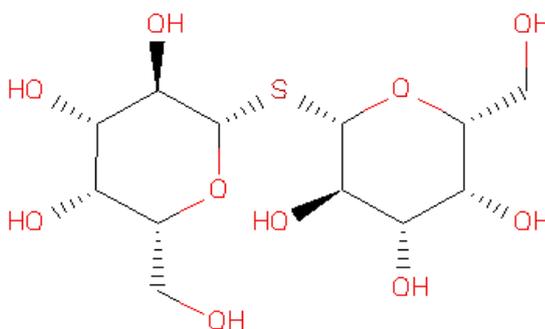


Figure 48 : La structure de la TDG, THIODIGALACTOSIDE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/TDG>

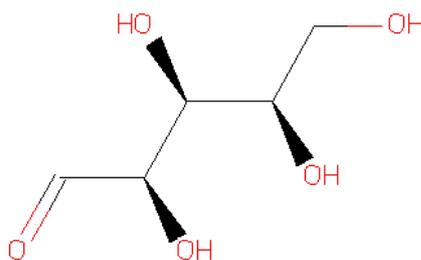


Figure 49 : La structure de la XLS, D-XYLOSE (LINEAR FORM).

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XLS>

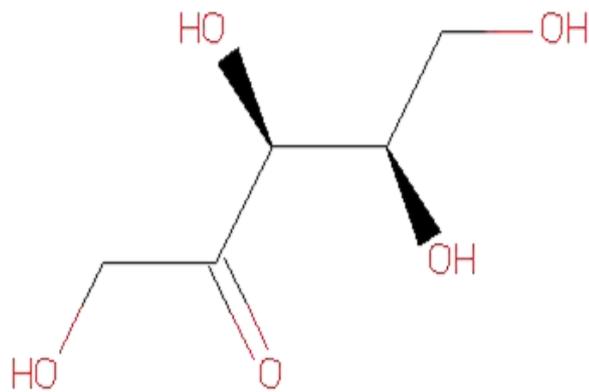


Figure 50 : La structure de la XUL, D-XYLULOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XUL>

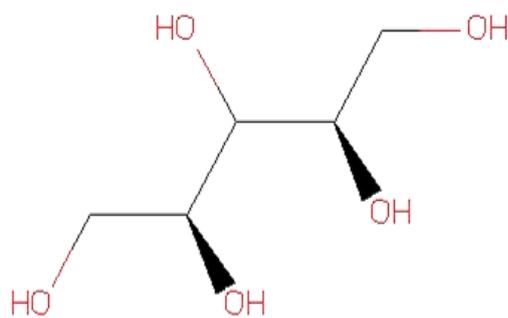


Figure 51 : La structure de la XYL, D-XYLITOL.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XYL>

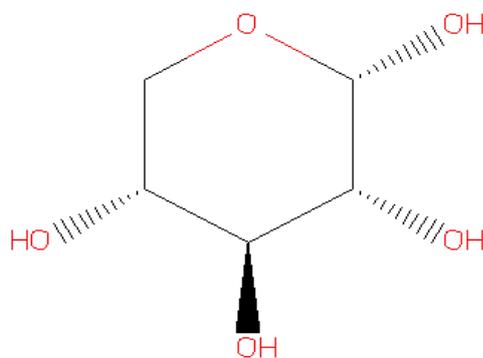


Figure 52 : La structure de la XYS, XYLOPYRANOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XYS>

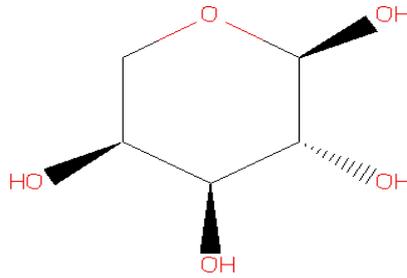


Figure 53 : La structure de la ARA, ALPHA-L-ARABINOSE.

Source: <http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ARA>

Chapitre III : matériels et méthodes

Ce projet est fait pour accomplir une étude structurale de l'environnement du site de liaison dans un ensemble des enzymes et protéines qui font des liaisons avec des glucides choisissant parmi la PDB.

Cette étude a nécessité de la programmation pour:

- La création d'une base de données.
- Faire les études et partager les résultats avec la communauté scientifique et universitaire.

Pour la réalisation de la base de données, les démarches suivantes sont optées dans la partie principale

- La sélection de la liste des codes PDB (représenter les protéines/enzymes) et la liste des ligands (représenter les Glucides). Voir les codes PDB et les Ligands dans le Chapitre 2
- L'utilisation d'une exploration de données (datamining) pour collecter les informations de PDB par les codes PDB et les ligands en utilisant le système SSFS (voir les détails ci-dessous) et scripts PHP écrits pour ce but
- La création des scripts en langage PHP et SQL qui permettent la création, l'accès et la recherche dans la base de données.
- La création de la base de données et lui intégrer les données collectées après avoir été traitées et analysées.
- Création d'une interface HTML pour faciliter la recherche et l'exploration de la base de données et pour afficher les résultats d'une manière qui est biologiquement pertinente.

1. Sélection de la liste des codes PDB et la liste des ligands:

Comme on a vu dans le chapitre précédent 100 codes PDB sont choisis parmi la PDB ces codes sont stockés dans un fichier nommé «LISTEPDB.txt».

Ces protéines font des contacts avec des 34 ligands glucidiques:

BGC, G6P, NAG, 2FP, FBP, F6P, MDM, MMA, MAN, LBT; TDG, LAT, R5P, RIP, 0MK, GAL, NGA, IUG, ARB, ARA, ALL, AOS, NA1, 3BU, A6P, 5RP, CAP, RUB, XUL, XYL, MBG, PAN, XLS, XYS

Stockés dans un fichier nommé «LISTELIGANDS.txt»

Pour les détails voir l'Annexe 2 (Section 1)

2. Collection (Datamining) des détails de calcul de l'environnement de liaison des ligands:

Cette section explique la technique de DataMining utilisé pour le calcul automatique et de collecte de données détaillées représentant l'environnement de liaison de ligand calculée pour les ligands dans les entrées pdb sélectionnés.

Le système SSFS ou Séquence, Structure and Function Server (**Abdelkrim Rachedi**) (<http://www.bioinformaticstools.org/ssfs/>) a été utilisé dans le calcul de l'environnement de liaison de chaque ligand trouvé dans chacune des structures PDB sélectionnés.

3 .Création de la base de données:

Dans ce projet un base de donnée a été crée sous le nom Glucides binding Structural & Functional Motifs (les motifs structurels et fonctionnels qui se lient des Glucides), nom d'abréviation "**GlcSFMs** database".

La création de la base de données « **GlcSFMs** » a passé par plusieurs étapes répartie en deux plates-formes une durant la période de développement (locale), et l'autre lors de production (web).

I. Plateforme Développement (local):

La base de données « **GlcSFMs** » a été crée par l'utilisation d'un système d'exploitation Windows 7 installée sur un ordinateur PC

I.1 Langage de programmation :

I.1.1. PHP (Hypertext Preprocessor):

PHP est un langage de programmation et est librement accessible et utilisé principalement sur Linux serveurs Web. PHP, à l'origine dérivé de pages d'accueil personnelles Outils, représente désormais PHP: Hypertext Preprocessor, dont le PHP FAQ décrit comme un «acronyme récursif."

PHP est une alternative à Microsoft de page ASP (ASP) de la technologie. Comme avec l'ASP, le script PHP est intégré dans une page Web avec son HTML. Avant la page est envoyé à un utilisateur qui l'a demandée, le serveur Web appelle PHP pour interpréter et exécuter les opérations prévues dans le script PHP.

I.1.2. Structured Query Language – SQL:

SQL est un langage de programmation standard pour les systèmes de gestion de base de données relationnelles tel que la **GlcSFMs**, voir section 2.2.2.5. **Le lien entre les tableaux.**

SQL a été utilisé sous la forme des instructions de requête dans les scripts PHP utilisés pour insérer les données dans la base de données et à extraire des résultats de celle-ci, voir ci-dessous.

I.1.3 HTML et CSS:

Les outils de développement web : HTML (HyperText Markup Language), CSS (Cascade Style Sheet) et même javas script seront aussi utilisés pour réaliser les interfaces du site.

Une page HTML qui inclut un script PHP est généralement donné un nom de fichier suffixe ".php" ".php3," ou ".phtml". Comme ASP, PHP peut être considéré comme "des pages HTML dynamiques», puisque le contenu variera en fonction des résultats de l'interprétation du script (searchenterpriselinix.techtarget.com/definition/PHP).

I.2. Bases de données MySQL:

MySQL est un système de gestion de bases de données relationnelles (SGBDR). Il est distribué sous une double licence GPL et propriétaire . Il fait partie des logiciels de gestion de base de données les plus utilisés au monde, autant par le grand public (applications web principalement) que par des professionnels, en concurrence avec Oracle, Informix et Microsoft SQL Server.

Son nom vient du prénom de la fille du cocréateur Michael Widenius, My. SQL fait allusion au Structured Query Language, le langage de requête utilisé.

I.3. Apache serveur:

XAMPP (www.apachefriends.org), une version d'Apache, est un ensemble de logiciels permettant de mettre en place facilement un serveur Web en gèrent l'ensemble de modules y compris développeur HTML, interpréteur de PHP, MySQL sous la version connue come phpMyAdmin et autres modules qui ne sont pas utiles dans ce projet. (Voir annexe 2 Section 2).

II. Création de la base de données « GicSFMs»:

Des scripts PHP sont à l' origine de

- La création de la base de donnée « **GicSFMs**» et l'insertion des données
- L'affichage des données sur une interface graphique facile et accessible par les utilisateurs

II.1.Les scripts PHP et SQL utilisée dans ce projet

II.1.1. Les scripts de la création de la base de données :

II.1.1.i Le premier script est « DB_connection.php » :

C'est le script qui fait la connexion a MySQL voir ci-dessous

```
1 <?php
2 $host="localhost";
3 $user="root";
4 $password="";
5 mysql_connect("$host", "$user", "$password") or die(mysql_error());
6
7 ?>
```

Figure 54: Capture d'écran représente le script « DB_connection.php »

II. 1.1.ii Le script de création « createDB.inc.php » :

C'est qui fait la création de la base de données

```
mysql_query("create database IF NOT EXISTS " . $dbname. " ") or die(mysql_error());

mysql_select_db("$dbname") or die ("unable to select dbase"); // Connexion a la base de donnée
mysql_query("CREATE TABLE IF NOT EXISTS `pdb_data` (
  `pdbk` int(11) NOT NULL AUTO_INCREMENT,
  `id_pdb` varchar(50) NOT NULL,
  `titre_pdp` varchar(50) NOT NULL,
  `determination_methode` varchar(50) NOT NULL,
  `resolution` varchar(10) NOT NULL,
  `r_factor` varchar(10) NOT NULL,
  PRIMARY KEY (`pdbk`))");
mysql_query("CREATE TABLE IF NOT EXISTS `ligand` (
  `ldpk` int(11) NOT NULL auto_increment,
```

Figure 55 : Capture d'écran de script « createDB.inc .PHP»

II.1.2.Les scripts de l'insertion des données dans la base de données :**II.1.2.i le script « insert_pdb_list.php » :**

Le script « insert_pdb_list.php » crée pour l'insertion des données a partir de fichier « LISTEPDB.txt » dans le tableau « pdb_list » voir ci-dessous

```

for($i=0;$i<count($val);$i+=1){
    $requete1 = " select * from pdb_data where    id_pdb ='".$val[$i][0]."'";
    $resultat1 = mysql_query($requete1);
    $testexiste = mysql_fetch_array($resultat1);

    if(isset($testexiste['id_pdb']))
    {
        ECHO "ID_PDP : ' ".$val[$i][0]."' est déjà enregistré dans la database <br>";
    } else{

        $reqinsert= mysql_query("INSERT INTO pdb_data (id_pdb , titre_pdp , determination
VALUES ('".$val[$i][0]."' , '".$addslashes($val[$i][1])."' , '".$val[$i][2]."' ,
") or die ("mysql Error read".mysql_error());
        echo '<br />';
        prt_arr($val[$i]);
    }
}

function prt_arr($myarray){
//-----
echo "<pre>";
print_r($myarray);
echo "</pre>";
}

```

Figure 56 : Capture d'écran de script « insert_pdb_list.php »

II.1.2.ii Le script « insert_ligands_list.php » :

C'est le script de l'insertion des données à partir de fichier

« LISTELIGANDS.txt » dans le tableau « ligands » voir ci-dessous

```

for($i=0;$i<count($val);$i+=1){
    $requete1 = " select * from ligand where    lig_id ='".$val[$i][0]."'";
    $resultat1 = mysql_query($requete1);
    $testexiste = mysql_fetch_array($resultat1);

    if(isset($testexiste['lig_id']))
    {
        ECHO "lig_id : ' ".$val[$i][0]."' est déjà enregistré dans la databa
    } else{
        $reqinsert= mysql_query("INSERT INTO ligand (lig_id, lig_name, lig_fo
VALUES ('".$val[$i][0]."' , '".$addslashes($val[$i][1])."' , '".$val[$i]
    }
}

function prt_arr($myarray){
//-----
echo "<pre>";
print_r($myarray);
echo "</pre>";
}
?>
</body>
</html>

```

Figure 57 : Capture d'écran représente le script « insert_ligands_list.php »

II.1.2.iii Le script « insert_details.php » :

Ce script est crée pour l'insertion des détails de contacts dans le tableau « prot_lig_contacts »

Son insertion est fait d'un manière automatique a partir de site web www.bioinformaticstools.org/prjs/lgb/lgb.php?

```

1  <?PHP
2  // création de data_base et connection à la base de donnée
3  require_once("createDB.inc.php");
4  mysql_select_db("$dbname") or die ("unable to select dbase");
5
6  // this is needed especially when adding new entries ----- see $jump below -----
7  $pdpk_var= mysql_query(" select distinct pdpk from prot_lig_contact");
8  $pdpks=array ();
9  while ($pdpk_val = mysql_fetch_array($pdpk_var) {$pdpks[] = $pdpk_val['pdpk'];}
10 // $pdpks=array_values(array_unique($pdpks));
11 //prt_arr($pdpks);

```

Figure 58 : Capture d'écran représentatif de script « fetch_insert_lgct_dtls.php »

II.1.2.vi. Script « gnr_rswn_script .php» :

C'est le script qui fait la régénération du nom et le fichier de l'image

Le fichier de l'image peut exécuter par le Rasmol pour plus d'informations sur le Rasmol voir (annexe 2 : Section 2)

```

$idpdbn= mysql_query(" select id_pdb from pdb_data where pdpk ='".trim($sourc[0])."' ");
$codpdb = mysql_fetch_array($idpdbn);
$title= "zap\nload ../pdb_entries/".$codpdb['id_pdb'].".pdb\n set ambient 60\nset speccpower 25\n s
$fin= "select ".trim($sourc[5]).":".trim($sourc[6])."\n spacefill 100 \n wireframe 30 \n color cpk\n#c
// $lgdots=trim($sourc[5]).":".trim($sourc[6]);
$avfin= "#select :".trim($sourc[1])."\n #backbone 10\n #color structure \n zoom 250\n #set Backgr

if ($sourc[2]=="No SSE"){

    // $dots.=trim($sourc[3]).":".trim($sourc[1]).",";
    $Wderv.= trim($sourc[3]).":".trim($sourc[1]).",";

    $rasmol="select ".trim($sourc[3]).":".trim($sourc[1])."\n";
    $rasmol.="wireframe 90 \n";
    // $rasmol.="spacefill 120 \n";
    $rasmol.="color cpk \n";

    // This addition has been made to create a small grey[PCP] "loop" around each of the

```

Figure 59 : Capture d'écran représente les détails de script « gnr_rswn_script .php»

II.1.2.v Le script « insertion_des_images.php» :

Ceci est le script qui fait l'insertion des images dans le tableau « file_images»

```

$fil0=$codpdb['id_pdb']."_".trim($sourc[0])."_".trim($sourc[1])."_".trim($sourc[6])."_".trim($so
$flsrc = @fopen($fil0, "r");
$donnees = fread($flsrc,filesize($fil0));
$VAR=addslashes($donnees);
$titre = $codpdb['id_pdb']."_".trim($sourc[0])."_".trim($sourc[1])."_".trim($sourc[6])."_".trim(
$type = "GIF";
//echo"$titre".<br>.$type.<br>.$donnees.<br>.<br>";
echo "<img src=\"data:image/gif;base64,\" . base64_encode($fil0).\" height=\"200\" width=\"250\"
$reqinsert= mysql_query("INSERT INTO file_image ( pdbk, name, type, data )
VALUES (\".trim($sourc[0]).\", \".$titre.\", \".$type.\", \".$VAR.\")" ) or die ("mysql Erreur reqt".my

$fill=$codpdb['id_pdb']."_".trim($sourc[0])."_".trim($sourc[1])."_".trim($sourc[6])."_".trim($so
$flsrc = @fopen($fill, "r");
$donnees = fread($flsrc,filesize($fill));
$VAR=addslashes($donnees);
$titre = $codpdb['id_pdb']."_".trim($sourc[0])."_".trim($sourc[1])."_".trim($sourc[6])."_".trim(
$type = "GIF";
//echo"$titre".<br>.$type.<br>.$donnees.<br>.<br>";

```

Figure 60 : Capture d'écran représente le script « insertion_des_images.php »

II.1.3 Les scripts de l'affichage des données local sur un navigateur

II.1.3.i Le script « menu.php » :

Ce script utilise le langage php et html pour afficher les donnée sur un navigateur internet (chrome ,mozilla ,internet explorer ...etc.)

```

42 <?php
43     require_once("createDB.inc.php");
44     $sql = " select * from ligand where 1";
45     $sql1 = " select * from pdb_data where 1";
46     .....
47     $req = mysql_query($sql) or die('Erreur SQL !<br>'.$sql.'<br>'.mysql_error());
48     $req1 = mysql_query($sql1) or die('Erreur SQL !<br>'.$sql1.'<br>'.mysql_error());
49     $req2 =" SELECT * FROM prot_lig_contact " ;
50     .....
51     // echo " <select id=\"lg\" name=\"ligand\">
52     echo " <select id=\"lg\" name=\"ligand\">
53         <option value=\"\"> </option>";
54     while($resultat=mysql_fetch_assoc($req) {
55         echo "<option value=\"\" $resultat[lig_id1] \"$\" $resultat[lig_id1] </option>";

```

Figure 61 : Capture d'écran représente le script « menu.php »

II.1.3.ii Le script « query.php » :

C'est un script pour afficher le résultat sur un navigateur web

```

17 <nz>coenzym Ligands Binding Motifs -- GLBMS</nz>
18 <?php
19 require_once("createDB.inc.php");
20 mysql_select_db("$dbname") or die ("unable to select dbase");
21
22 echo"<center>REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE <br>
23     MINISTERE DE L'ENSEGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE<br>
24     Université Dr.Moulay Tahar <br>Faculté de BIOLOGIE<br></center>";
25
26 $switche=$_POST["switch"] ;
27

```

Figure 62 : Capture d'écran représente le script « query.php »

II.2. La création des fichiers et les tables :

II.2.1. Création des fichiers :

Les fichiers sont écrits sous Notepad++

Pour plus d'informations sur Notepad++ voir (annexe 3)

II.2.1.1. Le fichier « LISTEPDB.txt » :

Les informations stocker dans ce fichier sont extraite de site www.pdb.org .

Ce fichier contient le code PDB le nom de PDB, méthode, résolution, R_factor

```

4BZG#Crystal structure of galactose mutarotase GalM from Bacillus subtilis in complex with malto
1T3Z#Formyl-CoA Transferase mutant Asp169 to Ser#X-RAY#2.30#0.179
1YQZ#Structure of Coenzyme A-Disulfide Reductase from Staphylococcus aureus refined#X-RAY#1.54#0
2Q1L#Design and Synthesis of Pyrrole-based, Hepatoselective HMG-CoA Reductase Inhibitors#X-RAY#2
2vjk#FORMYL-COA TRANSFERASE WITH ASPARTYL-COA THIOESTER INTERMEDIATE DERIVED FROM OXALYL-COA#P
3rba#Phosphopantetheine adenylyltransferase from Mycobacterium tuberculosis complexed with DPCoA
1YQ4#Avian respiratory complex ii with 3-nitropropionate and ubiquinone#X-RAY#2.33#0.202
1Z0Y#Crystal Structure of Mitochondrial Respiratory Complex II from porcine heart#X-RAY#2.40#0.2
3CWB#Chicken Cytochrome BC1 Complex inhibited by an iodinated analogue of the polyketide Crocaci
3H1H#Cytochrome bcl complex from chicken#X_RAY#3.16#0.253
3L73#Cytochrome BC1 complex from chicken with triazolone inhibitor#X-RAY#3.04#0.260
1oqc#The crystal structure of augmenter of liver regeneration: a mammalian FAD dependent sulfhyd
2f5v#Reaction geometry and thermostability mutant of pyranose 2-oxidase from the white-rot fungu
2f6c#Reaction geometry and thermostability of pyranose 2-oxidase from the white-rot fungus Penic
2i0z#Crystal structure of a FAD binding protein from Bacillus cereus, a putative NAD(FAD)-utiliz
3mbq#Crystal Structure of Human Augmenter of Liver Regeneration (ALR)#X-RAY#1.85#0.191

```

Figure 63 : Capture d'écran représentatif du fichier « LISTEPDB.txt » sous Notepad++. Voir (annexe 2 Section 1 (1))

II.2.1.2. Le fichier « LISTELIGANDS.txt»:

Ce fichier est écrit sous le Notepad++ et il contient les informations important sur les ligands (le code de ligand, nom de ligand, structure, site pour plus d'information sur le ligand)

```

8 MMA#01-METHYL-MANNOSE#C7 H14 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MMA
9 MAN#ALPHA-D-MANNOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MAN
10 LBT#ALPHA-LACTOSE#C12 H22 O11#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LBT
11 TDG#THIODIGALACTOSIDE#C12 H22 O10 S#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/TDG
12 LAT#BETA-LACTOSE#C12 H22 O11#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LAT
13 R5P#RIBOSE-5-PHOSPHATE#C5 H11 O8 P#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/R5P
14 RIP#RIBOSE (PYRANOSE FORM)#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/RIP
15 OMK#L-RIBOPYRANOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/OMK

```

Figure 64 : Capture d'écran représentatif du fichier « LISTELIGANDS.txt » sous Notepad++ Voir (annexe 2 Section 1 (2)).

II.2.1.3. Le fichier Ras Win :

Les fichiers «Ras Win» est crée par le script « gnr_rswn_script.php » Après l'exécution de script PHP on obtient les fichiers Ras Win

Ces fichiers sont l'origine de trois images (image _1, image_2, image_S)

Exemple :

Le fichier 1ZAI_12_D_D_3004

Sera Enregistrer sous les nomes suivants

1ZAI_12_D_D_3004_0

1ZAI_12_D_D_3004_1

1ZAI_12_D_D_3004_S (qui est sous forme stéréo)

II.2.2. Création des tableaux :

Dans le période de la création de la base de donnée «GlcSFMs»

4 tableaux sont crée pour le stockage des informations

Table	Action	Lignes	Type	Interclassement	Taille	Perte
file_image	Afficher Structure Rechercher Insérer Vider Supprimer	147	InnoDB	latin1_swedish_ci	3,5 KiB	-
ligand	Afficher Structure Rechercher Insérer Vider Supprimer	24	InnoDB	latin1_swedish_ci	16 KiB	-
pdb_data	Afficher Structure Rechercher Insérer Vider Supprimer	100	InnoDB	latin1_swedish_ci	16 KiB	-
prot_lig_contact	Afficher Structure Rechercher Insérer Vider Supprimer	18 161	InnoDB	latin1_swedish_ci	3,1 MiB	-
4 tables	Somme	18 442	InnoDB	latin1_swedish_ci	6,7 MiB	0 0

Figure 65 : Capture d'écran représentatif des tableaux de la base de données sous phpMyAdmi.

II.2.2.1. Structure de tableau «file_image» :

Comme en a vue précédemment l'insertion des données dans le tableau est fait par le script « insertion_des_images.php »

	idimg	pdpk	name	type	updated_date	data
Modifier Copier Effacer	1	1	2cdb_1_A_A_1368_0	GIF	2015-09-25 18:13:07	[BLOB - 9,1 Kio]
Modifier Copier Effacer	2	1	2cdb_1_A_A_1368_1	GIF	2015-09-25 18:13:07	[BLOB - 9,2 Kio]
Modifier Copier Effacer	3	1	2cdb_1_A_A_1368_S	GIF	2015-09-25 18:13:07	[BLOB - 16,2 Kio]
Modifier Copier Effacer	4	1	2cdb_1_B_B_1368_0	GIF	2015-09-25 18:13:07	[BLOB - 8,4 Kio]
Modifier Copier Effacer	5	1	2cdb_1_B_B_1368_1	GIF	2015-09-25 18:13:08	[BLOB - 8,8 Kio]
Modifier Copier Effacer	6	1	2cdb_1_B_B_1368_S	GIF	2015-09-25 18:13:08	[BLOB - 15 Kio]

Figure 66 : Capture d'écran représentatif de la structure de tableau file_imege sous phpMyAdmin

Le tableau file_image est composé de 06 colonnes :

La colonne «idimg» : représente le numéro d'identification pour chacun des images.

La colonne «pdpk» : le numéro d'identification pour les pdb Entries.

La colonne «name» : les nomes des images.

La colonne «type» : le type D'image ici est "GIF".

La colonne «upload_date» : la date de l'insertion d'image.

La colonne «data» : lieu de stockage de l'image.

II.2.2.2 .Structure de tableau «pdb_data» :

L'insertion des données dans ce tableau est faite par le script « insert_pdb.php »

En utilisant le fichier « list_pdb.txt » .

	pdpk	id_pdb	titre_pdp	determination_methode	resolution	r_factor
Modifier Copier Effacer	1	2:cd	SULFOLOBUS SOLFATARICUS GLUCOSE DEHYDROGENASE 1 IN	X-RAY	1.60	0.192
Modifier Copier Effacer	2	1:c:1	Soluble Quinoprotein Glucose Dehydrogenase from Ac	X-RAY	1.90	0.190
Modifier Copier Effacer	3	1up6	STRUCTURE OF THE 6-PHOSPHO-BETA GLUCOSIDASE FROM T	X-RAY	2.55	0.202
Modifier Copier Effacer	4	3aci	Crystal Structure of Carbohydrate-Binding Module F	X-RAY	1.60	0.144
Modifier Copier Effacer	5	4qfn	Structure of a glucose-6-phosphate isomerase from	X-RAY	1.80	0.140

Figure 67 : Capture d'écran représentatif de la structure de tableau pdb_data sous phpMyAdmin.

Le tableau `pdb_data` est composé de 06 colonnes :

La colonne «`pdpk`» : le numéro d'identification pour les `pdb` Entries.

La colonne «`id_pdb`» : `pdb` structure id.

La colonne «`name`» : le titre de `pdb` entry.

La colonne «`determination_methode`» : la méthode de détermination utilisée qui est `x_ray`

La colonne «`resolution`» : la distance correspondant à la plus petite caractéristique observable.

La colonne «`r_factor`» : une mesure de l'accord entre le modèle cristallographique et les données d'origine de diffraction des rayons X.

II.2.2.3. Structure de tableau «ligands» :

L'insertion des données dans le tableau « ligand » fait par le script « `insert_ligands_liste.php` » en utilisant le fichier « `list_ligand.txt` »

	idpk	lig_id	lig_name	lig_formul	lig_chemistry
	1	BGC	BETA-D-GLUCOSE	C6 H12 O6	http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCom
	2	G6P	ALPHA-D-GLUCOSE-6-PHOSPHATE	C6 H13 O9 P	http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCom
	3	NAG	N-ACETYL-D-GLUCOSAMINE	C8 H15 N O6	http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCom
	4	2FP	1,6-FRUCTOSE DIPHOSPHATE (LINEAR FORM)	C6 H14 O12 P2	http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCom
	5	FBP	BETA-FRUCTOSE-1,6-DIPHOSPHATE	C6 H14 O12 P2	http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCom

Figure 68 : Capture d'écran représentatif du tableau de ligands sous phpMyAdmin.

Le tableau `file_image` est composé de 05 colonnes :

La colonne «`idpk`» : Le numéro d'identification de ligand.

La colonne «`lig_id`» : représente le triple code de ligands utilisée.

La colonne «`lig_name`» : le nom de ligand.

La colonne «`lig_formul`» : la structure de ligand.

La colonne «`lig_chemistry`» site pour savoir plus sur le ligand.

II.2.2.4. Structure du tableau «prot_lig_contact» :

L'insertion des données dans le tableau «prot_lig_contact » est fait automatiquement à partir du site web : www.bioinformaticstools.org/prjs/lgb.php?qry par le script « fetch_insert_lgct_dtls.php » à l'aide d'un parseur d'html appeler DOM parser (voir annexe 2 Section 2)

	id	pdpk	pdb_ch	res_sselmt	res_nb	res_nm	res_at	ldpk	lig_at	lig_nb	lig_ch	bond_length	bond_type
Modifier Copier Effacer	1	1 A	39-47 H.1	39	CYS	SG	1 O1	1368	A	3.85			
Modifier Copier Effacer	2	1 A	39-47 H.1	41	ALA	CB	1 O1	1368	A	3.71	van der Waals		
Modifier Copier Effacer	3	1 A	No SSE	66	HIS	NE2	1 O1	1368	A	3.5	HBond		
Modifier Copier Effacer	4	1 A	No SSE	89	ASN	OG	1 O3	1368	A	3.93	van der Waals		
Modifier Copier Effacer	5	1 A	No SSE	89	ASN	OD1	1 O3	1368	A	3.9	HBond		

Figure 69 : Capture d'écran représentatif de la structure de tableau prot_lig_contacts sous phpMyAdmin

Le tableau file_image est composé de 13 colonnes:

La colonne «id» : numéro d'identification.

La colonne «pdpk» : le numéro d'identification pour les pdb Entries.

La colonne «pdb_ch» : la chaine de résidu.

La colonne «res_sselmt» : la structure secondaire de résidu.

La colonne «res_nb» : le numéro de résidu.

La colonne «res_nm» : le nom de résidu.

La colonne «res_atm» : l'atome de résidu.

La colonne «ldpk» :Le numéro d'identification de ligand.

La colonne «lig_at» : l'atome de ligand.

La colonne «lig_nb» : le numéro de ligand.

La colonne «lig_ch» : la chaine de ligand.

La colonne «bond_length» : le longueur de la liaison.

La colonne «bond_type» : le type de la liaison.

II.2.2.5. Le lien entre les tableaux :

La base de données **GlcSFMs** est du type connu comme **base de données relationnelle**, dans lequel les tables composent sont liés les uns aux autres d'une manière qui rend les données dans les différents tableaux correctement lié, efficacement modifiable et correctement en mise à jour.

La figure n°36 indique que les 4 tableaux de la base de données sont reliés avec eux par des clés primaire et secondaire.

Les liens sont instrumental dans la préservation de l'intégrité des donnée lors de l'insertion et pour facilite leurs mise à jour avec une fidélité complète.

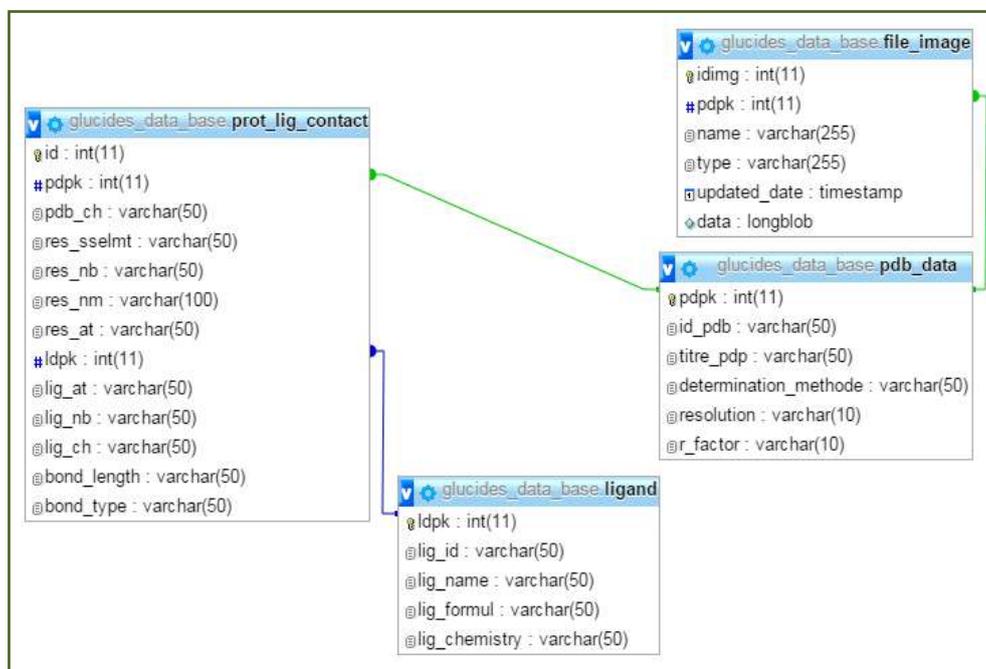


Figure 70: Capture d'écran représentatif le lien entre les tableaux

II.3. Voie Générale sur le Processus de la création de la base de données «GLcSFMs» et l'interface pour la recherche

La figure 37 présente les étapes essentielles de la création de la base de données et la création de l'interface de recherche par l'utilisation des données détaillées et calculé pour l'environnement de site de liaison, les images utilisé et même les scripts PHP et SQL qui font :

- La création de la base de données «GLcSFMs» et ces tableaux
- L'insertion des données
- Et l'affichage de ces données sur une interface de recherche accessible pour les utilisateurs.

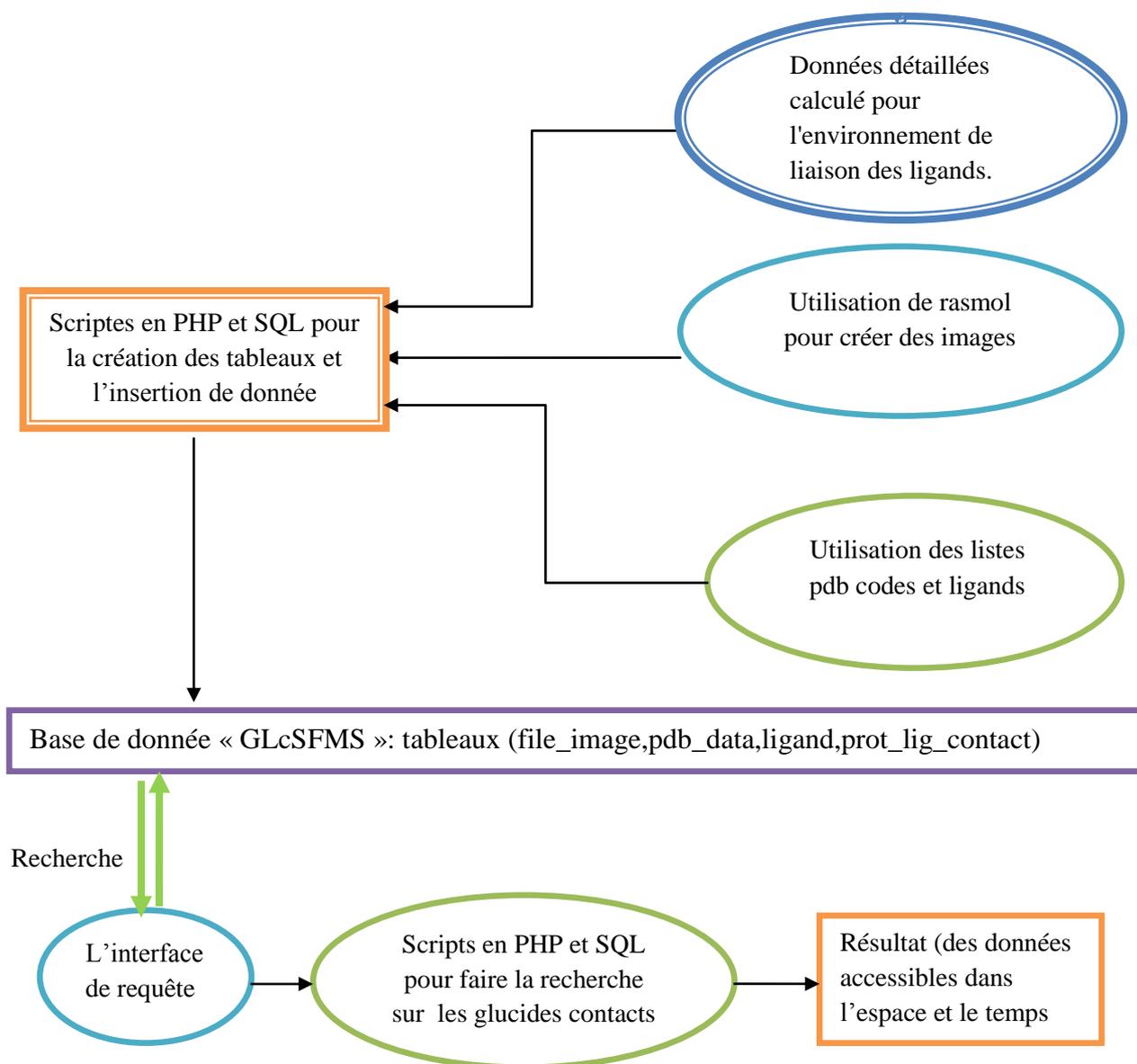


Figure 71 : Schéma représentatif des processus de la création et exploitation de la base de données.

III. Plateforme Production (web):

Après avoir finalisé la création de la version locale de base de données « **GlcSFMs**», Cette dernière est transit sur un support network.

Toutes les fonctions qui existent dans la version locale sont maintenues dans une version unique.

Pour l'accès à la base de données « **GlcSFMs**» cette adresse url est usée

www.bioinformaticstools.org/prjs/glsfms/

Cette version online est important pour simplifier l'accès a la base de donnée et partager les information avec le communauté scientifique et universitaire.

1. Présentation des résultats

1.1. Présentation des résultats sur le serveur locale :

L'interface HTML présenté dans la figure 72 a été créée pour l'accès a les donnée enregistré dans la base de données locale et aussi pour faciliter l'exploration de ces données par les utilisateurs.

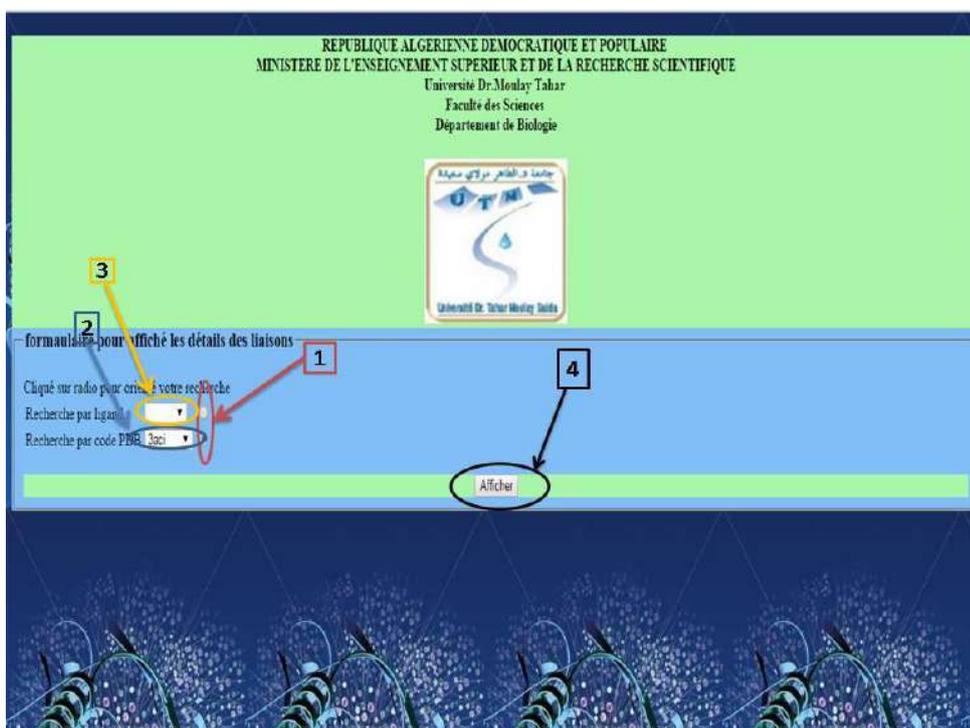


Figure 72 : Capture d'écrans représente la page menu.

La figure 72 représente les points importants pour orienter le mode de recherche :

1. La zone 1 présente l'utilisation de radio pour choisir le mode de recherche soit par Recherche par PDB soit Recherche par Ligands.
2. la zone 2 présente la sélection de code PDB qui permet de choisir un code PDB parmi les 100 codes stocké dans la base donnée « **GlcSFMs** ».
3. La zone 3 présente la sélection d'un ligand parmi les 34 ligands Glucidiques stocké dans la base de données « **GlcSFMs** ».
4. La zone 4 présente le bouton afficher qui fait le passage de la page de menu vers la page de résultats.

La page de résultat représente des informations concernant le code PDB ou ligand sélectionnée.

1.2. Cas de sélection par PDB :

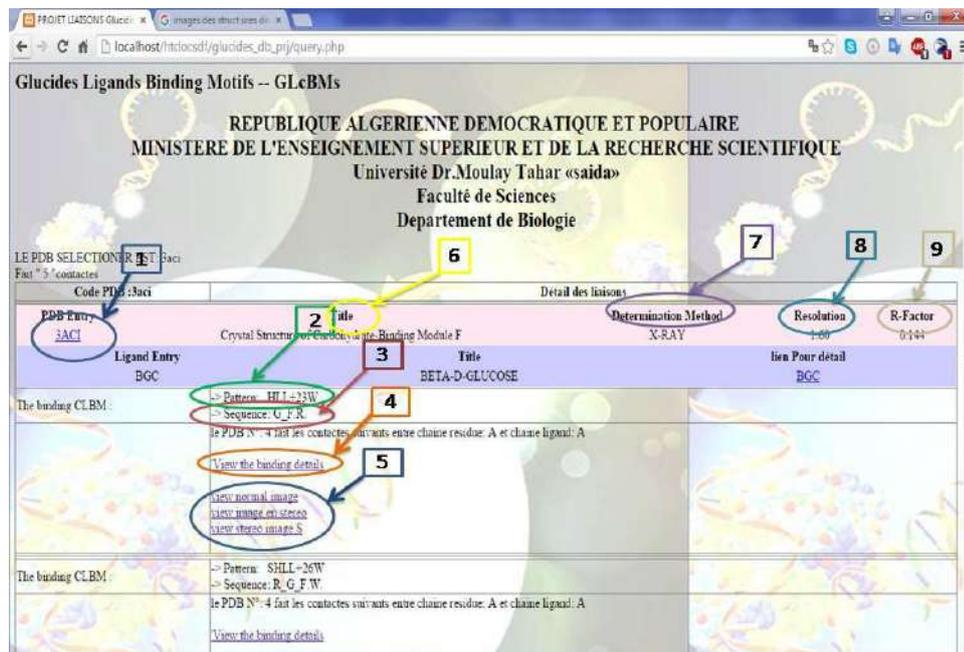


Figure 73: Capture d'écran représente la page résultat après la sélection du PDB 3ACI.

La page de sélection par PDB présente les informations important concernant le code PDB sélectionnée

1. La zone 1 présente le PDB entry ou code pdb sélectionné.
2. La zone 2 présente le pattern qui représente l'enchainement des hélices alpha, feuillets beta, parts non structurés ou boucle et aussi le nombre de molécules d'eaux.
3. La zone 3 présente la séquence des acides aminés existant dans les éléments de la zone 2 responsable des liaisons avec les ligands glucidiques.
L'ensemble (pattern-séquence) présente ce qui est dénommé comme **motif** dans cette étude.
4. La zone 4 présente le tableau de détails de contacts (voir figure 75).
5. La zone 5 présente les images de motif et (image 1, image 2, image S «stéréo» voir figure 75).
6. La zone 6 présente le titre de la PDB entrée.
7. La zone 7 présente la méthode de détermination qui est généralement dans ce projet « X-RAY »
8. Zones 8 et 9 présentent respectivement la résolution et r-factor.

La résolution et r-factor représentent la qualité de la structure de la protéine dans l'entrée PDB. La baisse de la valeur de ceux-ci mieux la structure est.

1.3. Cas de sélectionne par Ligand

REPUBLICQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
 MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
 Université Dr.Moulay Tahar «saïda»
 Faculté de Sciences
 Département de Biologie

LE LIGAND SELECTIONNER EST : 2FP
 Fait "14" contacts

Code PDB en contact avec le ligand 2FP: Détail des liaisons

PDB Entry	Titre	Determination Method	Resolution	R-Factor
1MMT	Crystal structure of fructose bisphosphate aldolase	X-RAY	2.35	0.179

The binding CLBM has the following:
 -> Pattern: SLLHSSLLLLSH-17W
 -> Sequence: A D E S T K K R E K L S G S Y G R
 ligand "2FP" N°: "350" chaîne: "A" fait contacts avec la chaîne proteique: "A"

The binding CLBM has the following:
 -> Pattern: SLLHSSLLLLSH-12W
 -> Sequence: A D E S T K K R E K L S G S Y G R
 ligand "2FP" N°: "350" chaîne: "B" fait contacts avec la chaîne proteique: "B"

Figure 74: Capture d'écran représente la page résultat après la sélection du ligand 2FP.

La Figure 74 présente la page de résultat après la sélectionné par ligand:

1. La zone 1 présente le ligand entrée et qui est le code de ligand sélectionné.
2. La zone 2 présente le code pdb en contact avec le ligand sélectionné.
3. La zone 3 présente le Title ou le nom scientifique de ligand.
4. La zone 4 présente le lien url pour plus d'exploration de ligand.

Dans les deux cas A et B, une ligne s’affiche qui contient le pattern et la séquence des acides aminés ainsi un tableau sur les détails de contacts et trois images qui présentent la structure 3D des motifs.



Figure 75 : Capture d’écran représente la partie d’affichage des données sur l’entrée PDB, Le ligand en contact, le motif de contact (sa pattern structurel et sa séquence d’acides aminés), détails de contact et trois images-3D montrant les éléments de contact.

2. Discussion et interprétation des résultats :

En raison du peu de temps donnée pour faire les projets masters et les ressources limitées disponibles, seulement 100 pdb codes ont été utilisés pour explorer l’environnement structurel impliqué dans la liaison de protéines à des ligands de nature glucidique. Les résultats peuvent être présentés comme indiqué dans ce qui suit:

Id (dans PDB)	Classes et Codes PDB	le nombre de contact	Les Différentes Motifs
BGC	1CQ1, 2CDB, 3VKV, 3VPH 3PQD, 1GPE, 3ACI, 1ZM1, 1BG3	21	HSSH, HH, H, SH, S, SSS, SSSS, SS, SHSSH
G6P	1BG3, 1UP6, 4QFH	8	HHHH, HSHSSH, SHSSH
NAG	1GPE, 1HKI, 9LYZ, 1LOG, 4XBN 4XBQ	16	H, HH, SHH, SS, HSHSSH, SSSSSH

2FP	3MMT, 2QDG, 1FDJ, 1ZAI	14	SHSSSSH, HSSSH, SSSSH
FBP	3VKV, 3VPH, 3PQD	9	HHS, HH
F6P	2CXS, 1T10, 1HOX	5	HHHH, HSH
MDM	2EF6, 1NIV, 4K1Y	14	HH, S, SS, H
MMA	4GX7, 1KUJ, 4OIZ, 1QDC, 1QDO	24	SS, S, H
MAN	1LOG, 1QDC, 2FMD, 1QDO	12	H, HH
LBT	4OJO, 3AP4, 4BME, 4GA9, 2NN8	12	SSS, SSSS, SH
TDG	1PV7	2	HHHHH, HHHHHH
LAT	1HKI, 4NO4, 4YM3, 3VT0, 2NN8	19	SSS, SSSS, SHSH
R5P	4M8L, 1ZHA, 4ADU	3	SSSSH, SHSSSSH, H
RIP	1DRK, 1OGD, 21 1DRJ, 4Q0P	8	HHSS, H
0MK	4Q0P, 3WW1, 4DUX	15	H, SSSSSH, SSSSS, SH
GAL	4LO1, 1CIW, 4XBQ, 4XBN	9	SHSH, H, HH, SSS
NGA	3COR, 4I4V, 4G9N	11	S SSH, SHSH
IUG	3ZGF, 3ZGG	2	SHHH, HH
ARB	1ABE, 6ABP, 3TB6, 1MMZ, 4QDP, 1BAP	8	HHH, SSSS, HSSSHS, SHSSHSS, SHSS, HSSH
ARA	1ABE, 6ABP, 3A22, 4RXT, 1BAP, 2ARC	15	SHSSHSS, SHHSS SSSSS, HSHSS, HSHS HS, HSHH, HHH
ALL	1RPJ	1	SHHS
AOS	3PH4, 2I57, 4GJJ, 3M0M	13	HSSSS, HSSHS, HSSSHS, H, HS
NA1	3FY1, 1HKJ	4	HH
3BU	3WW4	2	SSSSSSH, SSSSSS
A6P	2VVO	5	HSH, HSHH, HSHH
5RP	2VVP, 1TKU, 3OVQ, 4M8L, 4P77, 1ZHA, 2ISS, 3GND	38	SSSSH, HSSSSH, HSSH, HSSSH, HSSSSH, HSHH, HSHH, HSH, SSSHS, SSSS, HSHH, HSHH
CAP	1RBL	8	SSSH
RUB	1RXO, 1RCX	12	SHSSH, SSSH
XLS	3XIS, 1XIC, 8XIA, 4XIS	4	HSSSHS, HS
XYS	4XIS	2	HS, S
XUL	1XII	1	HSHS
XYL	4GNA, 3XIS, 1XLM, 1XIG, 2XIS, 5XIA	8	SSSS, SHS, HSSHS, HSSSHS, HS, HSHS, HSSHSS
MBG	1AFA	3	S
PAN	1KOJ	2	HHHHH, HHHHHH

Tableau 6 : Tableau représente le nombre de contacts des ligands à les PDB IDs ainsi que les motifs liés à chacun de ces ligands

En utilisant l'Excel une analyse a été effectuée sur l'ensemble des ligands glucidiques utilisée dans ce projet et on a constaté qu'il existe 53 motifs différentes sur les structures secondaires et quelque structure irrégulières ou boucles voir (tableau 07).

Numéro	Motif	Numéro	Motif	Numéro	Motif	Numéro	Motif	Numéro	Motif
1	HH	13	SHSSSS	25	HSSHS	37	HSSHH	49	HSSSSH
2	H	14	SHSSH	26	SSSSS	38	SHHSS	50	HSSSSSH
3	SH	15	HSSSH	27	SSH	39	HSHSS	51	HSHHH
4	S	16	SSSSH	28	SHHH	40	HS	52	SSSHS
5	SSS	17	HHS	29	SHHHHI	41	HSHH	53	SSHS
6	SSSS	18	HSH	30	HHHHH	42	SHHS	54	SSHSSSSH
7	SS	19	HHHHH	31	HHH	43	HSSSS	55	SSSH
8	HHHH	20	HSH	32	HSSSHS	44	SSSSSSSF	56	HSHS
9	HSHSSH	21	HHSS	33	SHSSHS	45	SSSSSSS	57	SHS
10	HSSH	22	SSSSH	34	SHSHSS	46	HSHS		
11	LL	23	SSSHS	35	SHSSHI	47	SHH		
12	LLL	24	LLL	36	L	48	HSSHSS		

Tableau 07 : Les principaux types de motifs en liaison avec les ligands du (tableau 06).

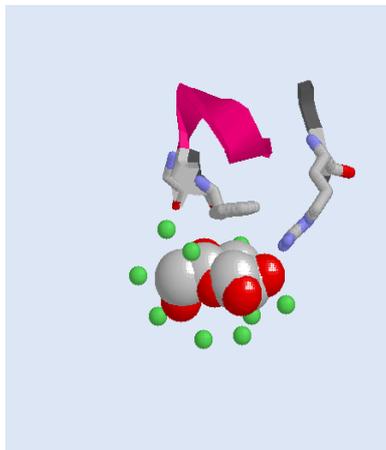
L'ensemble des motifs présentée dans le tableau 07 sont des motifs qui décrivent les éléments de la structure secondaire (en liaison avec les ligands glucidiques) à l'exception des motifs 11,12, 24 et 36 qui sont des boucles ou structures irrégulières

Les chiffres ci-dessus montrent que la plupart des motifs sont constitués par des éléments de la Structure Secondaires (éSS) qui indiquent clairement le rôle important de ces éléments dans la fonction biologique des protéines. En outre, les motifs entièrement composés de boucles bien qu'ils semblent avoir un rôle de liaison de ligand, mais parce qu'ils sont très peu nombreux par rapport à l'autre type de motifs renforce encore l'importance des éSS à la fois dans le maintien de la morphologie structurale des protéines et à l'exécution de la fonction biologique.

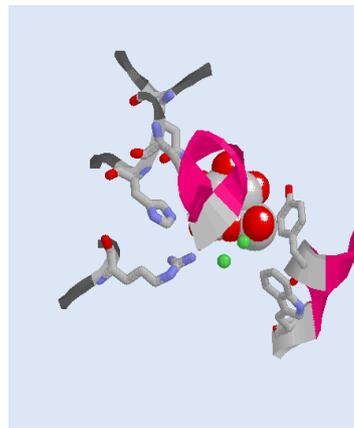
Il est important de noter que les cas où les motifs sont constitués entièrement de molécules d'eau ne sont pas importants, car ils sont sans signification biologique.

Il est également important de noter que presque tous les autres motifs listé dans le tableau 07 ont des molécules d'eau impliquées dans la liaison des ligands. Il est pour des raisons de simplicité que les molécules d'eau ont été exclues de l'annotation utilisé dans le tableau pour décrire les motifs.

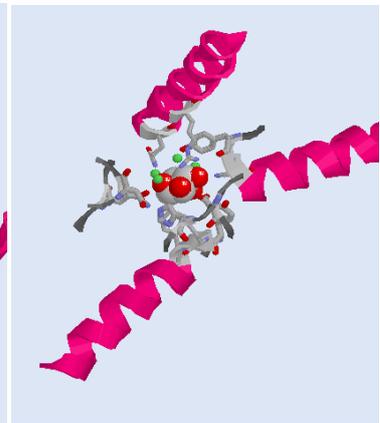
Les principaux types de motifs qui se lient les ligands glucidiques peuvent être visualisés en 3-D, comme indiquent les images présentées dans la Figure 76:



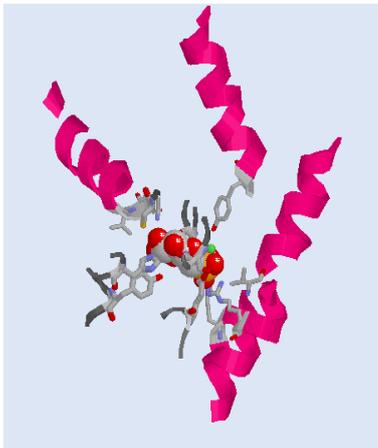
Motif « H »



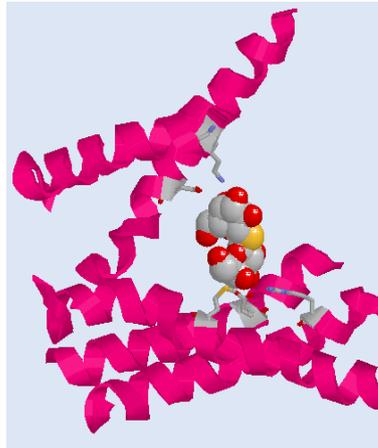
Motif « HH »



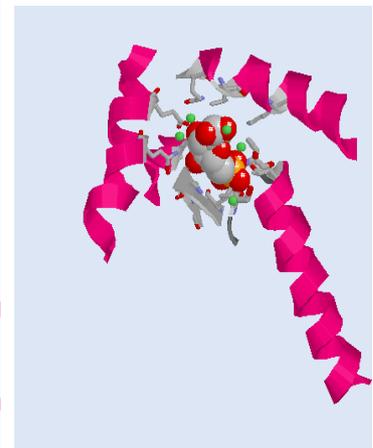
Motif « HHH »



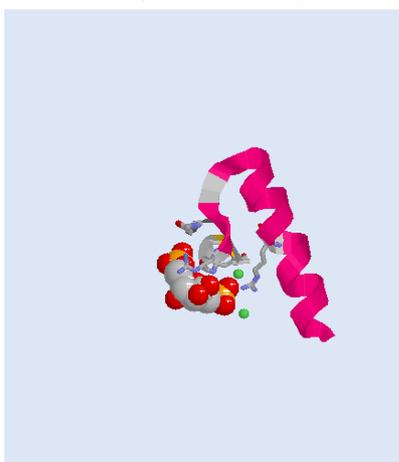
Motif « HHHH »



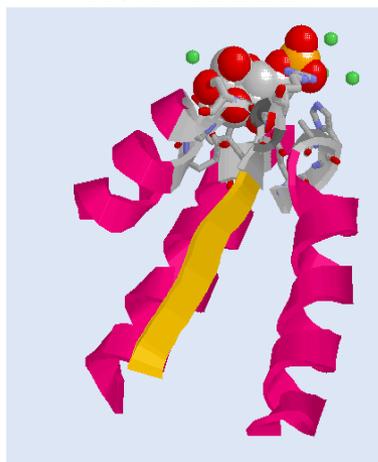
Motif « HHHHH »



Motif « HHHHHH »



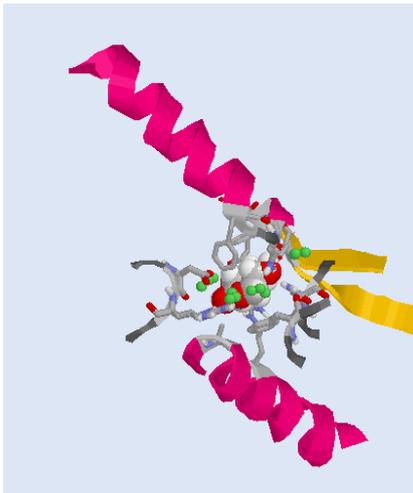
Motif « HHS »



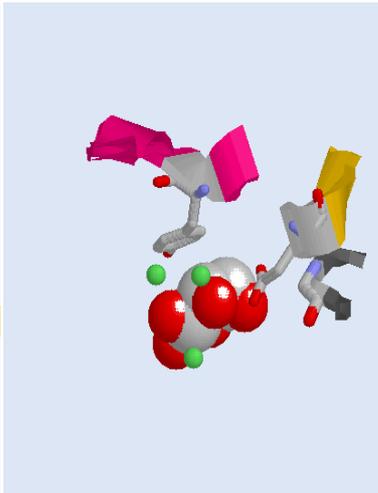
Motif « HSSH »



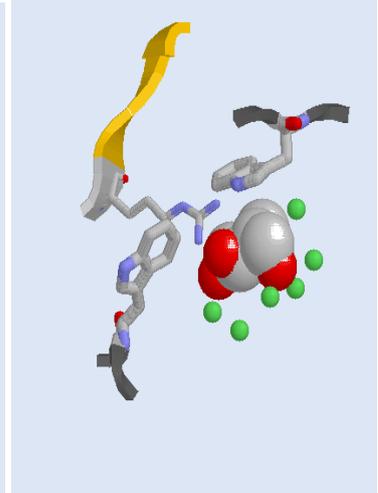
Motif « HSHH »



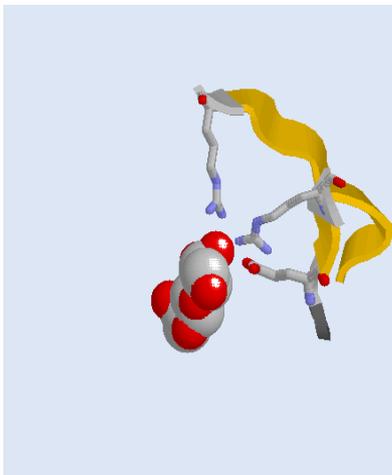
Motif « HSS »



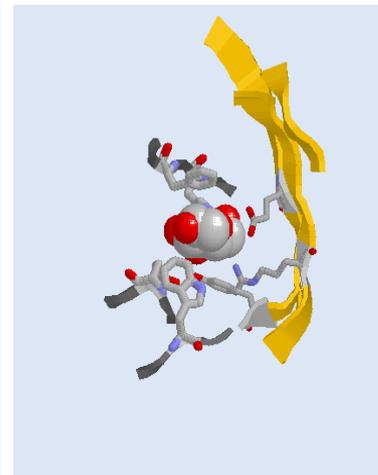
Motif « HS »



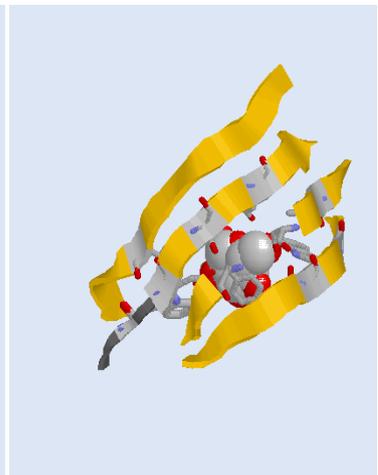
Motif « S »



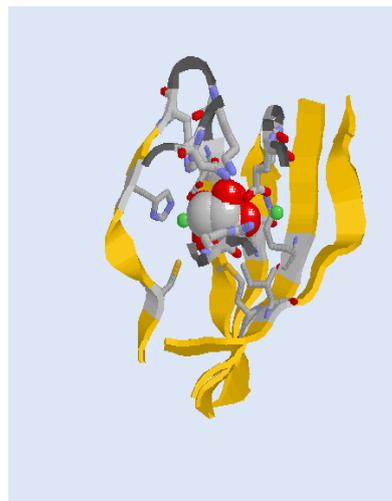
Motif « SS »



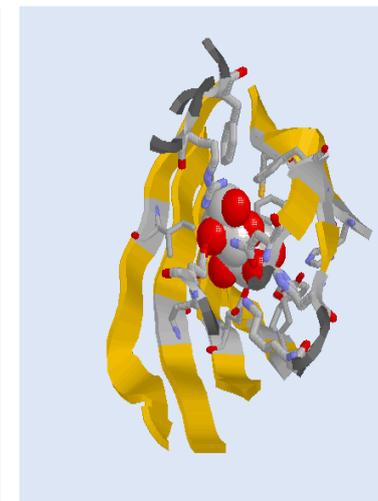
Motif « SSS »



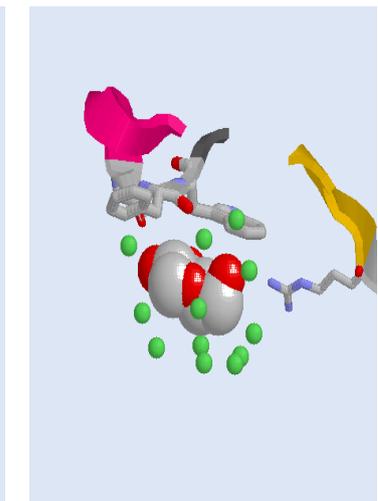
Motif « SSSS »



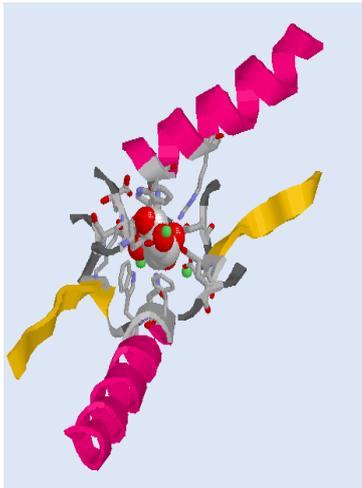
Motif « SSSSS »



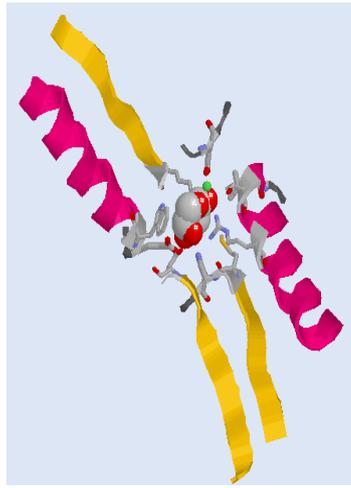
Motif « SSSSSSS »



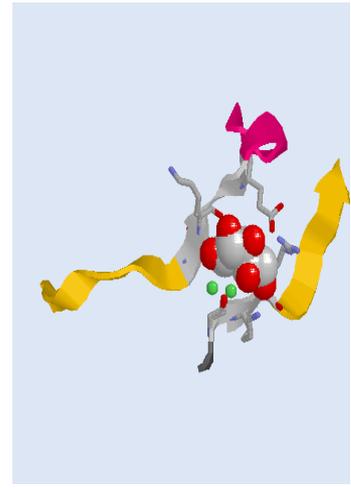
Motif « SH »



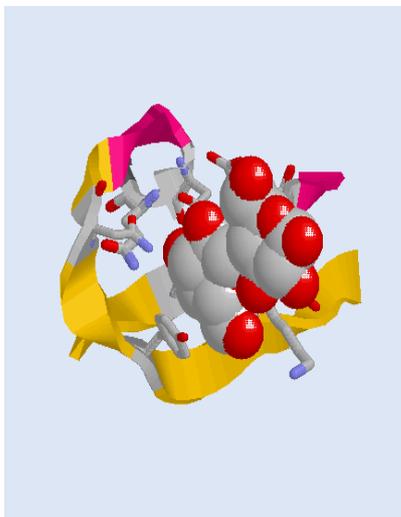
Motif « SHHS »



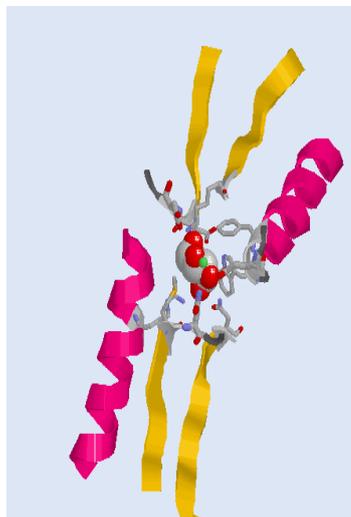
Motif « SHHSS »



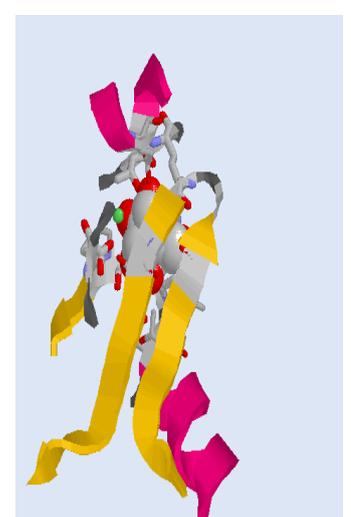
Motif « SHS »



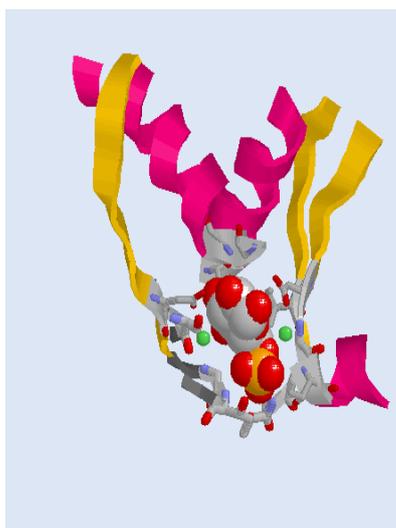
Motif « SHSH »



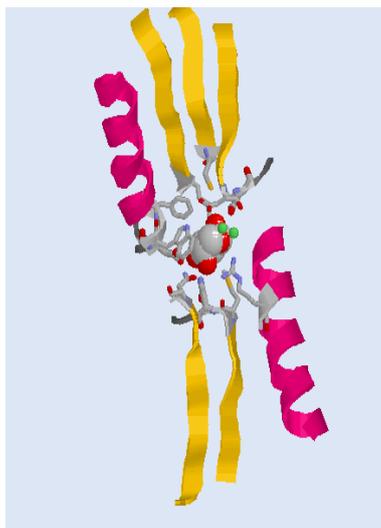
Motif « SHSHSS »



Motif « SHSSH »



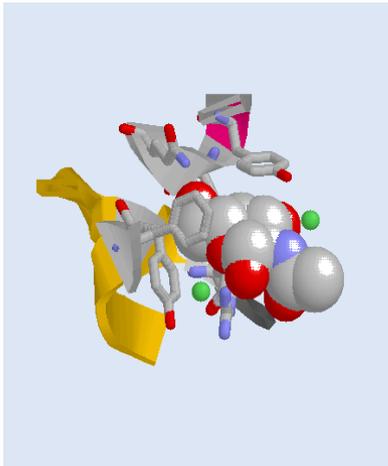
Motif « SHSSHH »



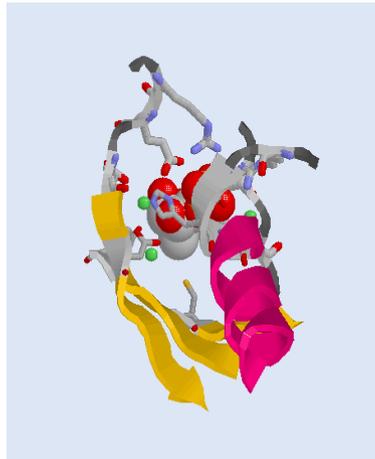
Motif « SHSSHSS »



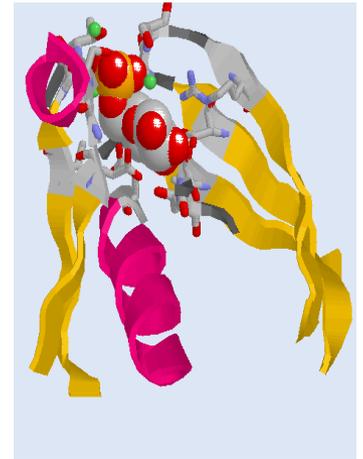
Motif « SHSSSSH »



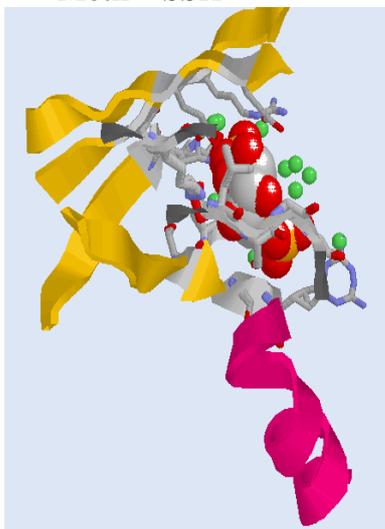
Motif « SSH »



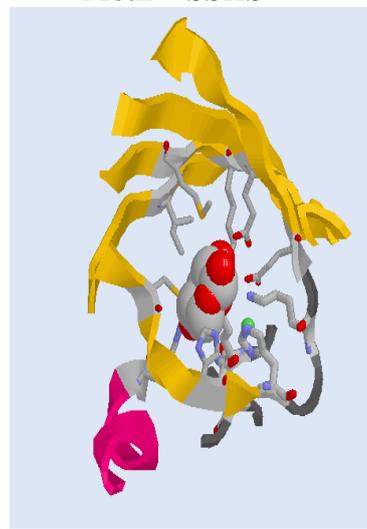
Motif « SSHS »



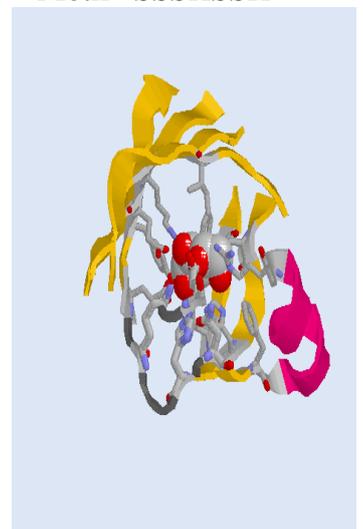
Motif « SSSHSSH »



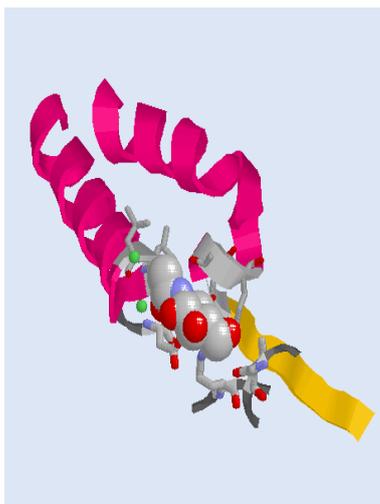
Motif « SSSSH »



Motif « SSSSSH »



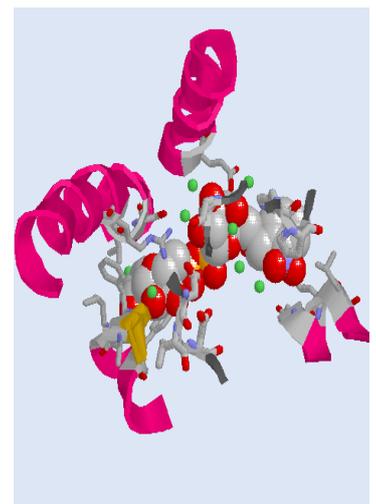
Motif « SSSSSSH »



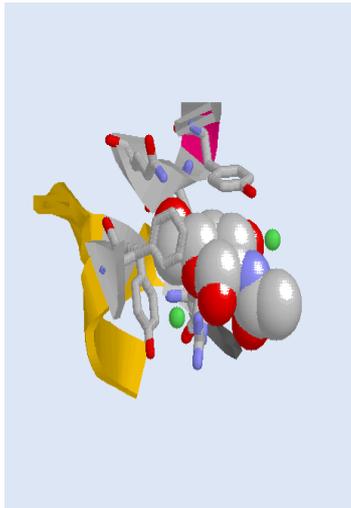
Motif « SHH »



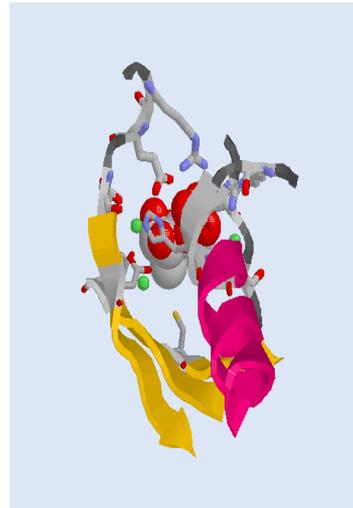
Motif « SHHH »



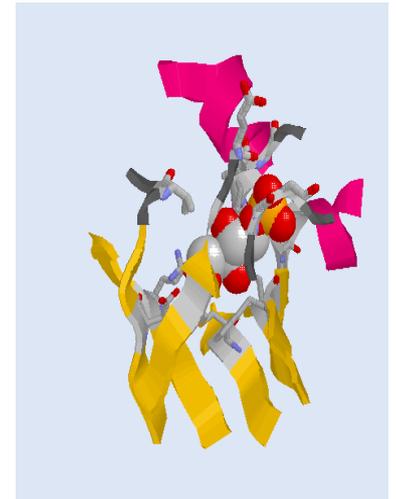
Motif « SHHHHH »



Motif « SSH »



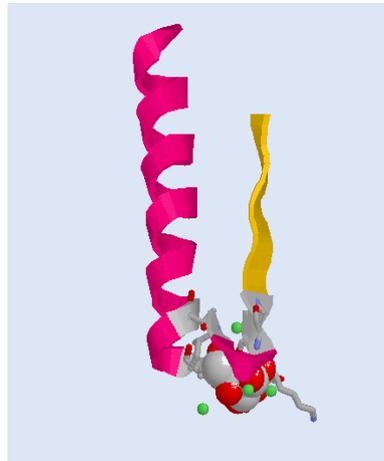
Motif « SSHS »



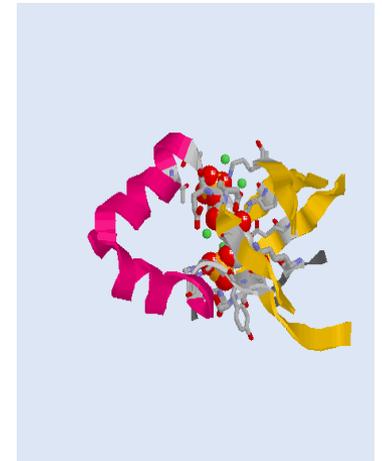
Motif «SSHSSSSH»



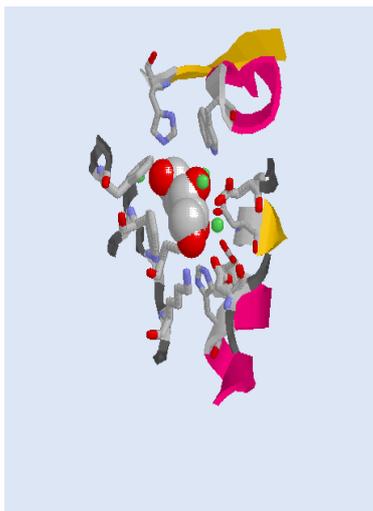
Motif « HHSS »



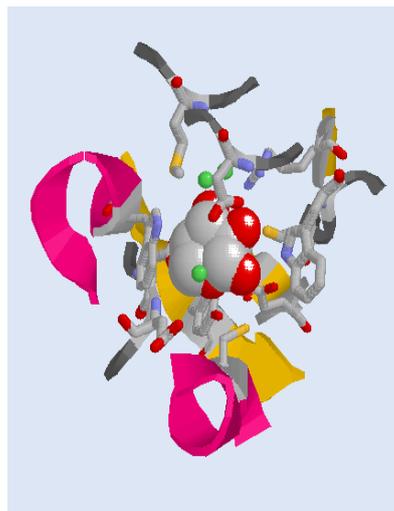
Motif « HSH »



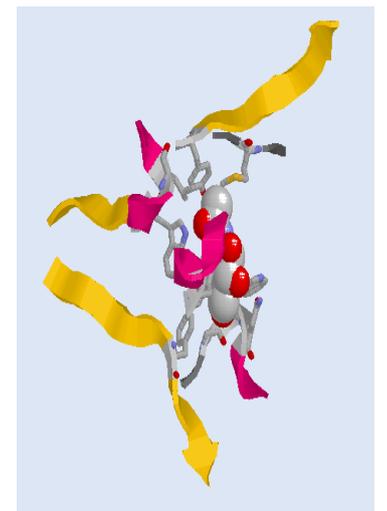
Motif «SHSSSSH»



Motif « HSHS »



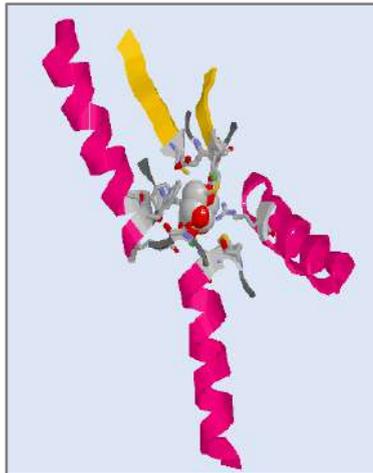
Motif « HSHSS »



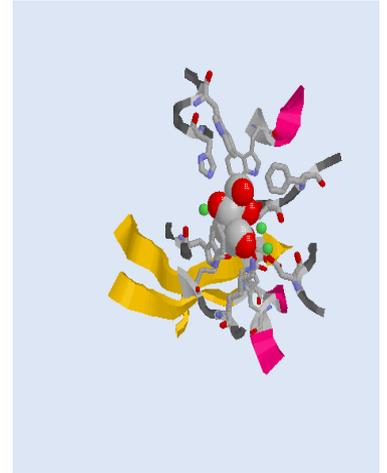
Motif « HSHSSH »



Motif « HSSH »



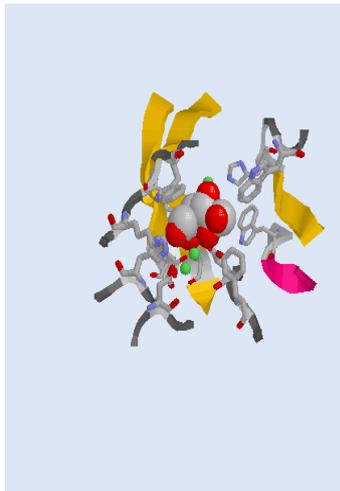
Motif « HSSHH »



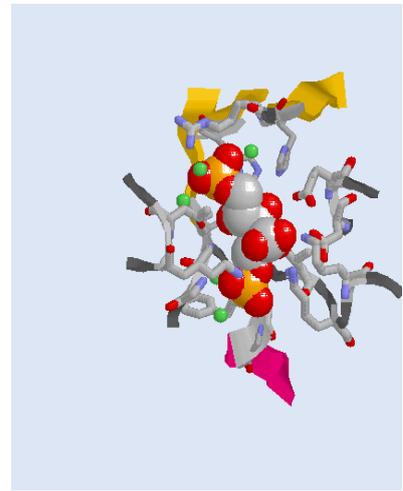
Motif « HSSHS »



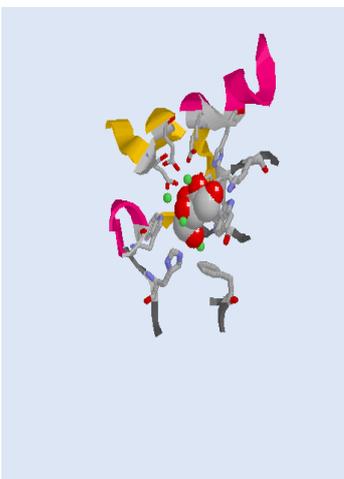
Motif « HSSSH »



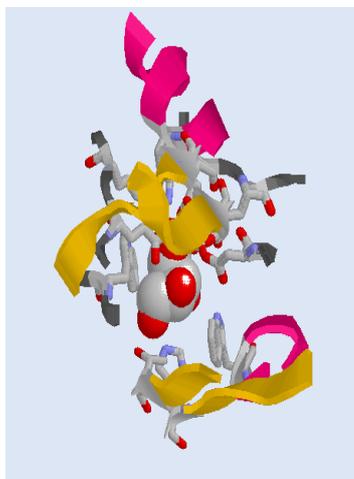
Motif « HSSSS »



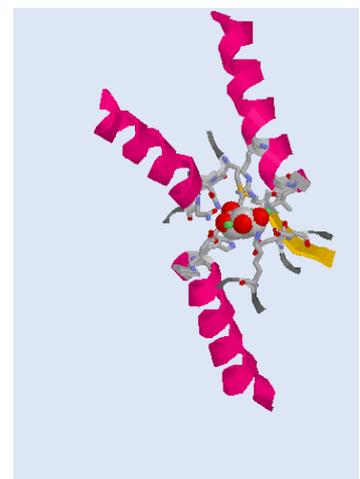
Motif « SSSH »



Motif « HSSHSS »



Motif « SSSHS »



Motif « HSHH »

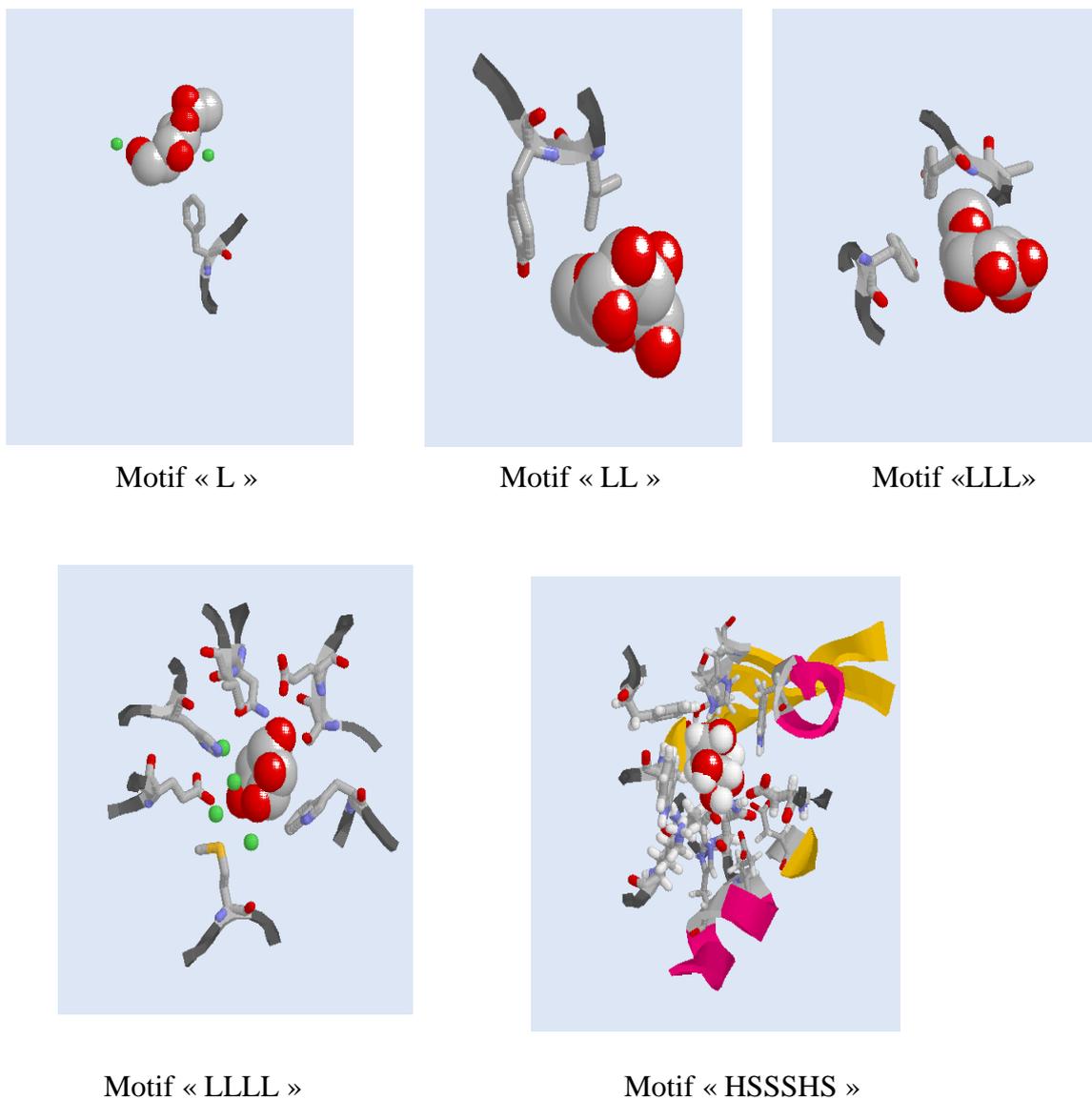


Figure 76 : Images 3-D des motifs présentés dans le tableau 07.

Pour savoir les motifs les plus fréquentes parmi les 53 motifs qui concernent les éléments de la structure secondaire pourcentage relatif de ces motifs à été calculée et mentionnée dans le graphe présentée dans la figure 43.

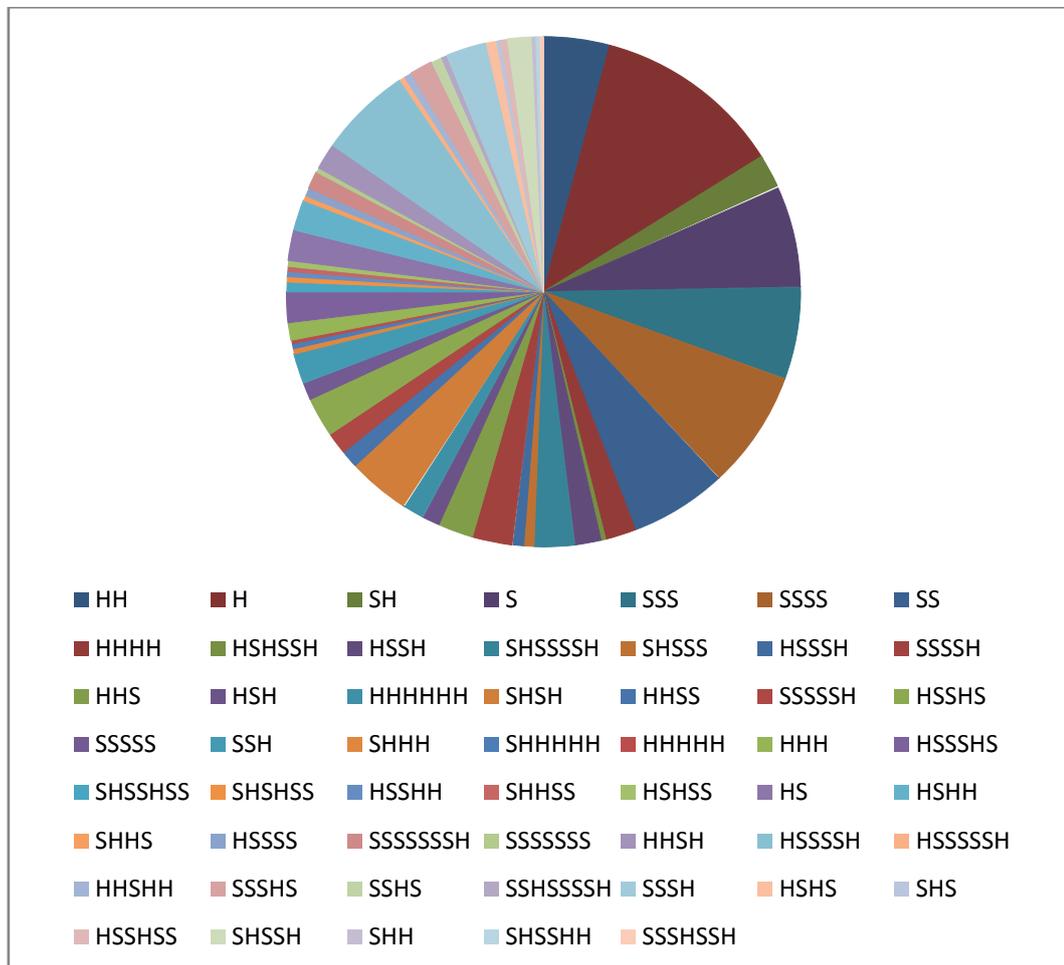


Figure 77 : Représentation graphique à secteur des taux de fréquence des motifs.

Les motifs contenant les éléments de structure secondaire ont une fréquence de plus de 93% des motifs obtenue dans cette étude.

A partir de figure 77 en peut répartir les motifs contenant des éléments de structures secondaires en trois niveaux en premier lieu le motif de structures secondaires le plus abondante est «H» (hélice α) avec une fréquence de plus de 12% , Les motifs « S » (feuille β) « HH » « SHSH » « SSS » « HSSSSH » « SS » « SSSS » ont un fréquence varie entre 4% et 7,5%. Les autres motifs ont une fréquence d'apparition moins de 3% pour chacun d'ils.

A partir de cette étude en peut déduire que les éléments de structure secondaires ont un rôle important dans la fonction biologique des protéines

Cette étude suggère que les motifs des structures secondaires peuvent faire des liaisons avec des ligands de nature glucidique pour former des complexes biologiquement actives

A partir de l’histogramme on peut observer que les motifs de petite taille c’est à dire ceux composés d’un seul ou deux éléments de structure secondaire sont les plus abondants par rapport aux motifs complexes et que ces motifs font des liaisons avec les ligands glucidiques pour former des complexes actifs ce qui explique le rôle structural dans la fonction biologique

Puisque la figure 77 ne présente que seules les fréquences de répartition globales des motifs, une analyse a été nécessaire de faire pour que l’étude sera plus objective cette analyse est basée sur la fréquence des motifs par ligand. Les résultats sont représentés sur le graphique suivant, Figure 78:

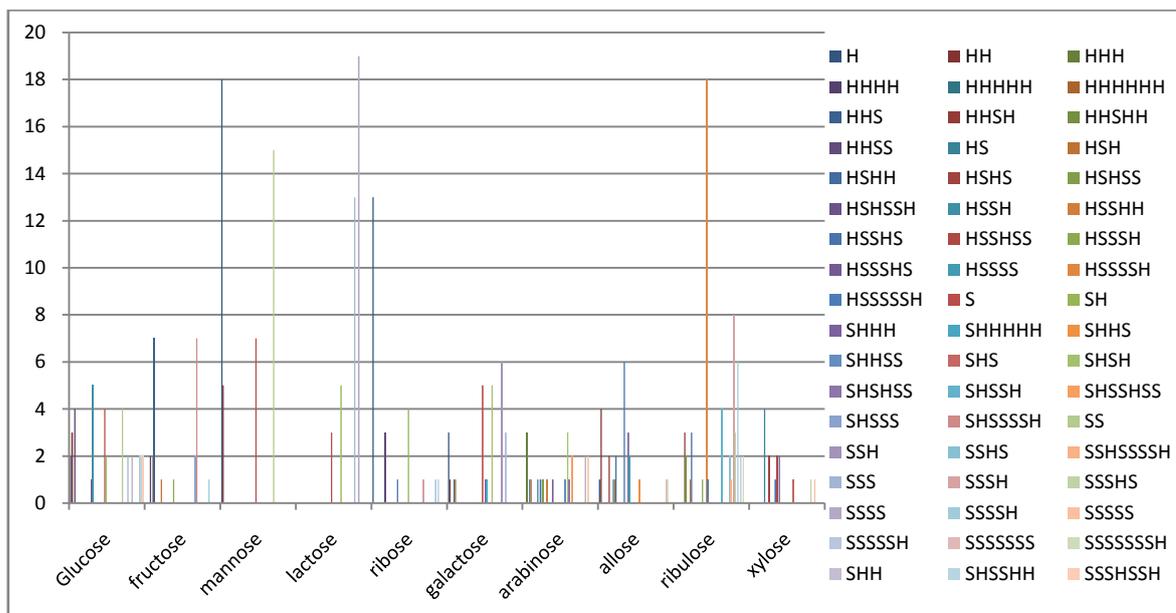


Figure 78 : La fréquence des motifs par rapport aux ligands.

A partir de l’histogramme (figure 78), on peut retirer les observations suivantes :

Pour le **glucose** les motifs de structure secondaire les plus abondants sont «HSSH» «HHHH» «S» «SS» qui varie entre 17% et 14%, les motifs «HH» «H» «SH» «S» «SSS» «SSSS» présentent moins de 10%, et le motif le moins abondant est «HSHSSH» avec une présence de 4%, par rapport au **fructose** qui a une présence majeure de deux motifs «SHSSSSH» «HHS» plus de 60% ,et

Les motifs «HHHH» «HSHSSH» «HSSH» «SHSS» «HSSSH» «SSSSH» «HSH» «HHHHHH» sont moins abondantes varie entre 4% et 8%, le **mannose** a une exception que ne contient que les hélices α ou les feuilles β pour chaque motif comme le cas des deux motifs les plus abondantes «H» «SS» qui présentent plus de 73% les motifs «S» «HH» sont moins abondantes ne présentent que seule 26% ,les deux motifs «SSSS» «SSS» présentent 80% de totalité des motifs de **lactose** par a port des motifs «SHSH» «S» qui ne présentent que 20%,pour le **ribose** le motif le plus abondante est «H» 54% les motifs «SH» «HHSS» varie entre 16% et 12% les motifs inférieures à 4% sont «SHSSSSH» «SSSSH» «SSSSSH» «HSSH», pour le **galactose** les motifs les plus abondantes sont «S» «SHSH» «SSH» qui présentent 60% et les motifs «HH» «H» «SSS» «HHHHHH» «SHHH» «SHHHHH» «HHHHH» présentent 40% ,concernant **l'arabinose** les motifs les plus abondantes sont «SSSS» «SHSH» «SSSS» «HHH» «SHSSHSS» leur présence est varies entre 9% et 14% les Motifs les moins abondantes «HHHHHH» «HHHHH» «HSSSHS» «SHSHSS» «HSSH» «SHHSS» «HSHSS» «HS» «HSHH» avec un présence de 43%,en ce qui concerne **l'allose** les deux motifs «HSSH» «HH» sont les plus présente 40% tandis que les motifs «H»«HSH»«HS» «HSSSHS» «HSHH» «SHHS» «HSSSS» «SSSSSSH» «SSSSSS»

«HSSH» varient entre 4% et 12%, pour le **ribulose** le motif le plus fréquent est «HSSSSH» il présente 31,5% de totalité des motifs, les motifs «SSSSH» «SSH» présentent 24% les motifs «SHSSH» «SSHSSSH» «SSHS» «SSSHS» «HSHH» «HSSSSH» «HSSH» «SSSSSSH» «HSHH» «SSSSH» «HSH» «HSSSH» sont mois abondantes varient entre 1,75% et 7% ,Et enfin le motif «SH» est le plus abondant dans le ligand **xylose** par un présence de plus de 26% les motifs «S» «HSSH» «SSSS» «HSSSHS» «SSSHS» «HSHS» «SHS» «HSSHSS» varient entre 6% et 13%.

Cette analyse suggère clairement que chaque ligand à une spécificité qui va influencer a la conformation spatiale locale dans la structure globale des protéines.

Parmi les 53 motifs de structures secondaires il y'a 06 motifs qui sont les plus représentés dans les structures protéiniques avec une fréquence d'occurrence dépend du ligand en complexe avec la protéine : «H», «S», «SS», «SSSS», «SSS», et «HSSSSH».

Additivement au travail précédent qui été basé sur les motifs et leurs fréquences d’occurrence, le travail suivant est basé surtout sur les séquences et les acides aminés dominants sur les structures en rapport avec les ligands (voire, Figure 79) et en rapport avec les motifs (voire, Figure 80).

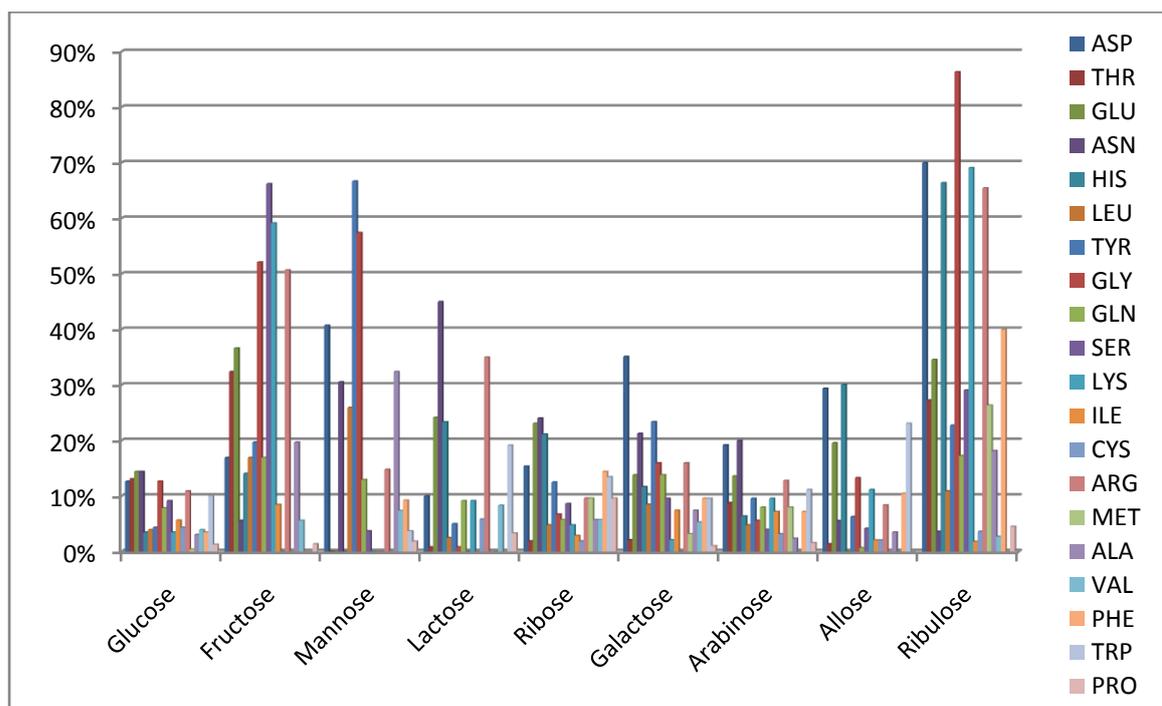


Figure 79 : Le taux de présence des acides aminés dans la séquence en contact avec les ligands par rapport aux ligands

Le graph présenté dans la figure 79 confirme le caractère d’affinité des acides aminés aux glucides, parmi les 20 acides aminés chacun d’eux à son affinité spécifique à un glucide bien déterminé.

A partir de figure 79 on peut observer que l’acide aspartique et l’asparagine sont les deux acides aminés les plus apparus dans les ligands glucidiques étudiés.

. Le tableau 08 récapitule les acides aminés qui présentent une affinité très élevée aux glucides dans les structures protéiniques étudiées:

LIGAND	Acide Aminé les plus abondant
glucose	ASP /THR/ GLY
fructose	GLY/GLU
mannose	GLY/ TYR

lactose	ASN/ARG
ribose	ASN/ GLU
galactose	ASP/ ASN
arabinose	ASP/ASN
allose	ASP/ HIS
ribulose	HIS/ASP/LYS
xylose	ASP/TRP/GLU

Tableau 08 : Les acides aminés les plus abondants par rapport aux ligands.

Pour que l'étude soit plus claire et approfondie, Un autre analyse sur l'Excel a été faite pour savoir le taux des acides aminés par apport aux motifs obtenue dans cette étude.

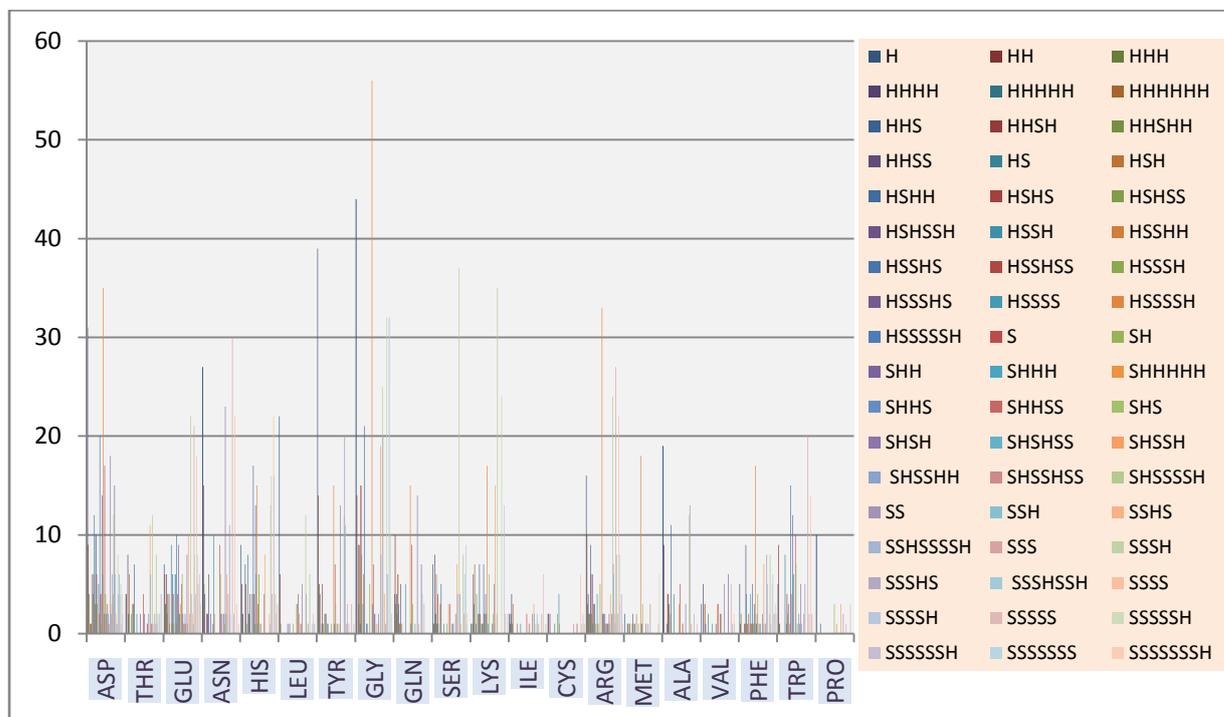


Figure 80: Le taux de présence des acides aminés en contact avec les ligands par rapport aux Motifs.

La figure 80 suggère que les grands motifs c'est à dire les motifs qui contiennent plusieurs éléments de structure secondaire sont ceux qui présentent un taux élevé des acides aminés donc on peut déduire que les grands motifs ont une complexité structurale par rapport aux petits motifs.

En outre, la tendance de certains types d'acides aminés d'être responsables de la liaison des ligands, comme on le voit ci-dessus, suggère que certains acides aminés pourraient être plus enclins à exécuter des rôles fonctionnels et d'autres peuvent être plus associés à un rôle structurel dans le maintien et la stabilité structurale des protéines.

Conclusion:

Ce projet est pour but de comprendre la nature structurale des sites de contacts entre des ligands glucidiques et les protéines. Et voir aussi les interactions qui se produisent au niveau de ces sites de contacts qui permet de mieux comprendre les relations structure-fonction des protéines.

Les résultats et les discussions de cette étude (dans le Chapitre 4) montrent une tendance claire des ligands glucidiques à lier certains types de motifs structurales et se lie favorablement certains types d'acides aminés. Ceci consiste un témoignage de l'importance de ce genre d'étude en révélant les éléments structuraux impliqués dans la réalisation de la fonction des protéines, contribuant ainsi à la compréhension de la relation structure-fonction des protéines.

Comme vu dans les résultats, cette étude a **identifié**, **défini** et **caractérisé** les éléments de structure ou les motifs structurales qui sont impliqués dans la liaison des ligands glucidiques trouvés dans le complexe à un ensemble de structures de protéines pour lesquelles les glucides sont importants dans l'exercice de leur fonction biologique spécifique.

L'implication de ces types de motifs en soutenant les deux côtés structurales et fonctionnelles des protéines et enzymes justifie le nom qui leur est donné dans cette étude: Glucides Binding Structural and Functional Motifs - GLcSFMs.

Comparé à d'autres types de motifs structurels signalés dans Chapitre 1, GLcFMs ont des caractéristiques distinctes et importantes qui peuvent être résumés comme suit:

1. Dans la plupart des cas, les GLcSFMs sont de nature **discontinue** par rapport à l'espace **1-Dimensionnel** (de la séquence primaire de la protéine). Cela signifie que les éléments de la structure secondaire qui les composent peuvent appartenir à des régions éloignées de la protéine. Ceci constitue une différence majeure entre ces motifs de liaison et les motifs structurales classique, voir "IV- Les Motifs Structurales", Chapitre 1.
2. Les GLcSFMs sont de nature **continue** dans l'espace **3-Dimensionnel** car les acides aminés responsables de la liaison des ligands bien qu'ils puissent appartenir à des régions éloignées, ils se réunissent dans l'espace et forme la poche de liaison.

3. Les GLcSFMs peuvent être faites entièrement d'éléments de la Structures Secondaire (ESS), boucles, ou un mélange des ESS et des boucles. En outre, des molécules d'eau sont généralement impliquées dans le processus de liaison et donc ils sont inclus dans leur définition.

La connaissance accumulée de cette étude, y compris l'identification, la définition et la caractérisation précise des éléments structurelles impliqués dans la liaison des ligands qui sont importants pour la fonction biologique des protéines sélectionnées peut contribuer à un certain nombre de méthodes bioinformatique tels que la Conception Rationnelle des Nouveaux Médicaments et la Modélisation Moléculaire. Ceci est en plus de faire des contributions à la compréhension des fondamentaux derrière la relation entre la structure et la fonction des protéines.

A la fin, il convient de souligner que cette étude a été réalisée sur un échantillon relativement petit de structures de protéines en raison du temps et les ressources limitées disponibles pour un projet master. Un ensemble de structures de protéines beaucoup plus large et représentatif serait nécessaire, dans des études futures, pour clarifier davantage les remarques et les conclusions tirées de cette étude en ce qui concerne le nombre de types de motifs de liaison et les types d'acides aminés impliqués dans la liaison des ligands de nature Glucidique.

Annexe 1

Section 1 : les protéines

1. Classification des protéines

La classification des protéines peut être établie selon différents critères :

1.1. Chimique

1.1.1. Holoprotéines : protéines vraies.

1.1.1.1 Sphéroprotéines : Classée selon la solubilité d'après Osborne en 1907, citée par Cheftel et al. (1985) :

a)-Albumine : Soluble dans l'eau et difficilement précipitable par les sels.

b)-Globuline : Peu ou insoluble dans l'eau mais soluble dans les solutions salées
exemple : actine et myosine (viande).

c)-Prolamine : Soluble dans les solvants organiques (Ethanol à 70%).

Exemple : Gliadine de blé

d)-Gluténine : Insoluble dans l'eau, partiellement soluble dans les acides et bases dilués. Exemple : gluténine de blé

e)-protamine : Connus dans les poissons. Exemple : histone riche en histidine

1.1.1.2. Scléroprotéines : Insolubles dans l'eau, dans les solutions salées, dans les acides et bases dilués et les solvants organiques neutres. Indigestibles ; résistent à l'action des protéases .exemple : les kératines (poils, ongles, cheveux,)

1.1.2. Hétéroprotéines : Ce sont des protéines composées ou conjuguées ; association entre les acides aminés et groupe prosthétique, selon la nature de ce dernier, on distingue :

1.1.2.1. Les phosphoprotéines : Protéines et $H_3P_{O_4}$. Exemple : l'ovovitelline (protéine de jaune d'œufs liée à lesécithines).

1.1.2.2. Les nucléoprotéines : Protéines et acides nucléiques.

1.1.2.3. Glycoprotéines : Protéines et glucides.

1.1.2.4. Chromoprotéines : Protéines et pigments.

1.2. Conformation

1.2.1. Sphéroprotéines : Protéines globulaires.

1.2.2. Scléroprotéines : Protéines fibreuse.

1.3. Fonction

a)-Protéines structurales. Exemple : élastine, collagène.

b)-Protéines plasmatiques et de liquides biologiques.

c)-Protéines à activité biologique.

Exemple : enzymes, hormones, transporteurs,...

d)-Protéines alimentaires : protéines économiquement favorisées, digestible et savoureuses (Alias et Linden, 1994 ; Cheftel et al. , 1985)

2. Propriétés des protéines

Définition de la fonctionnalité des protéines alimentaires

Les propriétés fonctionnelles des protéines alimentaires sont l'expression des propriétés physiques et physico-chimiques au niveau des qualités sensorielles des aliments qui les contiennent. Elles peuvent être classées en trois groupes principaux :

2.1. Propriétés d'hydratation : ce groupe inclut les propriétés suivantes :

L'absorption et la rétention d'eau, la mouillabilité, le gonflement, l'adhésion, la dispersibilité, la solubilité et la viscosité

2.2. Propriétés dépendant des interactions protéines-protéines :

Ce groupe inclut les propriétés suivantes : agrégation, fibrillation, extrusion, texturation et la formation de diverses autres structures (fibres et pâtes protéiques,...). Aussi, la coagulation, synérèse et gélification appartiennent à cette catégorie mais d'une manière plus complexe, avec modification de l'affinité pour l'eau.

2.3. Propriétés de surface (propriétés aux interfaces) :

Ce groupe inclut le moussage, foisonnement, émulsification , formation de films.

3. La solubilité

La dissolution d'un corps (C) dans un solvant (S) ne se produit que si les forces d'attraction qui s'établissent entre les molécules de (C) et celles de (S) sont supérieures aux forces de cohésions auxquelles sont soumises les molécules de (C). dans le cas de protéines, ces forces de cohésion sont représentées par les attractions : interactions électrostatiques, liaison hydrogènes et attraction de van der waals (tableau 09) (Boulangier et al...1959).

La solubilité des protéines est la résultante des interactions protéine-eau (affinité pour l'eau). C'est l'expression de la propriété d'hydratation qui met en jeu surtout des interactions hydrogènes et de van der waals (Frénot et Vierling, 2006).

On trouve de grandes différences dans le monde des protéines, ce qui crée une disparité vis-à-vis des autres classes, glucides et lipides. On peut distinguer trois cas :

-Solubilité dans l'eau pure

-Solubilité seulement en présence des sels neutres .

-Insolubilité dans l'eau, solubilité en milieu alcalin, ou en présence de solvant organique (Alias et Lindes, 1994).

De ce fait, les facteurs influençant la solubilité des protéines sont les suivants :

4. Influence du pH

Lorsque le pH du solvant est supérieur ou inférieur au pH_i des protéines

($pH \neq pH_i$), la protéine porte une charge négative ou positive, cette ionisation de protéine provoque le déroulement de la molécule et crée des interactions de type dipole-dipole entre les groupes chargés des protéines et l'eau en contribuant à sa solubilisation. De plus, les chaînes protéiques portant des charges électriques de même signe ont tendance à se répulser et par suite de se dissocier et se déplier si on représente les variations de la solubilité d'une protéine donnée en fonction du pH, on obtient habituellement une courbe en forme V ou U, le minimum de celle-ci correspondant au pH_i

Lorsque le pH du solvant se rapproche ou équivaut au pH_i , les forces électrostatiques maximales (minimum d'interaction entre l'eau et la chaîne polypeptidique), il s'ensuit une diminution de la solubilité des protéines qui s'agrègent en polymères insolubles entrant en contact entre elles et se précipitent sous forme d'agrégats (Wagner et Anon, 1990 ; Boulanger et al., 1959).

5. Influence de la force ionique (u)

Selon Frénot et Vierling (2006) et Scopes (1994), la solubilité d'une protéine pour une force ionique donnée, varie en fonction de la nature des ions et la concentration en sels

6. Interaction et liaisons impliquée dans la structure protéique

Liaisons non covalentes

6.1. Interactions de Van Der Waals :

Les interactions de Van Der Waals sont des dipôles non permanents de faible rayon d'action (fig 12). Elles sont nombreuses et contribuent essentiellement dans la

recherche de la concordance stérique entre le ligand et la protéine réceptrice.
(Bensegueni. A) .



Figure 81. Interactions de Van Der Walls

6.2. Ponts hydrogène

Les ponts hydrogène avec une force de l'ordre de 2 à 4 Kcal/mol agissent dans les Systèmes biologiques, à très courte distance (0,8 à 2,8 Å) entre un H lié à un O ou un N avec un O ou un N. Ils sont peu nombreux et s'adaptent très bien à la flexibilité (l'angle peut varier de 120° à 180°). (Bensegueni. A) (fig13)



Figure 82 : ponts hydrogène

6.3. Interactions électrostatiques

Les interactions électrostatiques agissent à plus grande distance entre dipôles chargés (fig14). La distance pour une liaison saline, par exemple, est de 3,7 à 4,5 Å, et sa force est de l'ordre de 8 Kcal/mol. En présence de solvant, ces interactions sont favorisées du fait de la polarité et de la petite taille de la molécule d'eau. (Bensegueni. A)

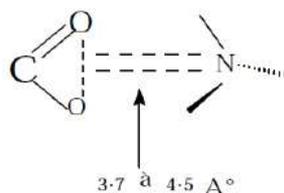


Figure 83 : Interactions électrostatiques

Liaisons covalentes

6.4. Pont-disulfure

Un pont disulfure (lien S-S) est une liaison covalente qui se forme par oxydation dans les protéines, de manière post-traductionnelle. Cette liaison se forme entre les atomes de soufre des fonctions thiol de deux cystéines dans une séquence peptidique (ou protéine) (voir la figure ci-dessous où les R représentent le reste de l'acide aminé).

La molécule résultante de la liaison de deux cystéines est la cystine.

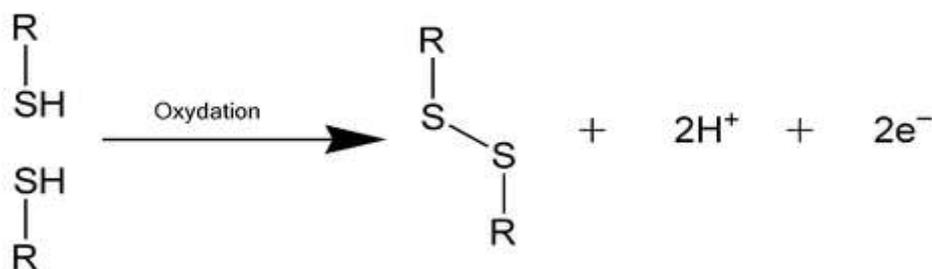


Figure 84 : Formation d'un pont disulfure.

Le pont disulfure est un élément des structures tertiaires (après le repliement de la protéine) ou quaternaire (lors d'association de sous-unités protéiques) de la protéine. La formation d'un pont disulfure à partir de deux cystéines s'accomplit spontanément en conditions oxydantes, en particulier en présence de dioxygène. Ils ne se forment en général pas dans le cytoplasme, qui est un environnement réducteur, mais lorsque les protéines sont sécrétées ou exposées à la surface cellulaire.

En biochimie, la réduction d'un pont disulfure peut se faire avec des réducteurs doux, tel le 2-mercaptoéthanol ou le dithiothréitol (DTT ou réactif de Cleland) (wiki/Pont_disulfure).

type	Energie (Kj /mol)	Distance de l'interaction (Å°)	Groupe fonctionnel impliqué	Agent de rupture	Agent de renforcement
Liaison covalente	330-380	1 - 2	Résidus de cystéine S-S	mercaptoethanol cystéines, dithiothréitol, sulfites	/
liaison hydrogène	8 - 40	2 - 3	amide, -NH...O=C hydroxyl phénol -OH...O=C	solution d'urée, chlorhydrate de guanine, détergents, chaleur	froid
interactions hydrophobes	4 - 12	3 - 5	résidus d'acides aminés à chaîne latérale aliphatique ou aromatique	détergents, solvants organiques	chaleur
interactions électrostatiques	24 - 84	2 - 3	groupe carboxyl (COO ⁻) aminé (NH ₄ ⁺) ...	solution saline pH faible ou élevés	/
Van der Waals	1 - 9		dipôles permanent, induit et instantané	/	/

Tableau 09 : interaction et liaisons impliquée dans la structure protéique (cheftel et al,1985),

Section 2 : les glucides

1. Classification des glucides

Les glucides sont généralement classés en fonction de leur degré de polymérisation (DP). Dans la littérature, cette classification n'est pas consensuelle (WHO/FAO, 1998), (Lehninger,1981), (Dobbing, 1989) en ce qui concerne les limites de degré de polymérisation à considérer dans chaque catégorie : les sucres (1 à 2 oses avec ou non les polyols), les oligosaccharides (2<DP<10 à 12) et les polysaccharides.

En effet selon Lehninger (1981), on distingue :

- Les oses ou sucres simples ou monosaccharides,
- Les oligoholosides ou oligosaccharides comportant 2 à 10 molécules d'oses reliées par des liaisons glycosidiques (ou osidiques),
- Les polyosides ou polyholosides ou polysaccharides, formés de longues chaînes linéaires ou ramifiées.

Dobbing mentionne que toutes les classifications des glucides sont arbitraires et

Particulièrement dans les distinctions entre glucides simples et glucides complexes. Pour sa part, il qualifie de « sucres » tous les sucres simples et leurs dérivés ainsi que les oligosaccharides contenant jusqu'à 5 unités mono saccharidiques (Dobbing, 1989). Dans le rapport de la WHO/FAO (1998), il est mentionné les classes suivantes :

- Les sucres (DP 1 et 2), regroupant les monosaccharides, les diosides ou

disaccharides et les polyols,

- Les oligosides ou oligosaccharides (DP 3 à 9) comportant 2 sous-groupes, les maltooligosaccharides (ou maltodextrines) et les autres oligosaccharides parmi lesquels, les α -galactosides et les fructo-oligosaccharides,

- Les polysaccharides (DP>9) qui comportent 2 sous-groupes : l'amidon et les Polysaccharides non amylacés.

1.1. Les oses

Les oses simples ou constitutifs des oligosaccharides et polysaccharides sont des hexoses (6 atomes de carbone) (glucose, galactose, fructose...)

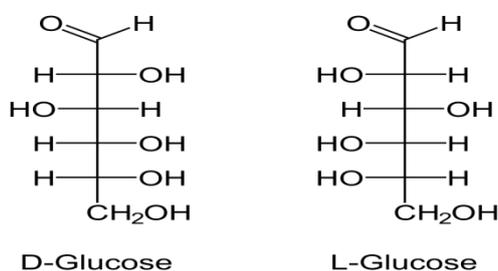


Figure 85 : exemple d'un hexose (glucose)

Et des pentoses (5 atomes de carbone) (arabinose, xylose...)

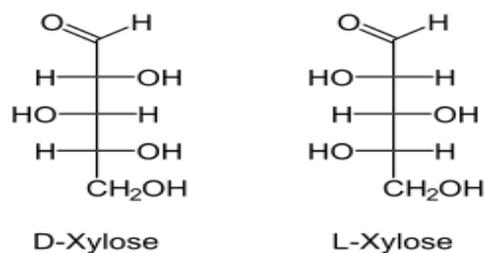


Figure 86 : exemple d'un pentose (xylose)

Qui se présentent sous la forme de cycles à 5 carbones + 1 oxygène (glucose, galactose...) ou à 4 carbones + 1 oxygène (fructose, arabinose, xylose...) ,D'une façon générale, les carbones du cycle sont liés à un atome d'hydrogène et à une fonction hydroxyle (OH) ou alcool (CH₂OH) et/ou à l'oxygène du cycle.

Le carbone 1 de nombreux oses (glucose, fructose) porte une fonction dite pseudoaldéhydrique qui possède un pouvoir réducteur élevé qui explique leur propension à donner lieu, lors de leur cuisson, en présence d'acides aminés à des réactions de Maillard

(Formation de composés colorés). La réaction de Maillard, appelée aussi brunissement non enzymatique, non-oxydatif, se produit à température élevée.

Elle permet par exemple d'obtenir le brunissement de la croûte du pain ou la coloration des aliments frits mais également le développement d'arômes comme lors de la torréfaction des graines de cacao. Les conséquences nutritionnelles de ces réactions ne sont pas négligeables puisque des acides aminés essentiels et en particulier, la lysine et à un moindre degré, le tryptophane sont rendus indisponibles par cette réaction avec les oses réducteurs. Stéréo-isomérisation des oses : la plupart des oses naturels sont de forme « D » (dextrogyre) mais il existe des oses naturels de forme « L » (levogyre). Cette lettre qui précède le nom de l'ose indique, par convention, la configuration du carbone asymétrique le plus éloigné du groupement carbonyle dans la représentation linéaire de Fischer. Mutarotation et formes anomères des oses : les oses dont la fonction pseudo-aldéhyde n'est pas impliquée dans une liaison peuvent être sous forme a ou b en fonction de l'orientation de la fonction hydroxyle par rapport au plan de la molécule (position équatoriale ou axiale).

Les différentes formes, du glucose par exemple, a ou b-D-glucose, sont interconvertibles, ce qui entraîne le phénomène de mutarotation.

1.2. Les diosides ou disaccharides

Ils sont composés de deux unités de monosaccharides unis par une liaison glycosidique. Sous forme polymérique, les liaisons sont figées en un état a ou b. Le maltose est ainsi l'a-D-glucopyranosyl-1,4-D-glucopyranose signifiant qu'il est constitué de 2 unités D-glucose reliées entre elles par une liaison a-1,4 impliquant le carbone 1 (la liaison C-O étant dans la position a) d'une unité a-D-glucose, non réductrice, et le carbone 4 d'une deuxième unité qui possède un groupement réducteur libre. Le lactose ou b-D-galactopyranosyl-1,4-D-glucopyranose

contient du D-galactose et du D-glucose liés par une liaison b-1.

L'autre dioside très répandu (dans le règne végétal) est le saccharose

(Fréquemment appelé « sucre »).

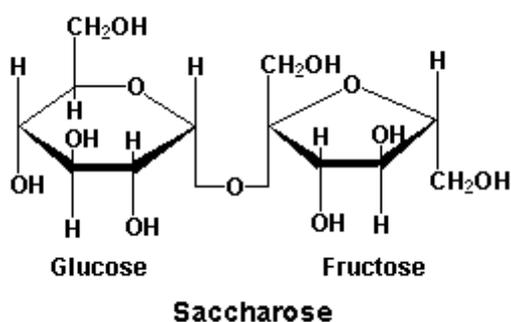


Figure 87 : structure de saccharose

Le tréhalose (α -D-glucopyranosyl-1,1- α -D-glucopyranose) est présent dans de nombreux champignons, levures ou algues.

1.3. Les polysides ou polysaccharides

L'amidon est formé d'une chaîne α -glucosidique. L'amylose et l'amylopectine sont les deux polymères constitutifs de l'amidon. Le premier résulte de l'enchaînement linéaire d'unités glucose liées par des liaisons α -1,4. L'amylopectine présente une structure branchée où des fragments linéaires de type amylose sont associés entre eux par des liaisons α -1,6.

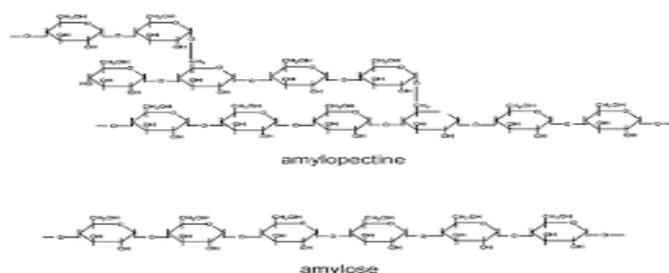


Figure 88 : image représente la structure de l'amidon

Le glycogène est le polysaccharide de réserve de l'organisme animal de structure voisine de l'amylopectine mais avec plus de ramifications.

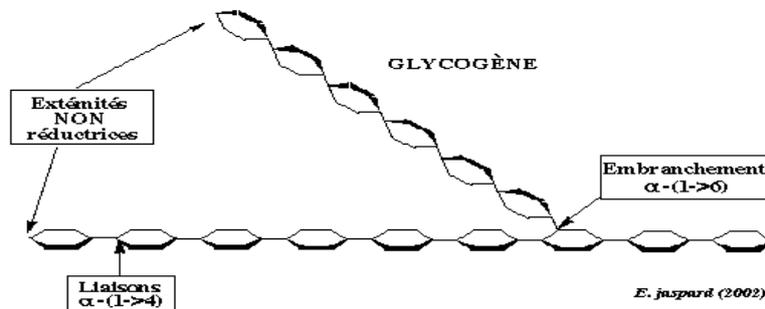


Figure 89 : structure de glycogène

Les autres polysaccharides sont, à l'exception des dérivés de l'amidon ou du glycogène, des composés indigestibles par les enzymes endogènes du tube digestif de l'homme ; ils sont donc pour beaucoup d'entre eux classés dans les fibres alimentaires (Afssa, 2002).

Si la cellulose présente des analogies avec l'amidon, les liaisons reliant les unités glucose sont de type β -1,4. Ces liaisons ne sont hydrolysées que par certaines des bactéries hébergées dans le côlon humain.

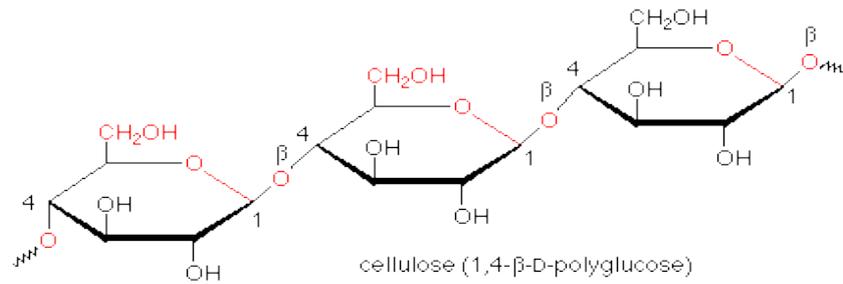


Figure 90 : structure de cellulose

La plupart des autres polysaccharides non digestibles (hémicelluloses, pectines, gommes) ont des structures complexes impliquant plusieurs types d'oses et de liaisons (Afssa, 2002).

Section 3 : Les glycoprotéines

1. Mannotriochitobiose

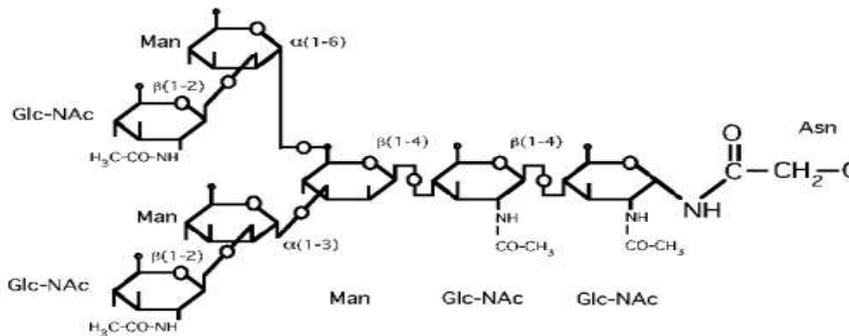


Figure 91 : structure de Mannotriochitobiose

- Les oligosaccharides des protéines N-glycosylées ont en commun un noyau central de deux

N-acétyl-glucosamines (Glc-NAc) et trois mannoses (Man).

- Sur les mannoses distaux on trouve des oses branchés qui forment des antennes : ici, il y a

deux antennes formées de mannose puis de N-acétyl-glucosamine liée en β 1-2.

- Il peut y avoir trois antennes ou plus lorsque d'autres oses viennent sur les autres carbones des

mannoses distaux du noyau central.

- Les antennes sont souvent terminées par des acides sialiques.

2. Transferrine

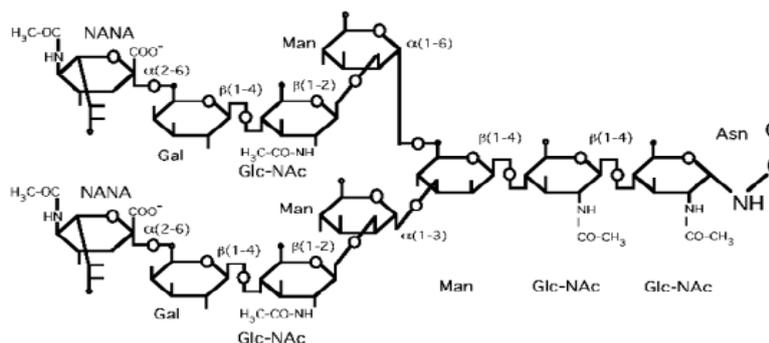


Figure 92 : structure de Transferrine

- La transferrine est une protéine plasmatique transporteuse de fer.
- La transferrine est une glycoprotéine de 88000 daltons dont les deux oligosaccharides sont liés à des Asn ; c'est une protéine N-glycosylée.
- Dans la structure de chaque oligosaccharide on retrouve le noyau central formé de 2 N-acétylglucosamines et de 3 mannoses. A cette structure de base, les enzymes de l'appareil de Golgi des hépatocytes ont ajouté des « antennes » liées aux mannoses distaux par des liaisons β 1-2. Chacune de ces antennes est formée de N-acétylglucosamine, de galactose et se termine par un acide N-acétylneuraminique (NANA).
- La transferrine est reconnue par son récepteur et peut jouer sa fonction de transporteur lorsqu'elle est pourvue de ses chaînons oligosaccharidiques.
- Le vieillissement de la molécule entraîne l'hydrolyse des acides sialiques terminaux. Une glycoprotéine plasmatique dépourvue de ses acides sialiques devient un ligand pour un récepteur hépatique qui la capte et le conduit vers une dégradation totale.

Annexe 2

Section 1 :

1. Liste PDB entries :

Las entrees PDB sont stocké dans un fichier nommée LISTEPDB.txt voir ci-dessous

Ce fichier contient les informations sur les PDB entrees le nom PDB, la méthode, résolution, R-factor séparer par le caractère # pou faciliter leur insertion dans la base de donnée «GlcSFMs» :

```
2cdb#SULFOLOBUS SOLFATARICUS GLUCOSE DEHYDROGENASE 1 IN COMPLEX WITH NADP AND GLUCOSE#X-RAY#1.60#0.192
1cq1#Soluble Quinoprotein Glucose Dehydrogenase from Acinetobacter Calcoaceticus in Complex with PQQH2 and Glucose#X-
RAY#1.90#0.190
1up6#STRUCTURE OF THE 6-PHOSPHO-BETA GLUCOSIDASE FROM THERMOTOGA MARITIMA AT 2.55 ANGSTROM RESOLUTION
IN THE TETRAGONAL FORM WITH MANGANESE, NAD+ AND GLUCOSE-6-PHOSPHATE#X-RAY#2.55#0.202
3aci#Crystal Structure of Carbohydrate-Binding Module Family 28 from Clostridium josui Cel5A in complex with
cellopentaose#X-RAY#1.60#0.144
4qfh#Structure of a glucose-6-phosphate isomerase from Trypanosoma cruzi#X-RAY#1.80#0.140
4EM5#Trypanosoma cruzi Glucose-6-P Dehydrogenase in complex with G6P#X-RAY#3.35#0.218
1BG3#RAT BRAIN HEXOKINASE TYPE I COMPLEX WITH GLUCOSE AND INHIBITOR GLUCOSE-6-PHOSPHATE#X-RAY#2.80#0.196
1zm1#Crystal structures of complex F. succinogenes 1,3-1,4-beta-D-glucanase and beta-1,3-1,4-celotriose#X-RAY#2.30#0.180
1GPE#GLUCOSE OXIDASE FROM PENICILLIUM AMAGASAKIENSE#X-RAY#1.80#0.164
9LYZ#X-RAY CRYSTALLOGRAPHY OF THE BINDING OF THE BACTERIAL CELL WALL TRISACCHARIDE NAM-NAG-NAM TO
LYSOZYME#X-RAY#2.50#0.230
3MMT#Crystal structure of fructose bisphosphate aldolase from Bartonella henselae, bound to fructose bisphosphate#X-
RAY#2.35#0.179
1ZAI#Fructose-1,6-bisphosphate Schiff base intermediate in FBP aldolase from rabbit muscle#X-RAY#1.76#0.155
2QDG#Fructose-1,6-bisphosphate Schiff base intermediate in FBP aldolase from Leishmania mexicana#X_RAY#2.20#0.185
1FDJ#FRUCTOSE 1,6-BISPHOSPHATE ALDOLASE FROM RABBIT LIVER#X-RAY#2.10#0.174
3VKV#Complex structure of Lactobacillus casei lactate dehydrogenase with fructose-1,6-bisphosphate#X-RAY#2.70#0.200
3PQD#Crystal structure of L-lactate dehydrogenase from Bacillus subtilis complexed with FBP and NAD+#X-RAY#2.38#0.215
1hox#CRYSTAL STRUCTURE OF RABBIT PHOSPHOGLUCOSE ISOMERASE COMPLEXED WITH FRUCTOSE-6-PHOSPHATE#X-
RAY#2.10#0.221
3VPH#L-lactate dehydrogenase from Thermus caldophilus GK24 complexed with oxamate, NADH and FBP#X-RAY#2.00#0.209
1T10#Phosphoglucose isomerase from Leishmania mexicana in complex with substrate D-fructose-6-phosphate#X-
RAY#2.35#0.222
2CXS#Crystal structure of mouse AMF / F6P complex#x-ray#1.50#0.173
4K1Y#Crystal structure of Canavalia boliviana lectin in complex with Man1-3Man-OMe#X-RAY#2.50#0.205
```

1NIV#MANNOSE-SPECIFIC AGGLUTININ (LECTIN) FROM SNOWDROP (GALANTHUS NIVALIS) BULBS IN COMPLEX WITH MANNOSE-ALPHA 1,3-METHYL-D-MANNOSE#X-RAY#3.00#0.223

2EF6#Canavalia gladiata lectin complexed with Man1-3Man-OMe#X-RAY#2.10#0.192

4GX7#Vibrio Cholerae Cytolysin Beta-Prism Domain With Methyl-Alpha-Mannose Bound#X-RAY#2.85#0.197

4OIZ#Structure, interactions and evolutionary implications of a domain-swapped lectin dimer from Mycobacterium smegmatis#X-RAY#3.40#0.244

1log#X-RAY STRUCTURE OF A (ALPHA-MAN(1-3)BETA-MAN(1-4)GLCNAC)-LECTIN COMPLEX AT 2.1 ANGSTROMS RESOLUTION#X-RAY#2.10#0.175

1KUJ#Crystal structure of Jacalin complexed with 1-O-methyl-alpha-D-mannose#X-RAY#2.00#0.191

1QDO#MAN(APLHA1-3)MAN(ALPHA1-O)METHYL CONCAVALIN A COMPLEX#X-RAY#2.80#0.164

1QDC#MAN(APLHA1-6)MAN(ALPHA1-O)METHYL CONCAVALIN A COMPLEX#X-RAY#2.00#0.175

2fmd#Structural basis of carbohydrate recognition by Bowringia milbraedii seed agglutinin#X-RAY#1.90#0.186

4GA9#Crystal Structure of Rat Galectin-1 in Complex with Lactose#X-RAY#1.88#0.218

4jo#Crystal Structure of Putative Tailspike Protein (TSP1, orf210) from Escherichia coli O157:H7 Bacteriophage CBA120 in Complex with Lactose#X-RAY#2.00#0.177

4bme#Crystal structure of the N terminal domain of human Galectin 8, F19Y mutant#X-RAY#2.00#0.197

3AP4#Crystal structure of the galectin-8 N-terminal carbohydrate recognition domain in complex with lactose#X-RAY#2.33#0.194

1pv7#Crystal structure of lactose permease with TDG#X-RAY#3.60#0.271

2nn8#Crystal structure of human galectin-3 carbohydrate-recognition domain with lactose bound, at 1.35 angstrom resolution#X-RAY#1.35#0.165

4no4#Crystal Structure of Galectin-1 L11A mutant#X-RAY#1.40#0.188

4ym3#Crystal structure of the human galectin-4 C-terminal carbohydrate recognition domain in complex with lactose#X-RAY#1.89#0.169

3vt0#Crystal structure of Ct1,3Gal43A in complex with lactose#X-RAY#2.91#0.253

2zgo#Crystal structure of AAL mutant H59Q complex with lactose#X-RAY#2.00#0.200

4adu#Crystal structure of plasmodial PLP synthase with bound R5P intermediate#2.44#0.196

1zha#A. aeolicus KDO8PS R106G mutant in complex with PEP and R5P#X-RAY#1.74#0.207

1dbp#IDENTICAL MUTATIONS AT CORRESPONDING POSITIONS IN TWO HOMOLOGOUS PROTEINS WITH NON-IDENTICAL EFFECTS#2.20#0.160

1drk#Crystal structure of the type 1 RIP bouganin#X-RAY#1.80#0.176

4q0u#Crystal structure of Acinetobacter sp. DL28 L-ribose isomerase mutant E204Q in complex with L-ribose#X-RAY#1.98#0.221

3ww1#X-ray structure of Cellulomonas parahominis L-ribose isomerase with L-ribose#X-RAY#1.95#0.206

4DUX#E. coli (lacZ) beta-galactosidase (N460S) in complex with L-ribose#X-RAY#2.30#0.170

4q0p#Crystal structure of Acinetobacter sp. DL28 L-ribose isomerase in complex with L-ribose#X-RAY#1.93#0.208

1ogd#THE STRUCTURE OF BACILLUS SUBTILIS RBSD COMPLEXED WITH D-RIBOSE#X-RAY#1.95#0.198

1drj#PROBING PROTEIN-PROTEIN INTERACTIONS: THE RIBOSE-BINDING PROTEIN IN BACTERIAL TRANSPORT AND CHEMOTAXIS#X-RAY#2.50#0.176

4LO1#HA17-HA33-Gal#X-RAY#2.25#0.200

4XBQ#Crystal Structure of Human Galectin-7 in Complex with Type 1 N-acetylglucosamine#X-RAY#2.23#0.197

4XBN#Crystal Structure of Human Galectin-3 CRD in Complex with Type 1 N-acetylglucosamine#X-RAY#2.21#0.168

4I4V#BEL beta-trefoil complex with N-acetylglucosamine#X-RAY#1.50#0.188

4G9N#Crystal structure of the Rhizoctonia solani agglutinin in complex with N'-acetyl-galactosamine#X-RAY#2.20#0.154

3COR#Crystal structure of the complex of peptidoglycan recognition protein (PGRP-S) with N-acetylglucosamine at 3.1 A resolution#X-RAY#3.10#0.208

1afa#STRUCTURAL BASIS OF GALACTOSE RECOGNITION IN C-TYPE ANIMAL LECTINS#X-RAY#2.00#0.221

3ZGG#Crystal structure of the Fucosylgalactoside alpha N- acetylglucosaminyltransferase (GTA, cisAB mutant L266G, G268A) in complex with NPE caged UDP-Gal (C222(1) space group)#X-RAY#1.90# 0.186

3zgf#Crystal structure of the Fucosylgalactoside alpha N- acetylglucosaminyltransferase (GTA, cisAB mutant L266G, G268A) in complex with in complex with NPE caged UDP-Gal (P2(1)2(1)2(1) space group)#X-RAY#0.183#0.217

1CIW#PEANUT LECTIN COMPLEXED WITH N-ACETYLGLUCOSAMINE#X-RAY#2.70#0.194

3tb6#Structure of the effector-binding domain of arabinose repressor AraR from Bacillus subtilis#X-RAY#2.21#0.188

1KOJ#Crystal structure of rabbit phosphoglucose isomerase complexed with 5-phospho-D-arabinonohydroxamic acid#X-RAY#1.90#0.223

1MMZ#Crystal structure of galactose mutarotase from Lactococcus lactis complexed with L-arabinose#X-RAY#1.80#0.172

4qdp#Joint X-ray and neutron structure of Streptomyces rubiginosus D-xylose isomerase in complex with two Cd²⁺ ions and cyclic beta-L-arabinose#X-RAY#2.00#0.231

1bap#A PRO TO GLY MUTATION IN THE HINGE OF THE ARABINOSE-BINDING PROTEIN ENHANCES BINDING AND ALTERS SPECIFICITY: SUGAR-BINDING AND CRYSTALLOGRAPHIC STUDIES#X-RAY#1.75#0.201

6ABP#SUGAR-BINDING AND CRYSTALLOGRAPHIC STUDIES OF AN ARABINOSE-BINDING PROTEIN MUTANT (MET108LEU) WHICH EXHIBITS ENHANCED AFFINITY AND ALTERED SPECIFICITY#X-RAY#1.67#0.160

2ARC#ESCHERICHIA COLI REGULATORY PROTEIN ARAC COMPLEXED WITH L-ARABINOSE#X-RAY#1.50#0.179

3A22#Crystal Structure of beta-L-Arabinopyranosidase complexed with L-arabinose#X-RAY##1.90#0.164

1ABE#NOVEL STEREOSPECIFICITY OF THE L-ARABINOSE-BINDING PROTEIN#X-RAY#1.70#0.

4RXT#Crystal structure of carbohydrate transporter solute binding protein Arad_9553 from Agrobacterium Radiobacter, Target EFI-511541, in complex with D-arabinose#X-RAY#1.35#0.157

1RPJ#CRYSTAL STRUCTURE OF D-ALLOSE BINDING PROTEIN FROM ESCHERICHIA COLI#X-RAY#1.80#0.194

3m0m#Crystal structure of Pseudomonas stutzeri L-rhamnose isomerase mutant S329F in complex with D-allose#X-RAY#1.45#0.177

3PH4#Clostridium thermocellum Ribose-5-Phosphate Isomerase B with d-allose#X-RAY#2.07#0.186

4GJJ#Crystal structure of Pseudomonas stutzeri L-rhamnose isomerase mutant H101N in complex with D-allopyranose#X-RAY#2.38#0.196

2I57#Crystal Structure of L-Rhamnose Isomerase from Pseudomonas stutzeri in Complex with D-Allose#X-RAY#1.97#0.172

1HKI#CRYSTAL STRUCTURE OF HUMAN CHITINASE IN COMPLEX WITH GLUCOALLOSAMIDIN B#X-RAY#2.55#0.223

3WW4#X-ray structures of Cellulomonas parahominis L-ribose isomerase with L-allose#X-RAY#1.95#0.217

1HKJ#CRYSTAL STRUCTURE OF HUMAN CHITINASE IN COMPLEX WITH METHYLALLOSAMIDIN#X-RAY#2.60#0.211

3FY1#The Acidic Mammalian Chitinase catalytic domain in complex with methylallosamidin#X-RAY#1.85#0.162

2VVO#CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS RIBOSE-5-PHOSPHATE ISOMERASE B IN COMPLEX WITH ALPHA D-ALLOSE 6-PHOSPHATE#X-RAY#1.85#0.162

3OVQ#Crystal Structure of hRPE and D-Ribulose-5-Phosphate Complex#X-RAY#2.00#0.149

3GND#Crystal Structure of E. coli LsrF in complex with Ribulose-5-phosphate#X-RAY#2.90#0.197

2VVP#CRYSTAL STRUCTURE OF MYCOBACTERIUM TUBERCULOSIS RIBOSE-5-PHOSPHATE ISOMERASE B IN COMPLEX WITH ITS SUBSTRATES RIBOSE 5-PHOSPHATE AND RIBULOSE 5-PHOSPHATE#X-RAY#1.65#0.168

1TKU#Crystal Structure of 3,4-Dihydroxy-2-butanone 4-phosphate Synthase of Candida albicans in complex with Ribulose-5-phosphate#X-RAY#1.66#0.184

4P77#Structure of ribB complexed with substrate Ru5P#X-RAY#2.04#0.176

1RBL#STRUCTURE DETERMINATION AND REFINEMENT OF RIBULOSE 1,5 BISPHOSPHATE CARBOXYLASE(SLASH)OXYGENASE FROM SYNECHOCOCCUS PCC6301#X-RAY#2.20#0.200

1rxo#ACTIVATED SPINACH RUBISCO IN COMPLEX WITH ITS SUBSTRATE RIBULOSE-1,5-BISPHOSPHATE AND CALCIUM#X-RAY#2.20#0.205

1rcx#NON-ACTIVATED SPINACH RUBISCO IN COMPLEX WITH ITS SUBSTRATE RIBULOSE-1,5-BISPHOSPHATE#X-RAY#2.40#0.224

2ISS#Structure of the PLP synthase Holoenzyme from Thermotoga maritima#X-RAY#2.90#0.213

4M8L#crystal structure of RpiA-R5P complex#X-RAY#2.37#0.239

8XIA#X-RAY ANALYSIS OF D-XYLOSE ISOMERASE AT 1.9 ANGSTROMS: NATIVE ENZYME IN COMPLEX WITH SUBSTRATE AND WITH A MECHANISM-DESIGNED INACTIVATOR#X-RAY#1.90#0.140

3xis#A METAL-MEDIATED HYDRIDE SHIFT MECHANISM FOR XYLOSE ISOMERASE BASED ON THE 1.6 ANGSTROMS STREPTOMYCES RUBIGINOSUS STRUCTURES WITH XYLITOL AND D-XYLOSE#X-RAY#1.60#0.151

4XIS#A METAL-MEDIATED HYDRIDE SHIFT MECHANISM FOR XYLOSE ISOMERASE BASED ON THE 1.6 ANGSTROMS STREPTOMYCES RUBIGINOSUS STRUCTURES WITH XYLITOL AND D-XYLOSE#X-RAY#1.60#0.135

4gna#mouse SMP30/GNL-xylytol complex#X-RAY#1.85#0.167

1xlm#D254E, D256E MUTANT OF D-XYLOSE ISOMERASE COMPLEXED WITH AL3 AND XYLITOL#X-RAY#2.40#0.179

2XIS#A METAL-MEDIATED HYDRIDE SHIFT MECHANISM FOR XYLOSE ISOMERASE BASED ON THE 1.6 ANGSTROMS STREPTOMYCES RUBIGINOSUS STRUCTURES WITH XYLITOL AND D-XYLOSE#X-RAY#1.71#0.149

1XIC#MODES OF BINDING SUBSTRATES AND THEIR ANALOGUES TO THE ENZYME D-XYLOSE ISOMERASE#X-RAY#1.60#0.152

1xig#MODES OF BINDING SUBSTRATES AND THEIR ANALOGUES TO THE ENZYME D-XYLOSE ISOMERASE#X-RAY#1.70#0.162

1xii#MODES OF BINDING SUBSTRATES AND THEIR ANALOGUES TO THE ENZYME D-XYLOSE ISOMERASE#X-RAY#1.70#0.147

5xia#STRUCTURES OF D-XYLOSE ISOMERASE FROM ARTHROBACTER STRAIN B3728 CONTAINING THE INHIBITORS XYLITOL AND D-SORBITOL AT 2.5 ANGSTROMS AND 2.3 ANGSTROMS RESOLUTION, RESPECTIVELY#X-RAY#2.50#0.135

Figure 93 : la liste de PDB « LISTEPDB.txt »

2. Liste des ligands :

Les ligands glucidique sont stocker dans un fichier Txt appeler
« LISTELIGANDS.txt ».

La liste de ligans contient le code PDB de ligand, le nom de ligand la structure et le site d'exploration de ligand.

Comme nous avons vue dans la liste PDB même cette liste est séparer par le caractère # pour le même but.

```
G6P#ALPHA-D-GLUCOSE-6-PHOSPHATE#C6 H13 O9 P#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/G6P
NAG#N-ACETYL-D-GLUCOSAMINE#C8 H15 N O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/NAG
2FP#1,6-FRUCTOSE DIPHOSPHATE (LINEAR FORM)#C6 H14 O12 P2#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/2FP
FBP#BETA-FRUCTOSE-1,6-DIPHOSPHATE#C6 H14 O12 P2#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/FBP
F6P#FRUCTOSE-6-PHOSPHATE#C6 H13 O9 P#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/F6P
MDM#METHYL-O3-(ALPHA-D-MANNOSE)-ALPHA-D-MANNOSE#C13 H24 O11#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/MDM
MMA#O1-METHYL-MANNOSE#C7 H14 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MMA
MAN#ALPHA-D-MANNOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MAN
LBT#ALPHA-LACTOSE#C12 H22 O11#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LBT
TDG#THIODIGALACTOSIDE#C12 H22 O10 S#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/TDG
LAT#BETA-LACTOSE#C12 H22 O11#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/LAT
R5P#RIBOSE-5-PHOSPHATE#C5 H11 O8 P#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/R5P
RIP#RIBOSE(PYRANOSE FORM)#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/RIP
OMK#L-RIBOPYRANOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/OMK
GAL#BETA-D-GALACTOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/GAL
NGA#BN-ACETYL-D-GALACTOSAMINE#C8 H15 N O6#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/NGA
IUG#1-(2-NITROPHENYL)ETHYL UDP-GALACTOSE#C22 H30 N3 O19 P2 (1+)#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/IUG
ARB#BETA-L-ARABINOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ARB
ARA#ALPHA-L-ARABINOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ARA
ALL#D-ALLOPYRANOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/ALL
AOS#D-ALLOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/AOS
NA1#METHYL N-ACETYL ALLOSAMINE#C9 H17 N O6#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/NA1
3BU#L-ALLOSE#C6 H12 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/3BU
A6P#6-O-PHOSPHONO-ALPHA-D-ALLOPYRANOSE#C6 H13 O9 P1#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/A6P
5RP#RIBULOSE-5-PHOSPHATE#C5 H11 O8 P#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/5RP
CAP#2-CARBOXYARABINITOL-1,5-DIPHOSPHATE#C6 H14 O13 P2#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/CAP
RUB#RIBULOSE-1,5-DIPHOSPHATE#C5 H12 O11 P2#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/RUB
XLS#D-XYLOSE (LINEAR FORM)#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XLS
XYS#XYLOPYRANOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XYS
XUL#D-XYLULOSE#C5 H10 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XUL
XYL#D-XYLITOL#C5 H12 O5#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/XYL
MBG#METHYL-BETA-GALACTOSE#C7 H14 O6#http://www.ebi.ac.uk/pdbe-srv/pdbechem/chemicalCompound/show/MBG
PAN#5-PHOSPHO-D-ARABINOHYDROXAMIC ACID#C5 H12 N O9 P#http://www.ebi.ac.uk/pdbe-
srv/pdbechem/chemicalCompound/show/PAN
```

Figure 94: Liste des ligans « LISTELIGANDS.txt. »

Section 2 :

1. XAMPP

XAMPP a pour but de construire une distribution facile à installer pour les développeurs qui souhaitent entrer dans l'univers d'Apache. Pour le rendre pratique aux yeux des développeurs, XAMPP est configuré avec toutes les fonctionnalités activées (<https://www.apachefriends.org/fr/about.html>).

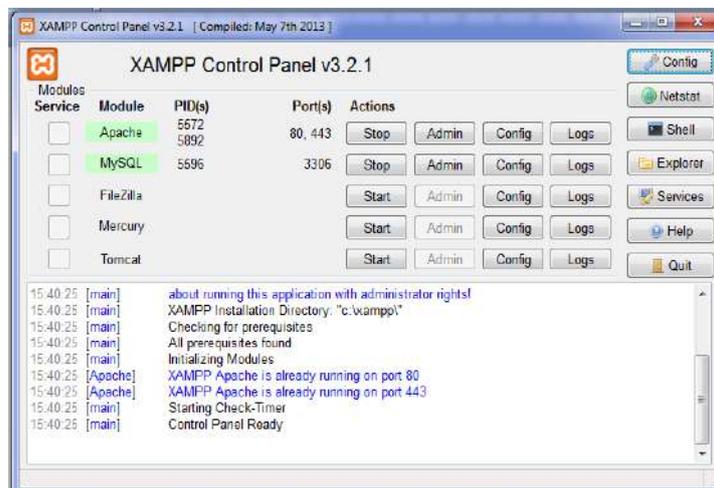


Figure 95 : Capture d'écran représentatif de XAMPP

2. Le RasMol :

qu'est-ce que le RasMol?

Disponible pour Windows, Macintosh, et UNIX, RasMol est un service gratuit, interactif spectateur moléculaire graphiques. Le programme lit dans les coordonnées du 3-D pour une molécule en utilisant le format de fichier pdb. Il affiche la molécule dans diverses représentations et permet une rotation de la molécule de manière interactive. RasMol-UCB permet la visualisation simultanée de plusieurs molécules (www.openrasmol.org).

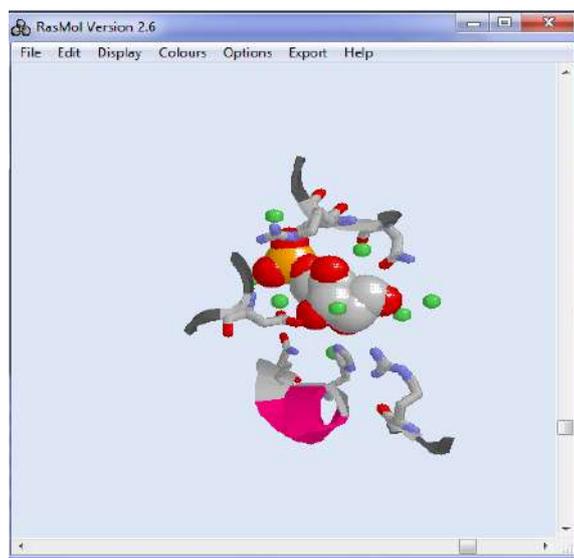


Figure 96 : Capture d'écran représentatif de RasMol.

3. Le Document Object Model

Le Document Object Model (DOM) est une interface de programmation d'application (*API*) pour valide *HTML* et bien formés *XML* documents. Il définit la structure logique des documents et la façon dont un document est accessible et manipulé. Dans la spécification DOM, le terme «document» est utilisé au sens large - de plus en plus, XML est utilisé comme un moyen de représenter différents types d'informations qui peuvent être stockées dans divers systèmes, et beaucoup de ceci serait traditionnellement être considéré comme les données plutôt que comme documents.

Néanmoins, XML présente ces données sous forme de documents, et le DOM peut être utilisé pour gérer ces données.

Avec le Document Object Model, les programmeurs peuvent créer des documents, naviguer leur structure, et d'ajouter, modifier ou supprimer des éléments et du contenu. Tout ce trouve dans un document HTML ou XML peut être consulté, modifié, supprimé ou ajouté en utilisant le Document Object Model, à quelques exceptions près - en particulier, les DOM *interfaces* pour les sous-ensembles XML internes et externes ont pas encore été spécifié (www.w3.org/TR/DOM-Level-2-Core/introduction).

Références Bibliographiques

- **Alias C., Linden G., 1994.** biochimie alimentaire. (Ed) Masson .Paris, Milan, Barcelone, 244 p.
- **Guillaume Chakroun.** Prediction de la structure d'une protéine .guillaume chakroun@hotmail.com
- **Sultanem C., 2004.** Film et bulles de protéines soluble : structure, introduction et perméabilité au gaz. Thèse de doctorat. Université paris XI, UFR scientifique d'Orsay. France, pp 3-47
- **Bensegumi A., Chikhi A., 2006.** Biochimie 1 : les protéines. (Ed) licence STE, pp 1-54
- **Frénot M., Vierling E., 2006.** biochimie des aliments diététique du dujet bien portant.(Ed) Biosciences et techniques, Doin, Cndp.Paris .pp23-109.
- <http://searchenterpriselinux.techtarget.com/definition/PHP> 25/09/2015 00 :19
- **Cheftel J. C., Cuq J. L ., Lorient D., 1985.** Protéines alimentaires : Biochimie, propriétés fonctionnelles, valeur nutritionnelle, modification chimique. (Ed) Technique et Documentation-Lavoisier. Paris, 309 p.
- <https://www.apachefriends.org/fr/about.html> 20/10/2015 13 :12
- **Boulanger P., Lemoigne M., Javillier M., Plonovoski M., Florkin M., Roche J., Wurmer R., 1959.** Traité de biochimie générale. Tome : I. Composition chimique des organismes.(Ed) Masson et Cie. Paris, 1475p.
- **Scopes R. K., 1994.** Protein purification – principales and practice. (3^{ème} Ed) Springer-Verlag. New-York, pp 71-78.
- **Bensegueni. A.** Etude théorique des métabolites secondaires des végétaux et des composés de synthèse sur le plan de l'activité biologique : simulation par docking (arrimage) moléculaire
- <http://www.wwpdb.org/> .2/11/2015. 11:41
- https://fr.wikipedia.org/wiki/Pont_disulfure 21/10/2015. 13:56
- **S. Gillet.** rappels de chimie des sucres, p-1
- **Lehninger, A.L. (1981).** Biochimie. Paris, Flammarion M S.
- **World Health Organisation ,Food Administration Organisation. (1998),** Carbohydrates in human nutrition: report of Joint FAO/WHO Expert Consultation 14-18 april 1997. FAO Food and nutrition paper n°66. Rome, WHO/FAO.
- <http://www.w3.org/TR/DOM-Level-2-Core/introduction.html> 01/10/2015 22 :13
- **Dobbing, J. (1989)** ,Dietary starches and sugars in man: a comparison. London, Springer-Verlag.
- **Abdelkrim Rachedi,** SSFS: Sequence, Structure and Function Server, 2010, <http://www.bioinformaticstools.org/ssfs/>
- **Agence Française de Sécurité Sanitaire des Aliments. (2002),** Les fibres alimentaires : définitions, méthodes de dosage, allégations nutritionnelles. Rapport du Comité d'experts spécialisé nutrion humaine du 24 septembre 2002. Maisons-Alfort, AFSSA.

- **Wagner J. R.**, Anon M. C., 1990. Influence of denaturation, hydrophobicity and sulfhydryl content on solubility and water absorbing capacity of soy protein isolate *Journal of Food Science* , 55 : 765-770.
- **Hoenger A, Bouchet-Marquis C.** Cellular tomography. *Adv Protein Chem Struct Biol.* 2011;82:67–90.
- **Altken, A** protein consensus sequence motifs. *mal biotechnol* 1999, 12:241-253
- **Pr A. Raisonier**, « Structures fonctions », 2002.
- **Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J.** (1987). *Crystallographic Databases ± Information Content, Software Systems, Scientific Applications*, edited by F. H. Allen, G. Bergerhoff & R. Sievers, pp. 107±132. Bonn: IUCr.
- <http://www.openrasmol.org/> 09/09/2015 05 :19
- **Stanislas Drey, 2010**, Les fichiers PDB ,mail,:stanislas.dorey@ens-cachan.fr
- **Catherine, Alain:** «Nucleic acids coding for novel hect domain proteins, their diagnostic and therapeutic USE », (Cabinet Harle et Phelip, 7 rue de Madrid, Paris, F-75008, FR), 02.13. (2003).
- « **Signal Transduction through Prion Protéine "**, *Science*, 15 septembre (2000)
- **P .Kaiser &al, (2000)**
- **A. MART I &al :** « Interaction between ubiquitin-protein ligase SCFSKP2 and E2F-1 underlies the regulation of E2F-1 degradation . *Nat Cell Biol* (1999).
- **DM. Koeppe & al ,(1999)**