

Thèse de Doctorat

Benoît DAVID

*Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université de Nantes
sous le sceau de l'Université Bretagne Loire*

École doctorale : Biologie Santé

Discipline : Biochimie

Spécialité : Enzymologie

Unité de recherche : UFIP, UMR CNRS 6286

Soutenue le 16 mai 2017

Conception rationnelle d'enzymes Conversion de glycoside hydrolases en transglycosidases

JURY

Président du jury	CZJZEK Mirjam , directeur de recherche, Université Sorbonne Paris 6
Rapporteurs :	Fabio STERPONE , chargé de recherche, Université Paris-Diderot 7 Richard DANIELLOU , Professeur d'université, Université d'Orléans
Examineurs :	Mirjam CZJZEK , directeur de recherche, Université Sorbonne Paris 6 Isabelle ANDRE , directeur de recherche, INSA Toulouse
Directeur de Thèse :	Charles TELLIER , Professeur d'université, Université de Nantes
Co-directeur de Thèse :	Yves-Henri SANEJOUAND , directeur de recherche, Université de Nantes

Remerciements

Mes plus grands remerciements s'adressent à Charles Tellier et Yves-Henri Sanejouand pour m'avoir accueilli dans leurs laboratoires respectifs et pour m'avoir encadré en tant que superviseurs de thèse. Ce fût un plaisir de travailler avec vous. Votre pédagogie, votre bonne humeur, votre ouverture d'esprit, votre patience (face à mes retards fréquents) ainsi que vos conseils avisés m'ont grandement aidé dans mes moments de doutes récurrents au cours de ces trois années. Un grand merci également à Bernard Offmann pour son encadrement sur le projet d'ingénierie des (trans)sialidases en première année de thèse et pour son optimisme indéfectible.

Je remercie le docteur Fabio Sterpone, le docteur Mirjam Czjzek, le docteur Isabelle André et le professeur Richard Daniellou pour avoir accepté d'être membres de mon jury de thèse ainsi que pour leur évaluation minutieuse de ce manuscrit.

J'adresse également mes remerciements aux membres de mon comité de suivi de thèse, les docteurs Marc Baaden et Marc Delarue, pour avoir accepté de juger à deux reprises mon travail au cours de mes deux premières années de thèse.

Merci à tous les membres de l'équipe Ingénierie moléculaire et glycobiochimie ainsi que de l'équipe de conception de protéines in-silico pour leur assistance technique précieuse durant mes expérimentations réelles et virtuelles. Du côté des modélisateurs, je remercie tout particulièrement Johann Hendrickx, pour son assistance sur la maîtrise de PyMOL et pour m'avoir enseigné quelques subtilités du langage de programmation Python. Je remercie également tout particulièrement Philippe Arnaud, sans qui la simulation des intermédiaires covalents n'aurait pas été possible. Du côté des expérimentateurs, je tiens tout d'abord à remercier Romain Irague et Diane Jouanneau qui ont fourni la (quasi)totalité du travail expérimental sur l'ingénierie de l'AgaD et grâce à qui j'ai pu obtenir ma première publication. Un grand merci également aux autres contributeurs de ce papier pour leur travail. Je tiens ensuite à remercier chaleureusement David Teze, Amélie Saumonneau et Marine Goux pour leur aide dans mes premiers pas au laboratoire ainsi que Claude Solleux pour m'avoir formé à la maîtrise de l'électrophorèse capillaire. J'adresse également un immense merci à mes

deux stagiaires, Mélanie Alfonso de Araujo et Maxime Chevallereau, qui ont fourni un travail remarquable dans la conception et la caractérisation de mutants de Ttβgly et de TcTS et qui m'ont ainsi énormément aidé dans l'avancée de mon projet de thèse.

Je remercie enfin tous ceux avec qui j'ai pu partager de nombreux instants de détente bien mérité hors du labo et dans le labo et qui m'ont indéniablement aidé d'une manière ou d'une autre à réguler la pression qu'impose un projet de thèse au quotidien.

Je ne peux conclure ces remerciements sans en adresser un dernier (mais pas le moindre) à ma famille et plus particulièrement à mes parents, sans qui je n'en serai pas là, et à qui je dédie ce manuscrit.

Contributions scientifiques

Publication :

David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.H., and Tellier, C. (2017); Internal Water Dynamics Control the Transglycosylation/Hydrolysis Balance in the Agarase (AgaD) of *Zobellia galactanivorans*. ACS Catalysis 3357–3367.

Posters et communications :

Turning glycoside hydrolases into transglycosidases : a theoretical study of the internal water dynamics in the *Thermus thermophilus* β -glycosidase (poster)

David, B., Sanejouand, Y.H., et Tellier, C.

ED-BS Doctoral School meeting, La Chapelle-Sur-Erdre, France, décembre 2014

A theoretical and experimental study of the internal water dynamics in the *Thermus thermophilus* β -glycosidase (présentation orale)

David, B., Sanejouand, Y.H., et Tellier, C.

Congrès GGMM, Sète, France, 25-27 mai 2015.

Turning glycoside hydrolases into transglycosidases : a theoretical and experimental study of the internal water dynamics in the *Thermus thermophilus* β -glycosidase (présentation orale)

David, B., Sanejouand, Y.H., et Tellier, C.

ED-BS Doctoral School meeting, Nantes, France, le 12 octobre 2015

Turning glycosides hydrolases into transglycosidases: a theoretical and experimental study of the internal water dynamics in β -glycosidases (poster)

David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.H., and Tellier, C.

Conférence EMBO « The chemistry and biochemistry of biocatalysis: From understanding to design », Oulu, Finlande, 12-15 juin 2016.

Converting glycosidases into transglycosidases through internal water dynamics engineering (présentation orale)

David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.H., and Tellier, C.

Réseau Français de Chimie Théorique (RFCT), Rennes, France, le 7 octobre 2016.

Conférence GlycoOuest, Nantes, France, le 9 décembre 2016

Converting a β -agarase into a β -transagarase: an experimental and theoretical study (présentation orale)

David, B., Irague, R., Jouanneau, D., Daligault, F., Czjzek, M., Sanejouand, Y.H., and Tellier, C.

Assemblée générale IDEALG, Lorient, France, le 8 novembre 2016

Conférence EnzyBio, Le Croisic, France, 9-12 mai 2017

Sommaire

LISTES DES ABREVIATIONS ET UNITES UTILISEES	13
PREAMBULE.....	16
- CHAPITRE I - INTRODUCTION GENERALE	19
1 NOTIONS DE GLYCOBIOLOGIE.....	21
1.1 ROLES DES OLIGOSACCHARIDES ET DES GLYCOCONJUGUES.....	21
1.2 SYNTHÈSE CONTRÔLÉE D'OLIGOSACCHARIDES : IMPORTANCE ET ENJEUX.....	24
1.3 SYNTHÈSE ENZYMATIQUE D'OLIGOSACCHARIDES : LES GLYCOENZYMES	24
1.3.1 Les glycosyltransférases.....	25
1.3.2 Les glycosides hydrolases	27
1.3.3 Les glycosynthases, thioglycoligases, et thioglycosynthases	32
1.4 INGÉNIERIE DES GLYCOSIDASES POUR AUGMENTER LA TRANSGLYCOSYLATION	36
2 L'EAU INTERNE AUX PROTÉINES	39
2.1 LE RÔLE DE L'EAU INTERNE DANS LA FONCTION DES PROTÉINES	39
2.2 LES CANAUX D'EAU INTERNES AUX PROTÉINES	40
2.3 L'INGÉNIERIE DES CANAUX D'EAU INTERNES AUX ENZYMES	43
2.4 MÉTHODES BIOPHYSIQUES POUR L'ÉTUDE DE LA DYNAMIQUE DE L'EAU INTERNE	44
2.4.1 La spectrométrie de masse d'échange deutérium-proton	44
2.4.2 La spectrométrie de masse à radiolyse.....	45
2.4.3 La spectroscopie infrarouge.....	46
2.4.4 La Résonance Magnétique Nucléaire en micelle inverse	47
3 MODELISATION DE L'EAU INTERNE ET SIMULATION DE SA DYNAMIQUE.....	48
3.1 PRINCIPE DE LA MÉCANIQUE MOLECULAIRE	48
3.2 MODELES THEORIQUES DE L'EAU	49
3.3 INTERACTIONS MOLECULAIRES ET CHAMPS DE FORCE	51
3.4 LA DYNAMIQUE MOLECULAIRE.....	54
3.4.1 Principe	54
3.4.2 Algorithmes d'intégration des équations du mouvement.....	56
3.4.3 Ensembles thermodynamiques	57
3.4.4 Ergodicité	57
- CHAPITRE II - MATERIEL ET METHODES.....	59
4 MATERIEL	61
4.1 LISTE DES MILIEUX, TAMPONS, ET REACTIFS UTILISES	61
4.1.1 Appareillage utilisé.....	62
4.1.2 Souches bactériennes utilisées	63
4.1.3 Plasmides.....	63
5 METHODES.....	64
5.1 METHODES EXPERIMENTALES	64
5.1.1 La PCR mutagène.....	65
5.1.2 Transformation des souches bactériennes XL1-Blue et BL21.....	67
5.1.3 Repiquage des clones et séquençage.....	67

5.1.4	Préculture d'un clone positif.....	68
5.1.5	Culture principale et induction de l'expression de la protéine	68
5.1.6	Lyse des cellules.....	69
5.1.7	Purification des protéines par chromatographie d'affinité Ni-NTA	69
5.1.8	Analyse de la pureté des fractions d'éluion par SDS-PAGE	70
5.1.9	Echange de tampon et concentration des protéines purifiées	71
5.1.10	Test d'activité enzymatique par chromatographie sur couche mince	72
5.1.11	Suivi cinétique par spectrophotométrie d'absorption	74
5.1.12	Suivi cinétique par électrophorèse capillaire.....	75
5.1.13	Détection d'oligosaccharides par analyse HPAEC-PAD	79
5.2	METHODES DE BIOINFORMATIQUE ET DE MODELISATION MOLECULAIRE	79
5.2.1	Alignement multiple de séquences de GH1.....	79
5.2.2	Analyse des séquences de GH33 trypanosomales	80
5.2.3	Simulation de dynamique moléculaire de la Ttβgly	83
5.2.4	Simulation de dynamique moléculaire de l'AgaD	84
5.2.5	Simulations de dynamique moléculaire de TcTS et TrSA.....	84
5.2.6	Paramétrage de la liaison glycosyl-enzyme de Ttβgly.....	86
5.2.7	Modification du champ de force	88
5.2.8	Construction du glucosyl-enzyme de Ttβgly	88
5.2.9	Analyse des trajectoires de simulations	89
5.2.9.1	Calcul du RMSD	89
5.2.9.2	Calcul du RMSF.....	89
5.2.9.3	Liaisons hydrogène	90
5.2.9.4	Flexibilité conformationnelle des chaînes latérales	91
5.2.9.5	Reconstruction des canaux d'eau potentiels	91
5.2.9.6	Dynamique de l'eau au sein d'un canal.....	93
5.2.9.6.1	Calcul des temps de purge d'un canal.....	93
5.2.9.6.2	Détermination de la probabilité de survie de l'eau d'un canal.....	94
5.2.9.6.3	Détermination du coefficient de diffusion de l'eau interne.....	95
- CHAPITRE III - INGENIERIE DE TTBGLY EN TRANSGLYCOSIDASE		97
6	INGENIERIE DE LA TTBGLY EN TRANSGLYCOSIDASE.....	99
6.1	PRESENTATION DU MODELE D'ETUDE.....	99
6.2	RECHERCHE DE RESIDUS MODULANT L'ACTIVITE DE TRANSGLYCOSYLATION.....	101
6.3	CARACTERISATION CINETIQUE DE LA TTBGLY SAUVAGE ET DE SES MUTANTS.....	104
6.3.1	Evaluation de la cinétique de transgalactosylation de l'ONPG	104
6.3.2	Détermination des performances catalytiques des mutants.....	107
6.3.3	Effet des mutations sur la balance H/T: hypothèses	109
6.4	ANALYSE DE LA DYNAMIQUE DE LA PROTEINE	112
6.4.1	Stabilité globale et locale du squelette peptidique.....	113
6.4.2	Flexibilité du squelette peptidique.....	116
6.4.3	Flexibilité des chaînes latérales de résidus en contact avec le substrat	117
6.4.4	Flexibilité des chaînes latérales des résidus W120 et H178	123
6.4.5	Conclusions	125
6.5	ANALYSE DE LA DYNAMIQUE DE L'EAU INTERNE A LA PROTEINE	126
6.5.1	Diffusion de l'eau interne au sein de la protéine.....	126
6.5.2	Identification de canaux d'eau internes à la protéine	129
6.5.3	Comparaison aux précédents résultats publiés.....	134

6.5.4	Identification de l'origine des molécules d'eau catalytiques.....	135
6.5.5	Purges des canaux d'eau desservant le sous-site -1	140
6.5.6	Résidence de l'eau des canaux desservant le sous-site -1	142
6.5.7	Conclusions.....	143

- CHAPITRE IV - INGENIERIE DE L'AGAD EN TRANS-AGARASE 145

7	INGENIERIE D'AGAD EN TRANS-AGARASE	147
7.1	PRESENTATION DU MODELE D'ETUDE.....	147
7.2	ORGANISATION DE L'EAU DANS LA STRUCTURE CRISTALLOGRAPHIQUE D'AGAD.....	149
7.3	MUTAGENESE DE RESIDUS BORDANT LE CANAL.....	152
7.3.1	Mutagenèse des résidus Q342 et Y181	153
7.3.2	Mutagenèse des résidus D341 et S351	155
7.3.3	Conclusions.....	157
7.4	ANALYSE DE LA DYNAMIQUE DE LA PROTEINE	159
7.4.1	Stabilité globale et locale du squelette peptidique.....	159
7.4.2	Flexibilité du squelette peptidique.....	162
7.4.3	Flexibilité de la chaîne latérale du résidu acide-base catalytique.....	164
7.4.4	Flexibilité des chaînes latérales d'autres résidus du site -1	165
7.4.5	Caractérisation des liaisons H stabilisant l'acide-base catalytique	170
7.4.6	Conclusions.....	172
7.5	ANALYSE DE LA DYNAMIQUE DE L'EAU INTERNE A LA PROTEINE	173
7.5.1	Diffusion de l'eau interne au sein de la protéine.....	173
7.5.2	Identification des canaux et poches d'eau internes à la protéine.....	175
7.5.3	Purges du canal d'eau caractérisé.....	179
7.5.4	Résidence de l'eau au sein du canal	180
7.5.5	Conclusions.....	181

- CHAPITRE V - GENERALISATION DE L'ETUDE AU GH DE LA FAMILLE 33..... 183

8	GENERALISATION DE L'ETUDE AUX ENZYMES DE LA FAMILLE DES GH33	185
8.1	SIALIDASES ET TRANS-SIALIDASES DE LA FAMILLE 33.....	185
8.2	SIALIDASE ET TRANS-SIALIDASE DES TRYPANOSOMES <i>T. RANGELI</i> ET <i>T. CRUZI</i>	185
8.3	INGENIERIE DE LA SIALIDASE DE <i>T. RANGELI</i> EN TRANS-SIALIDASE	188
8.4	ANALYSE DE LA DYNAMIQUE DES STRUCTURES PROTEIQUES DE TcTS ET TrSA	189
8.4.1	Stabilité et flexibilité du squelette peptidique	190
8.4.2	Flexibilité des chaînes latérales de la triade catalytique	192
8.5	ANALYSE DE LA DYNAMIQUE DE L'EAU INTERNE A TcTS ET TrSA.....	194
8.5.1	Identification des canaux et poches d'eau internes à la protéine.....	194
8.5.2	Etude du canal 1.....	196
8.5.3	Etude du canal 2.....	198
8.5.4	Etude du canal 3.....	201
8.5.5	Diffusion de l'eau interne au domaine catalytique.....	203
8.6	CONCLUSIONS	206
8.7	IDENTIFICATION DE NOUVEAUX RESIDUS CONTROLANT LA BALANCE H/T	206
8.7.1	Recherche de résidus fonctionnels en bordure des canaux d'eau.....	207
8.8	MUTAGENESE DES RESIDUS S115, S118, ET L313	211
8.8.1	Caractérisation des produits de réaction par chromatographie.....	212

8.8.2 Détermination des constantes cinétiques de la réaction.....	213
8.9 VERS LA RECONSTRUCTION DE L'ANCETRE COMMUN A TCTS ET TRSA.....	214
8.10 CONCLUSIONS ET PERSPECTIVES	219
CONCLUSION GENERALE ET PERSPECTIVES.....	221
ANNEXE	227
BIBLIOGRAPHIE	235
LISTE DES ILLUSTRATIONS	252

Listes des abréviations et unités utilisées

Réactifs chimiques divers :

ANTS	Acide 8-aminonaphthalène-(1,3,6)-trisulfonique
ADN	Acide désoxyribonucléique
dATP	2-Désoxyadénosine triphosphate
dCTP	2-Désoxycytidine triphosphate
dGTP	2-Désoxyguanosine triphosphate
DMSO	Diméthyl sulfoxyde
PMSF	Fluorure de phénylméthylsulfonyle
IPTG	Isopropyl β -D-1-thiogalactopyranoside
NTA	Acide nitriloacétique
BSA	Sérum Albumine Bovine

Substrats et produits de réaction enzymatique:

ONP	Ortho-Nitrophénol ou 2-Nitrophénol
Sialyl-PNP	4-Nitrophényl-sialyl
ONPG	2-Nitrophényl- β -D-galactopyranoside
LNT	Lacto-N-tétraose
3'SL	3'sialyl-lactose
ONPGG / ONPG-Gal (LA-G) ₅	2-Nitrophényl- β -D-galactopyranosyl- β (1,3)-galactopyranoside Agaro-décaose

Techniques:

HPAEC-PAD	High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection
PCR	Polymerase Chain Reaction
CCM	Chromatographie sur couche mince
SDS-PAGE	Sodium dodecyl sulfate gel electrophoresis
DXMS	Deuterium eXchange Mass Spectrometry
RMN	Résonance magnétique nucléaire
MD	Dynamique moléculaire
MSA	Alignement multiple de séquences

Métriques et unités:

D.O.	Densité optique
kb	Kilobase
kDa	KiloDalton
kV	Kilovolt
mM	Millimolaire
nm	Nanomètre
fs	Femtoseconde
Å	Angström
MSD	Mean Square Displacement
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
qsp	quantité suffisante pour
v/v	volume/volume
k_{cat}	Constante catalytique
Km	Constante de Michaëlis
Vm	Vitesse maximale
rpm	Révolutions par minute
V_i	Vitesse initiale
V_H/V_T	Vitesse initiale d'hydrolyse / vitesse initiale de transglycosylation
MWCO	Molecular Weight CutOff

Enzymes et réactions:

CAZy	Carbohydrate Active enZymes
GH	Glycoside hydrolases
GH 1	Glycoside hydrolases de la famille 1
TcTS	<i>Trypanosoma cruzi</i> trans-sialidase
TrSA	<i>Trypanosoma rangeli</i> sialidase
Ttβgly	<i>Thermus thermophilus</i> β-glycosidase
Tmβgly	<i>Tenebrio molitor</i> β-glycosidase
Tmaβgly	<i>Thermotoga maritima</i> β-glycosidase
Abg	<i>Agrobacterium</i> β-glycosidase
ENGases	Endoglucanases

T/H Transglycosylation/Hydrolyse
WT Wild type (sauvage)

Autres:

LB Lysogeny Broth
NAD Nicotinamide Adénine dinucléotide
Liaisons H Liaisons hydrogène
PDB Protein Data Bank
OW Atome d'oxygène de molécule d'eau
AB Acide-Base
Nuc Nucléophile

Préambule

L'emploi de glycoside hydrolases (GH) mutantes à fort rendement de transglycosylation se présente comme une alternative intéressante pour la synthèse d'oligosaccharides complexes. Cependant, les bases moléculaires régissant l'équilibre entre ces deux activités restent difficiles à élucider. Le projet de thèse vise à aboutir à une meilleure compréhension des bases moléculaires gouvernant la balance entre l'activité d'hydrolyse et de transglycosylation chez quatre glycosides hydrolases modèles incluant deux GH des familles 1 (GH1) et 16 (GH16), ainsi qu'une paire de GH homologues de la famille 33 (GH33). Il s'inscrit donc dans la continuité du travail de Teze et coll.^{104,117} avec pour objectif de démontrer un possible rôle fonctionnel des canaux d'eau dans la modulation des activités d'hydrolyse et de transglycosylation chez les glycosides hydrolases. Le travail effectué dans le cadre de ce projet de thèse sera décrit en cinq chapitres distincts :

Le premier chapitre consistera en une introduction générale du manuscrit basée sur une étude bibliographique et sera divisé en trois parties. Il soulignera dans une première partie l'importance fonctionnelle des glucides et plus spécifiquement des oligosaccharides dans leur contexte biologique. Les voies de synthèse enzymatiques d'oligosaccharides seront également évoquées et l'accent sera mis sur l'utilisation des GH à des fins de synthèse. Cette première partie se conclura sur la présentation des stratégies d'ingénierie permettant d'augmenter l'activité synthétique (transglycosylation) des GH. La seconde partie de cette introduction se focalisera dans un premier temps sur le rôle de l'hydratation interne dans l'activité et la stabilité des protéines. L'importance fonctionnelle des canaux d'eau internes au sein des enzymes et plus particulièrement au sein des hydrolases sera largement décrite. Différents exemples d'ingénierie rationnelle visant à moduler l'activité de ces enzymes en modifiant la structure de ces canaux seront ensuite détaillés. Cette seconde partie se conclura sur une présentation de quelques méthodes analytiques permettant d'identifier les sites d'hydratation internes aux protéines et d'étudier leur dynamique. La troisième et dernière partie de cette introduction se concentrera sur la modélisation *in-silico* de l'eau et de sa dynamique. Une brève description des différents modèles physiques de l'eau et de ses interactions

intra/intermoléculaires sera tout d'abord présentée. Cette introduction générale se conclura sur une présentation du principe de la méthode de simulation de dynamique moléculaire utilisée dans ces travaux.

Le second chapitre sera consacré à la présentation des méthodes expérimentales et bioinformatiques utilisées. Un des principaux défis principaux soulevés par ce projet a été de déterminer une méthode de recherche systématique des canaux d'eau à partir d'une trajectoire de dynamique moléculaire donnée. De nombreux programmes informatiques visant à prédire la trajectoire des canaux internes aux protéines sont désormais disponibles ^{183,184,185}. Bien qu'ayant démontré leur efficacité dans le cadre de nombreux travaux, ces méthodes complexes tendent cependant à proposer un nombre exhaustif de solutions dont certaines peuvent parfois s'avérer physiquement peu probables ¹⁷¹. Basant leurs prédictions sur le calcul de volumes accessibles au sein d'une structure protéique donnée, les algorithmes sous-jacents ne permettent également pas d'isoler spécifiquement et de manière explicite les molécules d'eau prenant part à la formation de canaux au sein d'une protéine donnée. Afin de pallier aux limitations de ces méthodes, un algorithme de reconstruction des canaux d'eau à partir des données de simulation a donc été développé et sera présenté. Diverses analyses seront également évoquées.

Le troisième chapitre sera dédié à l'application de ces méthodes pour l'ingénierie de la Tt β gly en transglycosidase. Les travaux de Teze et coll. ¹⁰⁴ ont montré que certaines mutations favorisant de manière significative l'activité de transglycosylation chez Tt β gly impliquaient des résidus (F401 et N282) distants ($\sim 8 \text{ \AA}$) de la liaison glycosyl-enzyme. La résolution cristallographique des mutants correspondants n'a cependant pas permis d'élucider la contribution des résidus mutés dans l'activité de transglycosylation puisque aucune différence structurale significative par rapport à l'enzyme sauvage n'a pu être détectée. Ces résultats laissent alors supposer qu'une modification du repositionnement ou de l'acheminement des molécules d'eau impliquées dans l'activité d'hydrolyse pourraient être induite par la mutation de positions distantes du site catalytique.

Par ailleurs, de courtes simulations de dynamique moléculaire effectuées sur la structure cristallographique de la Tt β gly combinées à une analyse de données de Deuterium eXchange Mass Spectrometry (DXMS) ont révélé d'un canal d'eau communiquant avec le site actif¹¹⁷. Une analyse de la conservation des molécules d'eau cristallographiques chez un ensemble d'homologues issus de la même famille a également montré que ce canal contenait plusieurs molécules d'eau très conservées¹¹⁷. En vue de confirmer les résultats précédents tout en réalisant une simulation plus proche du temps fonctionnel des GH (~ 1 ms), une simulation de la dynamique moléculaire de l'intermédiaire covalent de la Tt β gly a été réalisée sur une durée de 500 ns. Sur la base d'alignements de séquences, des expériences de mutagenèse sur les résidus bordant le canal ont conjointement été effectuées afin de caractériser des résidus potentiellement impliqués dans la modulation de la balance H/T chez Tt β gly.

Teze et coll. ont suggéré que ce canal pourrait être impliqué dans la fonction des glycosides hydrolases de la famille 1. D'autres études ont également suspecté la contribution probable de canaux d'eau dans la modulation de l'activité hydrolytique des GH1¹⁶⁹, des GH68¹⁵⁹ et des GH117¹⁵⁸. Il est donc tentant de supposer que la présence persistante de canaux d'eau au sein de GH de familles distinctes pourrait être essentielle au transport de l'eau catalytique au site catalytique. Afin de vérifier cette hypothèse, les analyses théoriques menées sur la β -glycosidase de *Thermus thermophilus* (Tt β gly) ont été étendues à la β -agarase de *Zobellia Galactinovorans* (AgaD) ainsi qu'à la transsialidase de *T. cruzi* (TcTS). Issue respectivement de la famille des GH 16 et 33, ces GH ont été particulièrement étudiées au laboratoire. Les analyses de ces simulations ainsi que les résultats des travaux expérimentaux associés seront décrites dans les quatrième et cinquième chapitres.

Une conclusion générale résumant les principaux résultats, suivie d'une présentation des perspectives possibles ouvertes par ces travaux seront finalement présentées.

- Chapitre I -

Introduction générale

1 Notions de glycobiologie

1.1 Rôles des oligosaccharides et des glycoconjugués

Les glucides constituent une des quatre classes essentielles de biomolécules. Ubiquitaires dans le règne du vivant, on les rencontre essentiellement sous la forme de polymères dans la nature¹. Les oligomères et polymères de glucides (oligo/polysaccharides) résultent de l'enchaînement séquentiel d'unités monosaccharides liées entre elles par des liaisons glycosidiques. La présence d'hydroxyles libres liés aux atomes de carbone du cycle de chaque monosaccharide lui assure de multiples possibilités d'assemblages modulaires avec d'autres sous-unités saccharidiques par la formation d'une liaison osidique. De nombreux facteurs incluant entre autres la nature et l'orientation variable des substituants de part et d'autre de la liaison glycosidique, le nombre de branchements ainsi que la stéréochimie de la liaison sont à l'origine de l'importante hétérogénéité structurale des oligo/polysaccharides². Leur diversité structurale s'accompagne par ailleurs d'une grande diversité de fonctions biologiques. Ceci a notamment conduit à formuler l'hypothèse d'un glyco-code décrivant une relation liant la structure et l'agencement des oligo/polysaccharides à leur fonction biologique respective^{3,4}. Au sein des parois végétales et microbiennes, les oligo/polysaccharides remplissent essentiellement un rôle structural (hémicellulose, peptidoglycane, pectines, chitine). Dans un contexte métabolique, on les rencontre principalement sous la forme de composés énergétiques (nucléosides tri-phosphatés) ou de molécules de stockage (glycogène, amidon).

A l'état libre au sein de nombreux liquides physiologiques (sang, lait maternel⁵...), une grande majorité d'oligosaccharides nécessitent cependant d'être associés à d'autres biomolécules (métabolites, protéines, lipides) afin de remplir une fonction spécifique. Une importante variété de métabolites secondaires (antibiotiques⁶, glycoflavonoïdes⁷, glycoalcaloïdes⁸...) est par exemple directement impliquée dans les réactions de défenses de nombreux organismes (micro-organismes, plantes supérieures...) en réponse aux stress environnementaux^{9,10}. Essentiellement conjugués aux protéines^{1,11}, on les retrouve principalement liés aux hydroxyles des sérines et thréonines (*O*-glycosylation) ainsi qu'au groupement amide des asparagines (*N*-glycosylation) et plus

rarement associés à d'autres aminoacides¹² (arginines, tyrosines, tryptophanes...) (Figure 1). On estime actuellement qu'entre 50 et 70 % des protéines humaines sont glycosylées¹¹.

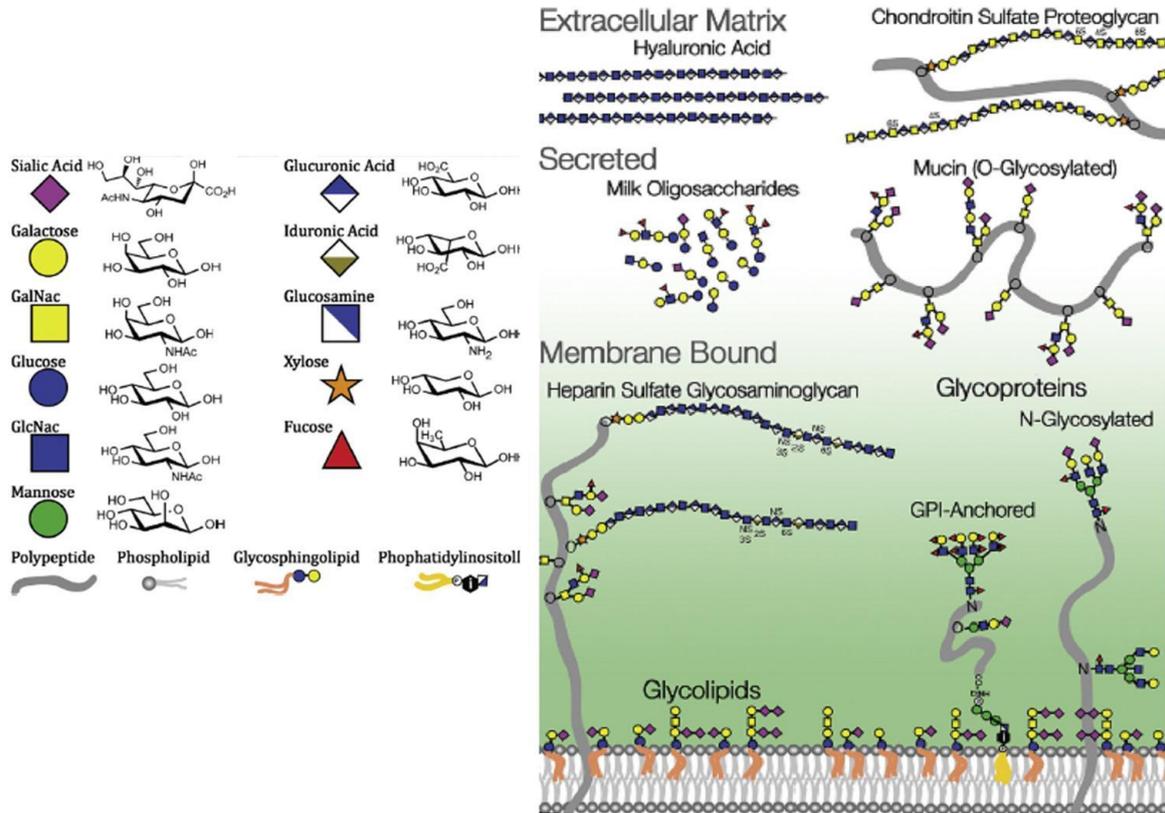


Figure 1. Exemples d'oligosaccharides rencontrés dans le monde vivant¹.

Les oligosaccharides conjugués aux macromolécules biologiques peuvent remplir un rôle essentiel dans de nombreux processus de signalisation¹³ (Figure 2). La glycosylation des protéines en cours de maturation au sein du réticulum endoplasmique des cellules eucaryotes joue par exemple le rôle de signal régulant le repliement et l'adressage des protéines au sein de compartiments cellulaires spécifiques². Au sein des matrices extracellulaires et des membranes, les glycoprotéines et les glycolipides sont des acteurs majeurs des processus de communications chimiques entre le milieu intracellulaire et extracellulaire associés à de nombreux événements biologiques (adhésion cellulaire¹⁴, réponses immunitaires¹³, fécondation¹⁵, interactions hôte-pathogènes¹⁶...). Constituants majeurs du glycocalice des cellules animales, les protéines porteuses de glycosaminoglycanes sont ainsi directement impliquées dans les processus de migration cellulaires indispensables aux processus de morphogenèse

(hyaluronate) et de réparation tissulaire (héparanes, chondroïtine sulfate) *via* leurs interactions avec de nombreux facteurs de signalisation (facteurs de croissance, cytokines, chimiokines...) ^{2,17}.

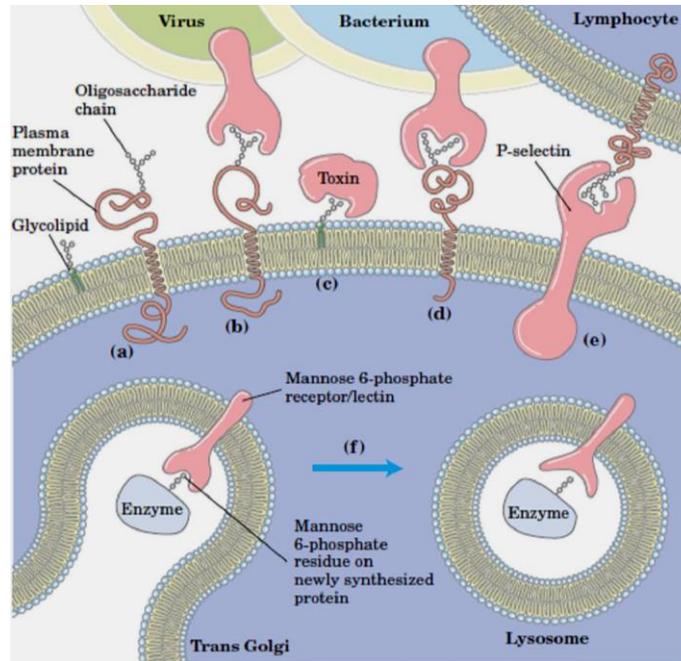


Figure 2. Quelques rôles majeurs exercés par les glycoprotéines ².

Au moins 60% des protéines thérapeutiques commercialisées à ce jour sont également glycosylées ¹⁸. Elles regroupent de nombreuses protéines sécrétées ¹⁹ incluant principalement des immunoglobulines, des interférons, des enzymes (ribonucléase, lysozyme...), ainsi que diverses hormones (lutéinisante, folliculo-stimulante, thyroïdienne...). La nature des oligosaccharides qui leur sont liés est connue pour affecter de manière significative leur demi-vie sérique, leur solubilité, leur mode d'action, ainsi que leur immunogénicité ^{18,20}. L'importante diversité de pathologies (diabète, cancers, dystrophies musculaires, troubles neurologiques, maladies auto-immunes...) résultant de la dérégulation des voies de glycosylation chez l'Homme illustre l'importance thérapeutique de cette modification post-traductionnelle ^{14,21,22}. Ces pathologies résultent en général de l'accumulation de mutations au sein de gènes codant des transporteurs ou des enzymes impliquées dans la synthèse ²² et la dégradation ^{23,24} des oligosaccharides. Communément appelés glycoenzymes, ces enzymes régissent la synthèse, le remodelage, ainsi que la conjugaison des oligosaccharides aux macromolécules biologiques.

1.2 Synthèse contrôlée d'oligosaccharides : Importance et enjeux

En dépit de leur importance biologique, la plupart des oligosaccharides et des glycoconjugués ne peuvent être obtenus qu'en faibles quantités à partir de sources biologiques. La purification difficile de ces composés à partir de l'extrait brut, associée au risque possible de contamination par des agents infectieux, rend cette approche peu adaptée pour l'obtention de ces molécules en grande quantités et à coût modéré. La conception de stratégies de synthèse a donc été nécessaire afin de répondre à ces besoins. Initialement développées, les méthodes de synthèse chimique^{4,25} sont toujours massivement utilisées aujourd'hui. La présence de nombreux groupements hydroxyles sur le cycle osidique d'une part, et d'autre part leur faible réactivité implique dans ces méthodes l'utilisation de groupements protecteurs en cours de synthèse⁴. L'addition et l'élimination sélective de ces groupements permet alors de contrôler la régiosélectivité de la synthèse de la séquence oligosaccharidique.

Initiée par Koenigs et Knorr en 1901²⁶, la synthèse chimique d'oligosaccharides a depuis bénéficié de l'acquisition de nouvelles connaissances en chimie et des dernières avancées technologiques. Le développement de stratégies de synthèse automatisées a ainsi permis de diminuer le temps nécessaire à l'assemblage contrôlé d'oligosaccharides à partir de monomères fonctionnalisés²⁷. Le coût important associé à l'obtention et à la maintenance d'un tel équipement n'est cependant pas toujours envisageable. L'utilisation alternative des méthodes de synthèse multi-étapes en phase homogène est également limitée par les pertes de matière occasionnées par les purifications successives des composés intermédiaires formés à chaque étape. Afin de pallier à ces limitations, les méthodes de synthèse chimique peuvent être remplacées et/ou complémentées par l'utilisation d'enzymes.

1.3 Synthèse enzymatique d'oligosaccharides : les glycoenzymes

Dans la nature, les enzymes catalysant la synthèse et la dégradation des composés glycosidiques appartiennent essentiellement aux familles des glycosyltransférases et des glycosides hydrolases et des polysaccharides lyases. L'utilisation de ces catalyseurs biologiques en synthèse offre de nombreux avantages par rapport aux méthodes

chimiques traditionnelles. Capables de catalyser la formation d'une liaison glycosidique dans des conditions physico-chimiques douces, la catalyse enzymatique assure également un contrôle fin de la régiosélectivité et de la stéréosélectivité de la réaction ²⁸. Les propriétés synthétiques des glycoenzymes permettent donc de s'affranchir de l'emploi de groupements protecteurs et de solvants organiques et offrent donc une alternative intéressante à la synthèse chimique traditionnelle.

1.3.1 Les glycosyltransférases

Les glycosyltransférases (GT) de type Leloir sont des enzymes ubiquitaires regroupées à ce jour en 145 familles distinctes dans la base de données CAZY (Carbohydrate Active enZymes) ²⁹. Ces enzymes sont connues pour catalyser la plupart des réactions de glycosylation trouvés dans la Nature ³⁰. La catalyse enzymatique implique l'intervention d'un sucre donneur activé (via un phosphate, un pyrophosphate, ou un nucléoside phosphate) à partir duquel sera transférée la partie osidique du donneur vers un accepteur spécifique (oligosaccharide, protéine, lipide, petite molécule organique...). Deux mécanismes catalytiques dits de rétention (*Figure 3*) et d'inversion (*Figure 4*) de configuration ont été à ce jour répertoriés ³¹. Ils se distinguent l'un de l'autre par la stéréochimie de configuration du carbone anomérique du sucre donneur qui peut être conservée ou inversée lors de sa conversion en produit de réaction.

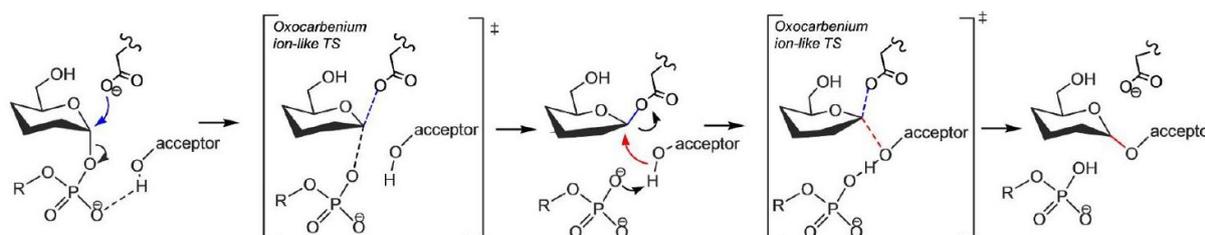


Figure 3. Mécanisme à rétention de configuration des GT ³¹.

Le mécanisme de la réaction d'inversion de configuration repose sur l'attaque nucléophile de l'hydroxyle du sucre accepteur sur le carbone asymétrique du sucre donneur. Cette attaque s'accompagne du départ concerté du groupement nucléoside selon une substitution nucléophile de type 2. Le départ du groupe partant est facilité par le caractère basique du résidu catalytique ainsi que par son activation par un acide de

Lewis (cation divalent) favorisant la formation d'un état de transition de type oxocarbenium ³².

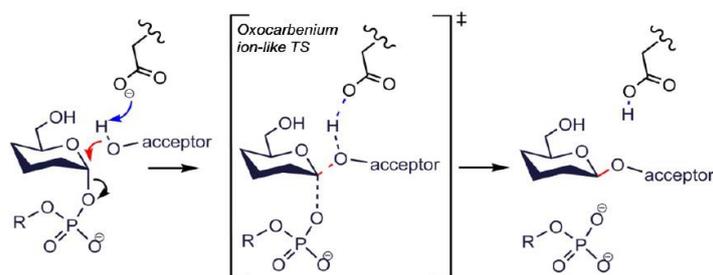


Figure 4. Mécanisme à inversion de configuration des GT ³¹.

Un mécanisme à double déplacement régissant le transfert de glycosyle avec rétention de configuration a tout d'abord été proposé. Ce mécanisme implique l'intervention d'un résidu catalytique du type nucléophile devant réaliser une attaque nucléophile du carbone asymétrique du substrat pour aboutir à la formation d'un intermédiaire covalent glycosyl-enzyme. L'attaque ultérieure de cet intermédiaire par un accepteur nucléophile conduirait alors à un produit de configuration identique à celle du substrat. Il a été cependant remarqué qu'un nombre non-négligeable de GT à rétention de configuration ne possédaient pas de résidu nucléophile conservé au sein de leur site catalytique et que le remplacement de résidus proposés comme catalytiques nucléophiles par une alanine n'a pas toujours permis d'abolir l'activité de transfert. Une analyse systématique de la géométrie des substrats au sein des sites actifs des structures de GT répertoriées dans la PDB tend finalement à proposer l'hypothèse d'un mécanisme de déplacement orthogonal à une seule étape ³³.

Les glycosyltransférases présentent un intérêt important de par leur capacité à synthétiser un grand nombre de composés naturels glycosylés ayant des propriétés pharmacologiques et/ou nutraceutiques ^{6,7,34,35}. Souvent difficiles à produire et à purifier, ces enzymes ont néanmoins été utilisées pour la synthèse de glycoconjugués par fermentation. L'ingénierie des voies de glycosylation de certaines plantes ^{36,37,38}, bactéries ^{39,40,41}, et levures ⁴² *via* l'introduction de gènes de glycosyltransférases exogènes ou la mutation de gènes endogènes s'est par ailleurs illustrée comme une approche prometteuse pour la production de biopharmaceutiques glycosylés et de glycoside activés ³⁵. L'emploi de cofacteurs activés indispensables à leur activité peut

cependant constituer une limitation à une synthèse à grande échelle en raison du coût lié à leur production et à leur recyclage⁴³. En raison de leurs capacités synthétiques ainsi que de leur plus grande stabilité *in-vitro* en comparaison aux glycosyltransférases, certaines glycosides hydrolases peuvent ainsi se présenter comme une alternative intéressante à celles-ci dans le cadre d'applications en synthèse.

1.3.2 Les glycosides hydrolases

Les glycosides hydrolases (GH ou glycosidases) sont les principales enzymes (avec dans une moindre mesure les polysaccharides phosphatases⁴⁴, lyases⁴⁵ et monooxygénases⁴⁶) impliquées dans la dégradation des oligo/polysaccharides au sein des organismes vivants. Ubiquitaires, ces enzymes sont regroupées en 133 familles au sein de la classification CAZy et seraient codées en moyenne par 1% du génome des organismes les exprimant⁴⁷. De même que les glycosyltransférases, la catalyse de la réaction hydrolytique peut suivre un mécanisme à rétention (*Figure 5*) ou à inversion (*Figure 6*) de configuration anomérique⁴⁸. Les deux mécanismes font en revanche dans la plupart des cas intervenir deux résidus catalytiques acides (souvent glutamate ou aspartate)⁴⁹.

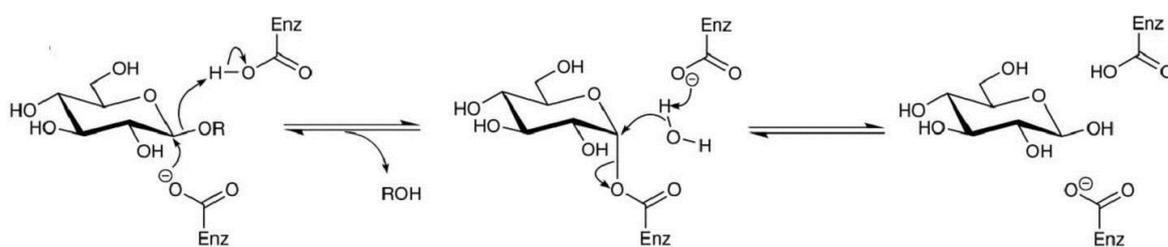


Figure 5. Mécanisme catalytique à rétention de configuration des GH⁵⁰.

Dans le mécanisme à inversion de configuration, un résidu joue le rôle de base en activant une molécule d'eau par déprotonation. Rendue particulièrement nucléophile, celle-ci va pouvoir réaliser une attaque concertée sur le centre anomérique du cycle osidique. La protonation simultanée de l'oxygène anomérique par le second résidu catalytique (acide) va également faciliter l'hydrolyse de la partie aglycone du substrat (groupe partant).

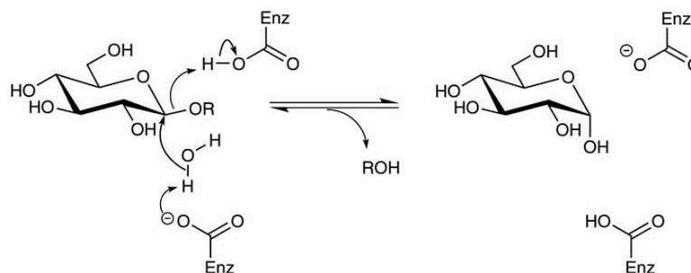


Figure 6. Mécanisme catalytique à inversion de configuration des GH ⁵⁰.

Le mécanisme à rétention de configuration comporte en revanche deux étapes impliquant la formation d'un intermédiaire covalent. Lors de la première étape, l'un des deux résidus réalise une attaque nucléophile sur le centre anomérique du glycosyl donneur pour former un intermédiaire covalent glycosyl-enzyme. Le second résidu catalytique joue dans un premier temps le rôle d'acide de Brønsted en protonant l'oxygène anomérique impliqué dans la liaison glycosyl-enzyme pour faciliter le départ du groupe partant. Possédant un caractère basique à l'état déprotoné, ce même résidu va pouvoir dans un second temps activer une molécule d'eau par déprotonation. L'attaque nucléophile de la liaison glycosyl-enzyme par cette molécule d'eau activée entraînera finalement l'hydrolyse. Dépourvues de résidu acide-base catalytique, les myrosinases représentent une exception au sein de la famille 1. Ces enzymes sont ainsi dépendantes d'un coenzyme, l'ascorbate, qui joue le rôle de base catalytique au cours de la réaction de déglycosylation ⁵¹.

Une variation du mécanisme classique proposé par Koshland pour les GH à rétention de configuration a cependant pu être observé dans le cas des *N*-acetyl- β -hexosaminidases appartenant aux familles 18, 20, 25, 56, 84, et 85 (*Figure 7*). Ces enzymes étant privées de nucléophile catalytique, ce rôle est joué par le groupe 2-acetamido du substrat. Le départ du groupe partant s'accompagne donc d'une attaque nucléophile intramoléculaire de l'oxygène du groupe *N*-acétamide sur le carbone anomérique provoquant la formation d'un ion oxazolinium. Stabilisé par un carboxylate vicinal, cet état de transition est finalement clivé pour conduire à la formation d'un produit d'hydrolyse avec rétention de configuration ⁵².

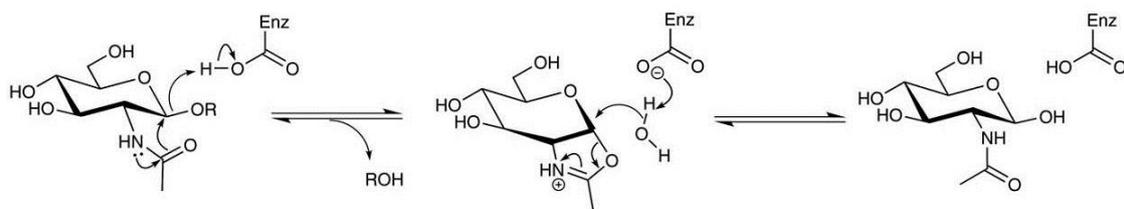


Figure 7. Mécanisme catalytique des *N*-acetyl- β -hexosaminidases avec assistance par le substrat ⁵⁰.

Enfin, un mécanisme alternatif a également été décrit pour les familles 4 ⁵³ et 109 ⁵⁴. Plus rare, il fait intervenir le cofacteur NAD⁺ qui est impliqué dans l'oxydation du substrat de départ. Générant un intermédiaire insaturé, cette étape d'oxydation est ensuite suivie de l'élimination de la partie aglycone du substrat.

La plupart des GH à rétention de configuration présentent également la particularité de posséder une activité de transglycosylation coexistant le plus souvent avec l'activité d'hydrolyse ⁵⁵. Suivant le même mécanisme, la réaction de transglycosylation a lieu lorsqu'un accepteur organique (type alcool) entre en compétition avec l'eau catalytique et réalise une attaque nucléophile sur l'oxygène anomérique pour former une nouvelle liaison glycosidique (*Figure 8*).

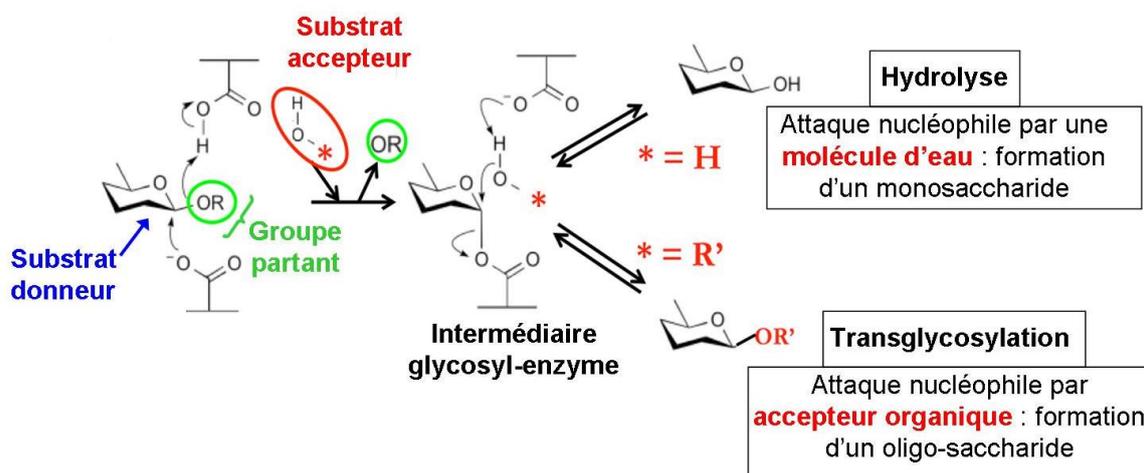


Figure 8. Mécanisme réactionnel de l'hydrolyse et de la transglycosylation.

R et *R'* désignent les parties aglycone du groupe partant et de l'accepteur.

Si la grande majorité des GH catalysent préférentiellement l'hydrolyse, la transglycosylation est dans certains cas l'activité biologique prédominante des GH. Présentes au sein de diverses familles de GH (*Tableau 1*), les transglycosidases

possèdent une activité principale de type glycosyltransférase (de type non-Leloir) au détriment d'une activité hydrolytique réduite ⁵⁶.

Famille	Clan	Transglycosylases	Structure
GH1	A	Os9BGlu31 ⁶⁸ , AA5GT ⁶⁷ , AA7GT ⁶⁷ , SFR2 ⁶⁶	-
GH5	A	mannan transglycosylase	1RH9
GH10	A	xylan endotransglycosylase	-
GH12	C	xylan endotransglycosylase	-
GH13	H	CGTase	1D3C
		Isomaltutose synthase	1CXL
		Trehalose synthase	4JCM
		Amylosucrase ⁶⁵	3BMW
		α -(1,4)-Glucan branching enzyme	IV8
		α -(1,3)-Glucan synthase	1G5A ⁵⁷
		Maltose- α -glucosyl transferase	4LXF
Maltooligosyltrehalose synthase	1ZJA,3AML		
GH16	B	XET ⁶² , Crh1 & Crh2 ⁶⁹ chitin- β (1,6)-glucanosyltransferase	1UMZ
GH17	A	β (1,3)-glucanosyltransglycosylase	4WTP ⁶³
GH23		lytic transglycosylase	-
GH27	D	Galactan	-
		Galactosyltransferase	-
GH31	D	Oligosaccharide- α (1,4)-glucosyltransferase isomaltosyltransferase	4BA0 ⁶⁰
GH32	J	sucrose 1-fructosyltransferase ⁶⁴	3UGF
		fructan 6G-fructosyltransferase ⁶⁴	
GH33	E	Trans-sialidases	2JKB
			1MS3
			1S0I
GH36	D	Stachyose synthase,	-
		Raffinose synthase ⁵⁸	-
GH57	(β/α) ₇	glycogen branching enzyme	3P0B,3N98
GH65	L	phosphate β -glucosyltransferase	4KTP
GH66		Cycloisomaltoseoligosaccharide	3WNO
GH68	J	Levan sucrose ⁶⁴	1PT2
GH70	H	Dextran sucrose ⁶⁵	3KLL
GH72	A	β -(1,3)-glucanosyltransglycosidase	2W61 ⁶¹
GH77	H	Amylomaltase	1ESW
		4- α -glucanotransferase	1X1N
GH102			-
GH103		peptidoglycan lytic transglycosylase ⁵⁹	1QUS
GH104			-

Tableau 1. Exemples (liste non-exhaustive) de transglycosidases répertoriées dans CAZY.

Dans la Nature, certaines transglycosidases catalysent la glycosylation d'accepteurs très variés dans le cadre de nombreux processus biologiques essentiels. Chez *Arabidopsis thaliana*, la β -transglycosidase SFR2 est impliqué dans le remodelage de la composition pariétale en galactolipides en réaction au gel ⁶⁶. D'autres espèces de plantes, *Dianthus caryophyllus* et *Delphinium grandiflorum*, ont recours à une transglycosidase pour la synthèse des anthocyanes ⁶⁷. Chez le riz, la transglucosidase Os9BGlu31 est essentielle à la glycosylation des phénylpropanoïdes et des phytohormones en général ⁶⁸. Enfin, d'autres transglycosidases intervenant dans la synthèse de la paroi cellulaire chez la levure ⁶⁹, les plantes ⁷⁰, les bactéries ⁷¹, ou dans

l'édition des ARN chez tous les êtres vivants ⁷² ont également été caractérisées. Contrairement aux glycosyltransférases, ces enzymes présentent en général une spécificité large pour l'accepteur et sont donc particulièrement attractives pour de nombreuses applications synthétiques. Capables d'utiliser un donneur naturel peu cher et très abondant, le glucopyranosyl-1,2-β-D-fructofuranoside (ou sucrose), les transglucosidases des familles 13 et 70 ont par exemple été employées pour la production de glucoflavonoïdes et d'oligosaccharides d'intérêt industriel ^{73,74}. Des dextransucrases de *Leuconosotoc citreum* ⁷⁵ et de *Leuconostoc mesenteroides* ⁷⁶ ont également été exploitées pour la production d'édulcorants alimentaires avec un rendement de synthèse compris entre 42 et 66 %. Impliquées dans la biosynthèse de fructooligosaccharides chez les plantes et les bactéries ⁶⁴, les transfructosidases des familles 32 et 68 présentent également un intérêt industriel. Ainsi, des transfructosidases recombinantes d'*Aspergillus sp.* sont aujourd'hui fréquemment utilisés dans l'industrie agroalimentaire pour la production d'oligomères de fructose ⁷⁷.

Néanmoins, leur spécificité restreinte vis à vis du donneur limite leur champs d'application. D'autre part, seul un nombre limité de transglycosidases ont à ce jour pu être caractérisées (*Tableau 1*). Par ailleurs, l'utilisation des GH en synthèse reste aussi fortement compromise par leur activité hydrolytique majoritaire; responsable de la dégradation des substrats et des produits de transglycosylation. En vue d'augmenter le rendement de transglycosylation, il est toujours possible d'employer des conditions de réaction favorisant le déplacement de l'équilibre thermodynamique de la catalyse vers la formation de produits de transglycosylation. Ce phénomène, appelé hydrolyse inverse, peut être favorisé en diminuant par exemple la concentration en eau du milieu réactionnel par l'ajout de cosolvants organiques, ou en augmentant la concentration de l'accepteur. Bien que simple à réaliser, cette technique ne permet cependant que d'obtenir des rendements de synthèse variant entre 15 et 65 %. D'autres inconvénients comme la faible régiosélectivité de la réaction et la difficulté liée à la séparation des produits finaux limitent considérablement l'application de cette approche à des fins industrielles ⁷⁸.

En amont, l'ingénierie de ces enzymes reste cependant une solution envisageable afin de réduire leur capacité hydrolytique au profit de la transglycosylation ⁵⁶.

Différentes stratégies d'ingénierie ont été mises en place pour augmenter les rendements de transglycosylation catalysée par les GH. La plupart d'entre elles consistent à faire évoluer la protéine naturelle *via* l'introduction de mutations (mutagenèse) au sein de la séquence protéique. La conception de substrats artificiels visant à favoriser la transglycosylation a également été envisagée ^{55,79}.

1.3.3 Les glycosynthases, thioglycoligases, et thioglycosynthases

Initiée par le groupe de Withers, la conception de glycosynthases et de thioglycoligases à partir de GH a permis d'illustrer une fois de plus l'importance des résidus catalytiques acide-base et nucléophile dans la catalyse de la réaction hydrolytique. Une glycosynthase est une GH dont le résidu nucléophile catalytique a été inactivé par mutagenèse. Ne pouvant plus former d'intermédiaire covalent, l'enzyme est alors dépourvue d'activité hydrolytique. La transglycosylation reste cependant possible à condition d'utiliser un donneur dont le groupe partant est activé et en configuration inversée par rapport au produit de réaction souhaité (*Figure 9*).

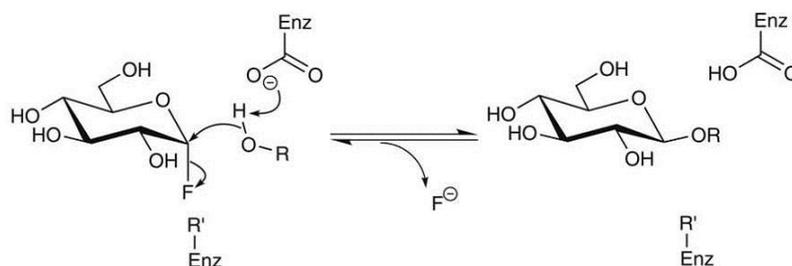


Figure 9. Mécanisme catalytique d'une glycosynthase ⁵⁰.

L'utilisation de fluorure d' α -glucosyl ou d' α -galactosyl en tant que donneurs activés en présence d'accepteurs de type para-nitrophenyl glycosides a permis d'atteindre des rendements de synthèse de l'ordre de 90% ⁵⁰. Testée pour la première fois avec le mutant E358A de la β -glucosidase d'*Agrobacterium* (Agb) ⁵⁰, cette stratégie initialement appliquée pour l'ingénierie d'exo-glycosidases à rétention de configuration issues des familles 1, 2, 5, 7, 8, 10, 16, 17, 26, 29, 31, 36, 52, 85, et 95, a par la suite été étendue aux endo-glycosidases (ENGases) dont les Endo- β -N-acetylglucosaminidases ^{80,81} ainsi qu'aux GH à inversion de configuration ⁸². Dans le dernier cas, la conversion en

glycosynthase est réalisée en mutant l'acide-base catalytique et en utilisant un donneur activé de configuration identique au produit d'hydrolyse de l'enzyme sauvage. L'inactivation du nucléophile empêchant la formation de l'intermédiaire oxazoline, l'ion oxazolium formé par assistance nucléophile du substrat peut néanmoins être attaqué par un accepteur nucléophile pour former un produit de transglycosylation (*Figure 10*).

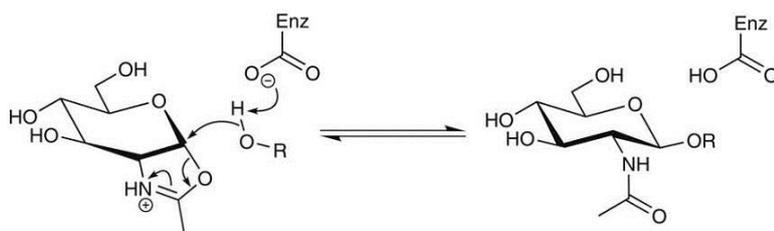


Figure 10. Mécanisme catalytique d'une glycosynthase avec assistance du substrat ⁵⁰.

Deux stratégies alternatives pour la conception de glycosynthases ont été introduites par Aragunde et coll. ⁸³ et Honda et coll. ⁸⁴. La première approche repose sur la substitution du glutamate catalytique nucléophile par un aspartate. Le remplacement de la chaîne latérale du glutamate par une chaîne plus courte a ainsi permis la conversion de la 1,3-1,4- β -glucanase de *Bacillus licheniformis* en glycosynthase. La seconde méthode consiste à introduire une mutation sur un résidu impliqué dans la stabilisation de la molécule d'eau catalytique à proximité de l'acide-base catalytique. Cette approche a été appliquée à une exo-oligoxylanase (Rex) de la famille 8 *via* la substitution Y198F ⁸⁵.

Les capacités synthétiques exceptionnelles des glycosynthases les rendent particulièrement attractives pour la synthèse d'oligosaccharides et de glycoconjugués. Les glycosynthases ont ainsi pu être employées pour la production de polysaccharides d'intérêt biologique comme la cellulose ⁸⁶, la chitine ⁸⁷, ou le xylane ^{88,89}. La conversion de l'endoglycoceramidase de *Rhodococcus sp.* en glycosynthase a également été exploitée pour la synthèse de glycosphingolipides ⁹⁰. Enfin, l'application de cette technologie à la glycosylation contrôlée des protéines d'intérêt pharmaceutique a su se montrer prometteuse (*Figure 11*). Les glycosynthases générées à partir des ENGases d'*Arthrobacter protophormiae* et de *Mucor hiemalis* ont ainsi été employées avec succès pour transférer un oligosaccharide sur le groupement GlcNAc d'un peptide accepteur ⁸¹.

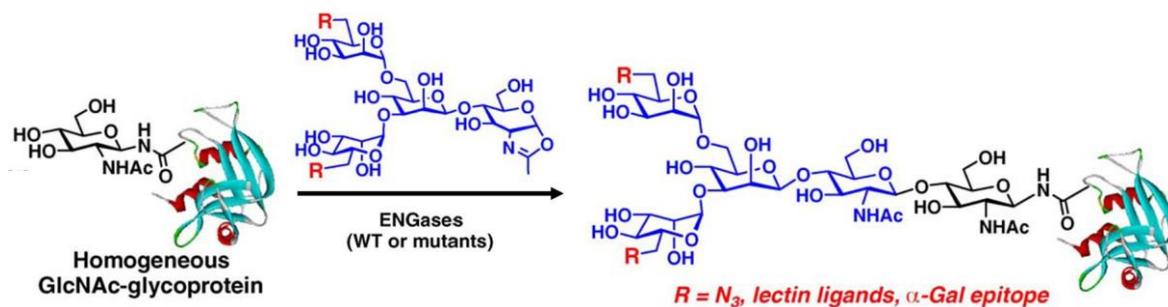


Figure 11. Transglycosylation d'une protéine catalysée par une ENGases ⁵⁵.

D'autres stratégies chimio-enzymatiques employant des sucres donneurs de type oxazolines ont en outre été conçues pour la synthèse de N-glycopeptides et la glycosylation contrôlée d'anticorps thérapeutiques ⁹¹. Non limitées à la glycosylation de macromolécules, les glycosynthases peuvent aussi être utiles pour la glycosylation de petites molécules d'intérêt industriel. Le mutant E197S de la cellulase Cel7B d'*Humicola insolens* s'est par exemple révélé particulièrement efficace dans la production de glycoflavonoïdes ⁹². Le répertoire synthétique des glycosynthases peut également être étendu *via* des techniques d'ingénierie complémentaires. Des approches d'évolution dirigée couplées à des méthodes de criblages adaptées ont par exemple permis d'étendre la spécificité de substrats de ces enzymes et d'augmenter considérablement leur activité de transglycosylation ^{93,94}. Dans le cas de la glycosynthase BGlu1, un contrôle fin de la régiosélectivité de la réaction de synthèse peut être assuré *via* l'introduction par mutagenèse rationnelle de mutations spécifiques au sein du site de liaison de l'accepteur ⁸⁶.

L'inactivation du résidu acide-base catalytique des GH à rétention de configuration a également été entreprise. Les mutants obtenus sont alors incapables d'hydrolyser des substrats nécessitant une catalyse acide mais restent aptes à cliver les substrats activés ⁹⁵. Cependant, en raison de l'absence de la base catalytique, l'étape de déglycosylation du nucléophile est très fortement ralentie. En revanche, l'utilisation d'accepteurs activés anioniques de type thiols accélère considérablement la catalyse et conduit à la formation d'une liaison thioglycosidique entre le donneur et l'accepteur (*Figure 12*). Ces mutants de GH sont pour cette raison appelées thioglycoligases.

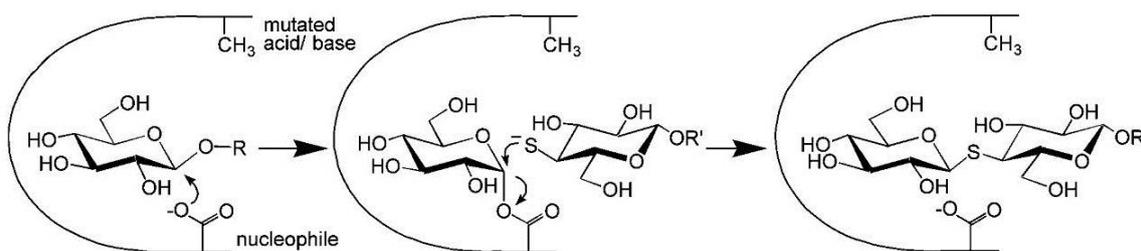


Figure 12. Mécanisme catalytique d'une thioglycosylase ⁹⁶.

Les thioglycosides produits peuvent être utilisés comme précurseurs pour la production de substrats synthétiques d'enzymes ⁹⁷ ou d'inhibiteurs ⁹⁸. Des rendements de synthèse entre 60 et 80% ont pu être obtenus avec ces mutants. Initiée sur le mutant E171A de l'Agb et E429A de la β -mannosidase de *Cellulomonas fimi* ⁹⁵ cette approche a par la suite appliquée à d'autres enzymes. La conversion de l'arabinofuranosidase Araf51 de *Clostridium thermocellum* en thioglycosylase a par exemple rendue possible l'application de cette approche à la synthèse de thiofuranosides ⁷⁷. Tout comme les glycosynthases, les thioglycosylases peuvent également être utilisées pour la production de thioglycoprotéines ⁹⁹. En vue de maximiser les rendements de transglycosylation, il reste cependant nécessaire de supprimer complètement l'activité hydrolytique. Dans cette logique, il est possible d'inactiver les deux résidus catalytiques acide-base et nucléophile (Figure 13). Cette approche présente également l'avantage de pouvoir utiliser des fluorures de glycosyles en tant que donneurs. Plus simples à produire, ils peuvent alors remplacer les donneurs de type dinitrophenyl glycosides traditionnellement employés comme substrat des thioglycosylases. Les doubles mutants obtenus sont appelés thioglycosynthases ⁹⁶.

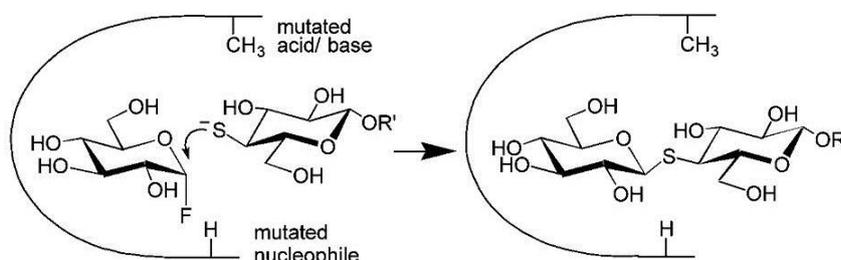


Figure 13. Mécanisme catalytique d'une thioglycosynthase ⁹⁶.

Malgré leur caractère prometteur, l'utilisation des glycosynthases et des thioglycoligases souffre de quelques inconvénients majeurs. Particulièrement réactifs, les fluorures de glycosyles utilisés comme donneurs des glycosynthases s'hydrolysent spontanément et doivent ainsi être préparés extemporanément¹⁰⁰. De plus, la préparation de l'accepteur de type thiol des thioglycoligases peut également nécessiter des étapes de protection/déprotection¹⁰¹ rendant alors cette approche plus délicate. Le champ d'application de la thioglycoligation reste également limité à la production de thioglycosides. Enfin, l'utilisation obligatoire de substrats synthétiques ne permet pas toujours d'exploiter ces enzymes dans des stratégies de synthèse impliquant des substrats naturels. Une méthode alternative consiste alors à employer diverses techniques de mutagenèse visant renforcer l'activité de transglycosylation des GH tout en conservant les résidus catalytiques acide-base et nucléophile intacts.

1.4 Ingénierie des glycosidases pour augmenter la transglycosylation

De nombreux exemples d'ingénierie de GH en transglycosidases peuvent être recensés dans la littérature. Par l'utilisation de stratégies de mutagenèse rationnelles, combinatoires, ou semi-rationnelles, ces travaux ont montré qu'il était possible d'altérer la partition hydrolyse/transglycosylation (H/T) de ces enzymes en mutant certains résidus conservés en bordure des sous-sites de liaison aux substrats donneurs et accepteurs. Ces sous-sites se définissent comme des portions du site donneur ou accepteur capables de lier une unité osidique spécifique au sein d'un substrat donné. Les sites donneur et accepteur des GH sont comportent au minimum un sous-site chacun, respectivement dénommés -1 (où a lieu la catalyse) et +1 (liant l'aglycone et l'accepteur). Afin d'accommoder des substrats oligosaccharidiques de tailles variables, les sous-sites -1 (chez les endo-glycosidases) et +1 sont bordés de sous-sites adjacents dans le cas de nombreuses GH (*Figure 14*).

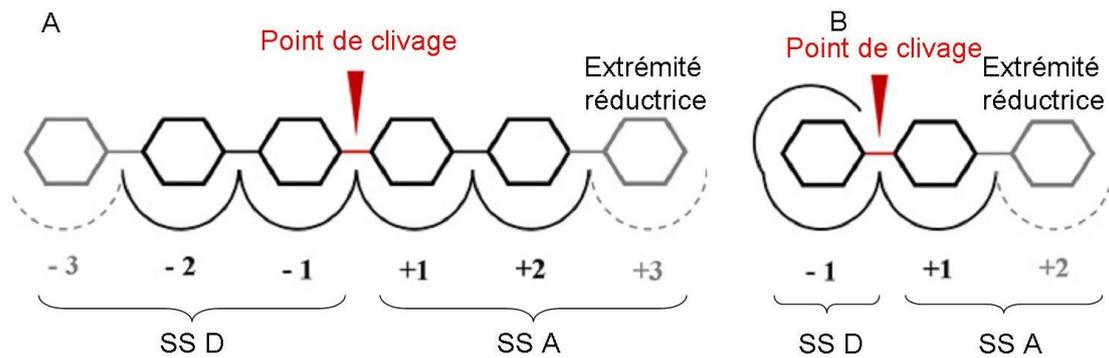


Figure 14. Nomenclature des sites de liaison des GH aux sucres donneurs et accepteurs ⁵⁶.

Les hexagones représentent des unités osidiques de type hexose pontés entre eux par une liaison glycosidique. Les encadrés A et B représentent respectivement la topologie des sites de liaisons dans le cas d'une endo-glycosidase (clivage à l'intérieur de la chaîne) et d'une exo-glycosidase (clivage à l'extrémité non-réductrice). SSD et SSA correspondent respectivement aux sous-sites donneurs (numérotation négative) et accepteurs (numérotation positive) du site catalytique. Les unités osidiques situées au sein des sous-sites de liaison au donneur et à l'accepteur illustrent respectivement la partie glycone et aglycone du substrat donneur.

Une des premières expériences réalisées selon cette approche concerne l'ingénierie du sous-site -1 de la β -glycosidase de *Thermus thermophilus* (Tt β gly, GH1) ¹⁰². Employant une méthodologie combinant évolution dirigée et mutagenèse rationnelle ¹⁰³, des mutants à fort rendements de transglycosylation (jusqu'à 70 % de gain par rapport à l'enzyme sauvage) ont pu être obtenus ¹⁰⁴. Le séquençage des deux meilleurs mutants a révélé la mutation de deux résidus conservés en seconde couche du sous-site -1 (F401 et N282). En ciblant les résidus conservés autour du sous-site -1, cette approche a pu être par la suite étendue à une α -galactosidase de la famille 36 et a permis de caractériser des résidus des sous-sites -1 et +1 fortement impliqués dans le contrôle de la balance H/T ¹⁰⁵. Une stratégie de mutagenèse semi-rationnelle analogue a de plus été récemment appliquée à une α -L-Arabinofuranosidase de *Thermobacillus xylanilyticus* (GH51) ^{106,107}. Dans l'ensemble des cas évoqués ci-dessus, l'importante diminution des performances catalytiques (k_{cat} /K_m) des mutants a suggéré la possible implication de ces résidus dans la stabilisation de l'état de transition. Mais un mutant à fort rendement de transglycosylation dérivé de la chitinase A de *Serratia marcescens* a également pu être conçu en mutant un tryptophane conservé au sein du sous-site -3 ¹⁰⁸. Il a été supposé que ce résidu pourrait contrôler la flexibilité conformationnelle de la chaîne

latérale du résidu D313. La mutation W167A contribuerait à diminuer sa flexibilité. Alors orienté vers le nucléophile, il stabiliserait l'intermédiaire glycosyl-enzyme favorisant ainsi la transglycosylation au détriment de l'hydrolyse. Enfin, la substitution d'une arginine très conservée et directement impliquée dans la liaison au galactose au sein du sous-site -1 de la β -galactosidase de *Geobacillus stearothermophilus* (GH42) a par ailleurs démontré que la déstabilisation de l'interaction du substrat donneur à son site de liaison contribue à augmenter l'activité de transglycosylation ¹⁰⁹.

Si la mutagenèse de positions conservées en bordure du sous-site -1 a su se montrer efficace pour favoriser la transglycosylation, il est également possible d'accroître le ratio T/H en introduisant des mutations au niveau du site de liaison de l'accepteur afin de renforcer son interaction avec la protéine. De nombreux travaux réalisés sur différentes familles de GH ont par ailleurs pu montrer l'importance de résidus aromatiques conservés dans la liaison du sucre accepteur. Ainsi, la substitution des résidus aromatiques W97 et F396 au sein de la chitinase A de *Serratia marcescens* (GH18) réduit considérablement l'activité de transglycosylation ¹¹⁰. Par analogie, le remplacement du groupement phényle de F396 par un cycle aromatique plus encombrant *via* la substitution F396W produit l'effet inverse. L'implication d'un résidu phénylalanine au sein du site +1 de la β -glycosidase TnBgl1A de *Thermotoga neapolitana* dans l'activité de transglycosylation a également été démontré ¹¹¹. Une étude comparée de deux homologues de la famille 33, la sialidase de *Trypanosoma rangeli* (TrSA) et de la transialidase de *Trypanosoma cruzi* (TcTS), a également suggéré des conclusions similaires. Présent au sein du site accepteur de TcTS et muté en sérine chez TrSA, le résidu Y119 s'est révélé être indispensable à l'activité de transialidation de TcTS et semblerait impliqué dans la stabilisation de l'accepteur ¹¹². Cette hypothèse a pu être confirmée ultérieurement par des travaux de modélisation moléculaire ¹¹³. Bien qu'essentielle pour promouvoir la transglycosylation chez de nombreuses GH, la présence de résidus aromatiques au sein du site accepteur n'induit pas systématiquement une augmentation de l'activité de transglycosylation. Ainsi, la mutation F116A introduite au sein du site +1 de la β -xylosidase de *Bacillus halodurans* (GH39) diminue l'activité de transglycosylation lorsque le pentanol est utilisé comme substrat accepteur ¹¹⁴. L'utilisation d'octanol, provoquant l'effet inverse, démontre alors que le choix du type de substrat accepteur doit également être pris en considération.

Ces résultats suggèrent qu'une multitude de facteurs moléculaires semblent être impliqués dans le contrôle de l'équilibre de la balance H/T. Les études évoquées précédemment ont souligné l'importance de résidus conservés au sein des sites donneur et accepteur dans la stabilisation de l'état de transition et de l'intermédiaire covalent glycosyl-enzyme ainsi que dans l'affinité vis-à-vis du substrat accepteur. D'autres travaux ont par ailleurs insistés sur le rôle de certains résidus dans la stabilisation de l'eau interne au voisinage de l'acide-base catalytique^{84,85}. L'eau étant indispensable à la réaction d'hydrolyse, il est en effet probable que la dynamique de l'eau interne à la protéine puisse jouer un rôle crucial dans la modulation de l'équilibre H/T.

2 L'eau interne aux protéines

2.1 Le rôle de l'eau interne dans la fonction des protéines

En dépit de leur densité importante, l'intérieur des protéines globulaires est en permanence hydraté^{115,116}. Des molécules d'eau spatialement conservées ont d'ailleurs été trouvées au sein des structures d'un grand nombre d'enzymes incluant la famille 1 des GH¹¹⁷, les ribonucléases¹¹⁸, des kinases¹¹⁹, ainsi que les sérines-protéases¹²⁰ et aspartate-protéases¹²¹. La solvatation interne des protéines joue un rôle important dans leur fonctionnalité ainsi que dans le repliement^{122,123}, la dynamique¹²⁴, la flexibilité^{125,126}, et la stabilité de leur structure biologique^{127,128}. L'eau, par son caractère de base de Lewis, peut catalyser diverses réactions chimiques et occupe ainsi une place essentielle dans de nombreuses réactions enzymatiques^{129,130,131} (*Figure 15*). Elle peut également jouer un rôle indirect dans la catalyse en régulant le degré d'ionisation des résidus avec lesquels elle interagit^{132,133} ou en participant à la création de réseaux d'interactions polaires et de liaisons hydrogène¹³⁴.

Jouant un rôle crucial dans le maintien de l'architecture protéique, ces interactions stabilisatrices peuvent également contribuer à stabiliser les interactions d'une enzyme avec son substrat^{135,136} ou l'état de transition¹³⁷. Par ailleurs, la solvatation d'un ligand au sein de son site de liaison conditionne fortement la thermodynamique de son interaction avec les résidus environnants¹³⁸. De manière complémentaire, cette interaction est souvent renforcée par l'augmentation de l'entropie liée au déplacement

par le substrat de molécules d'eau ordonnées au sein du site de liaison ¹³⁹. Une étude de la solvation du site catalytique des phosphoinositide kinases et de l'endonucléase *EcoRI* a d'autre part montré l'importance de la présence et de la disposition spatiale des molécules d'eau internes au site de liaison dans l'affinité et la sélectivité de l'interaction enzyme-substrat ^{138,140}.

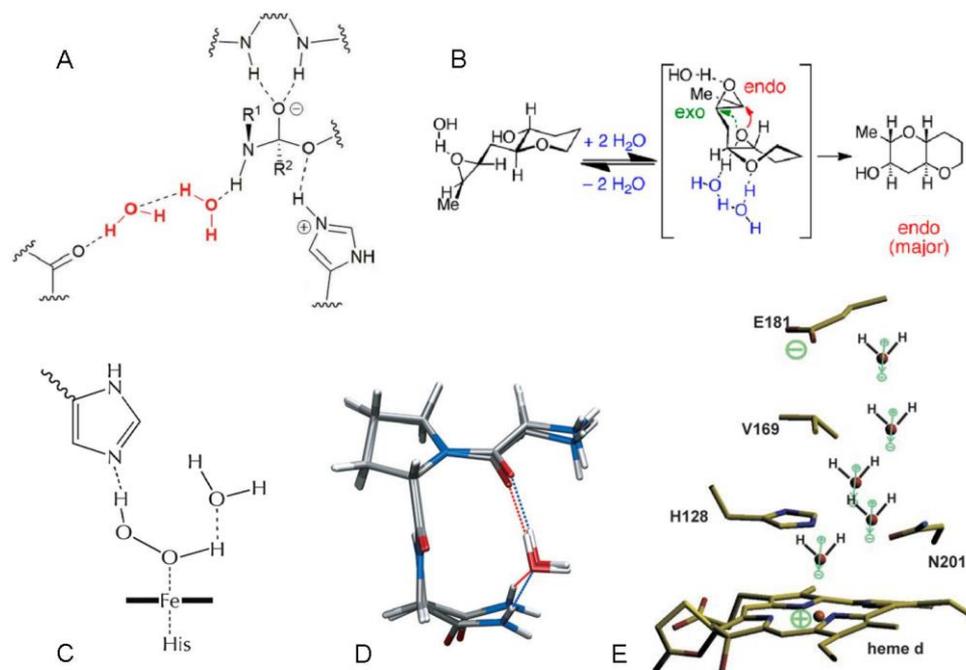


Figure 15. Quelques rôles biologiques de l'eau :

A : Stabilisation de l'état de transition dans le site actif de la lipase B de *C. antarctica* ¹³⁷.

B : Cyclisation des époxydes ¹³⁰.

C : Transfert de proton pour la dégradation d' H_2O_2 dans le site actif d'une peroxydase ¹⁴¹.

D : Stabilisation de coudes structuraux au sein des protéines ¹²².

E : Relais électrostatique impliqué dans la stabilisation de charges dans la catalase ¹⁴².

2.2 Les canaux d'eau internes aux protéines

Des simulations de dynamique moléculaire et des clichés de cristallographie ont pu montrer une organisation possible des molécules d'eau internes sous la forme de clusters et de chaînes ordonnées ¹⁴³ au sein de nombreuses enzymes. La capacité de l'eau interne à s'organiser de telles sortes résulte de la formation de crevasses ou tunnels au cœur des structures protéiques ¹⁴⁴. La forme allongée de certaines cavités

contribue très souvent à la formation de canaux d'eau internes. Débouchant parfois à la surface des protéines, ces canaux peuvent alors favoriser l'accessibilité de l'eau externe (« bulk ») aux sites structuraux enfouis (sites catalytiques, poches de liaison...) et remplissent par ce rôle une fonction essentielle chez un grand nombre d'enzymes.

La découverte de canaux d'eau internes aux protéines a pu être initiée par les premiers travaux sur l'étude des aquaporines ¹⁴⁵. Traversées par une cavité connectant le milieu extracellulaire au cytosol, ces protéines transmembranaires ubiquitaires ^{146,147} régulent les échanges d'eau au sein des tissus ¹⁴⁸. Si les aquaporines sont de par leur fonction dotées d'un large canal unique et bien défini, les enzymes peuvent en revanche disposer d'une plus grande diversité de canaux d'eau internes variant selon leurs formes, leurs dimensions et leurs localisations au sein de la protéine. Des canaux d'eau aux fonctionnalités diverses ont ainsi pu être répertoriés chez plusieurs classes d'enzymes dans l'ensemble du règne du vivant.

Un des rôles ubiquitaire majeur exercé par ces canaux est sans doute le transport de protons. Un enchaînement régulier des molécules d'eau au sein d'un canal peut en effet permettre le transfert ordonné (dit « de Grotthuss » ¹⁴⁹) de protons d'un point à l'autre de la protéine. Dans les cas des chaînes respiratoires membranaires, mitochondriales et chloroplastiques, les canaux d'eau impliqués dans le transport de protons contribuent à générer la force motrice nécessaire à la production d'ATP intracellulaire et occupent en conséquence une position centrale dans les processus de respiration cellulaire et de photosynthèse ¹⁵⁰. Des canaux d'eau transporteurs de protons associés à une importante diversité de fonctions biologiques (production d'hydrogène, oxydation métabolique, production d'ATP...) ont ainsi été mis en évidence chez diverses enzymes telles les hydrogénases ¹⁵¹, les déshydrogénases ¹⁵², les peroxydases ^{141,153}, les bacteriorhodospines ¹⁵⁴, et les cytochromes oxydases ^{155,156} (*Figure 16*).

Les canaux d'eau fonctionnels des hydrolases sont en revanche le plus souvent impliqués dans l'apport de molécules d'eau catalytiques au site actif.

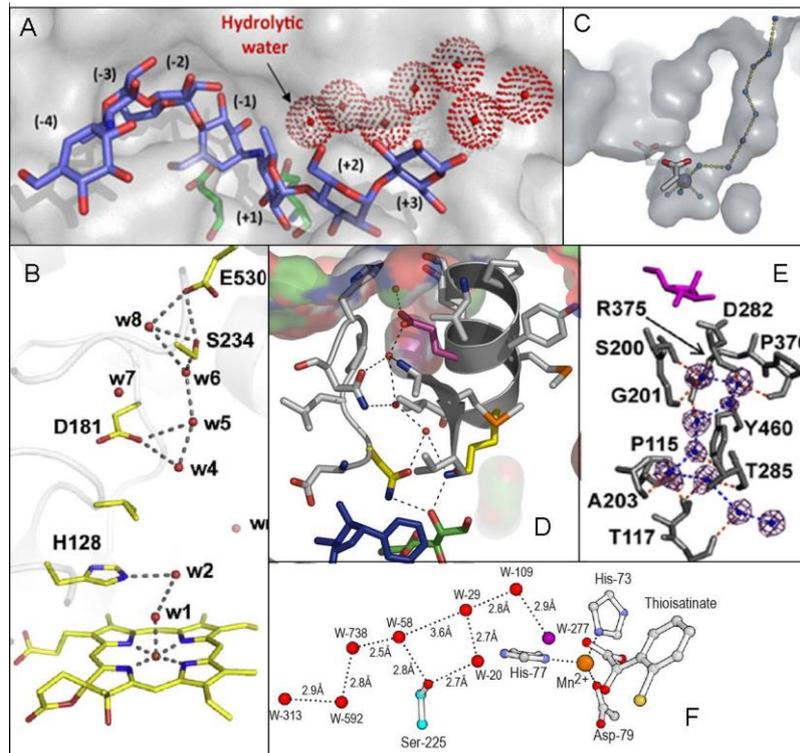


Figure 16. Quelques canaux d'eau cristallographiques rencontrés chez diverses enzymes.

A : Canal d'eau d'une α -amylase de *Pseudoalteromonas haloplanctis* ¹⁵⁷.

B : Canal d'eau de la catalase *KatE* d'*Escherichia coli* ¹⁵³.

C : Canal d'eau de la 1,3- α -3,6-anhydro-L-galactosidase de *Zobellia galactinovorans* ¹⁵⁸.

D : Canal d'eau de la D-mannitol 2-déshydrogénase de *Pseudomonas fluorescens* ¹⁵².

E : Canal d'eau de la β -fructofuranosidase K-1 d'*Arthrobacter sp.* ¹⁵⁹.

F : Canal de l'isatin hydrolase de *Labrenzia aggregata* ¹⁶⁰.

De nombreuses études de mutagenèse ont notamment pu confirmer le caractère fonctionnel de tels canaux d'eau dans l'activité des cytochrome oxydases ¹⁶¹, des squalène-hopène cyclases ¹⁶², des peroxydases ¹⁴² et d'hydrolases ^{160,163,164,165}. Ces expériences ont également permis de mettre en évidence des résidus enfouis contrôlant la circulation des molécules d'eau au sein du canal ¹⁶⁰, leur transfert vers ¹⁶⁶ et hors ¹⁶⁷ du site catalytique. Non limités aux canaux d'eau, une étude récente a pu répertorier au moins 129 résidus contrôlant la diffusion de petites molécules fonctionnelles au sein de cavités internes de 71 protéines issues des 6 classes fonctionnelles d'enzymes ¹⁶⁸. Sur l'ensemble des résidus recensés, 36 ont pu être caractérisés comme régulateurs du passage de molécules vers le site catalytique. A l'image de valves, la dynamique des chaînes latérales de ces résidus (principalement aromatiques) permettrait ainsi de réguler le trafic moléculaire au sein des enzymes.

2.3 L'ingénierie des canaux d'eau internes aux enzymes

Des canaux d'eau potentiellement impliqués dans l'acheminement de molécules d'eau catalytiques au site catalytique ont par ailleurs été mis en évidence chez une α -amylase humaine¹⁵⁷ ainsi que dans le cas de GH issues des familles 1¹¹⁷ et 117¹⁵⁸. Un canal communicant avec le site actif d'une β -fructofuranosidase d'*Arthrobacter sp.* (GH68) a quant à lui été suspecté de jouer un rôle dans le drainage de l'eau hors du site catalytique¹⁵⁹. Bien que n'ayant pas été démontrée, une telle fonction permettrait de limiter la solvataion du site catalytique et pourrait peut être expliquer la prédominance de l'activité de transglycosylation sur l'hydrolyse au sein des GH de la famille 68. D'autres travaux ont montré que l'introduction de mutations en bordure des canaux susceptibles de conduire l'eau du bulk au site catalytique pouvait significativement affecter l'activité de nombreuses enzymes. Chez les squalène-hopène cylases, l'obstruction par mutagenèse d'un canal suspecté de fournir l'eau catalytique nécessaire à la conversion du squalène-hopène en hopanol a permis de favoriser la réaction inverse¹⁶². Par analogie, l'interruption d'un canal impliqué dans la solvataion du site catalytique de l'haloalkane déhalogénase de *Rhodococcus rhodochrous* a eu pour effet de renforcer l'activité de cette enzyme vis à vis du 1, 2, 3-trichloropropane¹⁶⁶.

Soulignant également l'importance fonctionnelle de ces canaux d'eau chez les hydrolases, la mutation de résidus susceptibles de contrôler le passage d'eau du canal au site catalytique a conduit à diminuer l'activité d'hydrolyse chez la lipase B de *Candida antarctica*¹⁶³ ainsi que chez la β -glycosidase de *Tenebrio molitor*¹⁶⁹. L'ensemble de ces études démontre ainsi de manière remarquable que l'ingénierie de canaux d'eau internes aux enzymes constitue une technique prometteuse pour modifier leurs propriétés catalytiques. Compte tenu de l'importance de l'eau dans l'activité des GH, cette méthode pourrait donc être appliquée à l'identification de nouveaux résidus impliqués dans le contrôle de la balance H/T et permettre par la suite de concevoir de nouveaux catalyseurs glycosynthétiques. En vue de pouvoir cibler les résidus impliqués dans la formation des canaux d'eau internes aux GH, il est essentiel de localiser avec précision le parcours des molécules d'eau au sein de ces enzymes. Bien que restant une méthode de référence dans la détermination des structures moléculaires, l'utilisation de la diffraction aux rayons X n'est pas toujours adaptée pour répondre à cet objectif. Les

densités électroniques des molécules d'eau internes sont en effet souvent mal résolues à cause de leur mobilité trop importante. Bien que le développement de nouvelles méthodes de cristallographie à haute résolution puisse nous permettre à l'avenir de résoudre ce problème, d'autres techniques peuvent cependant être déjà envisagées.

2.4 Méthodes biophysiques pour l'étude de la dynamique de l'eau interne

2.4.1 La spectrométrie de masse d'échange deutérium-proton

La spectrométrie de masse d'échange deutérium-proton (DXMS) a par exemple été utilisé par Teze et coll. pour évaluer l'accessibilité du solvant à l'intérieur de la β -glycosidase de *Thermus thermophilus*¹¹⁷ (Figure 17). Le principe de cette technique consiste à incuber une protéine d'intérêt dans un tampon composé essentiellement d'eau deutérée (D₂O). Durant l'incubation, les deutérons du solvant s'échangent avec les protons des groupements amide accessibles de la protéine. Les régions internes bordant un éventuel canal d'eau enfoui seront ainsi marquées par le deutérium, ce qui permettra par la suite leur identification par spectrométrie de masse. Bien qu'ayant été appliquée avec succès dans l'étude menée par Teze et coll., cette méthode souffre néanmoins de quelques inconvénients majeurs. Tout d'abord, elle ne permet pas d'observer les échanges rapides des protons liés aux hétéroatomes des chaînes latérales et reste limitée à la détection des échanges des protons amidiques du squelette peptidique¹⁷⁰. De plus, un échange retour au cours duquel les deutérons sont à nouveau échangés contre des protons a souvent lieu dans le cas des régions les plus exposées au solvant. Conduisant à une disparition du marquage au deutérium, ce phénomène rend alors impossible la détection des régions concernées¹⁷¹.

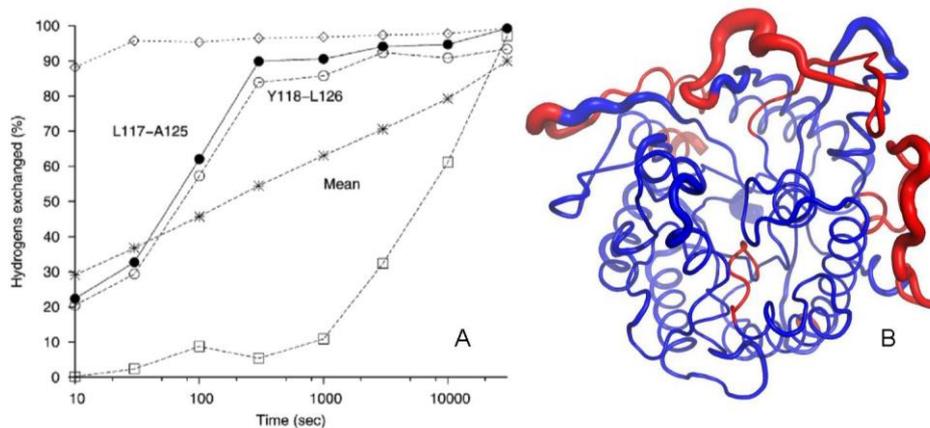


Figure 17. Mise en évidence de peptides de la Ttβgly accessibles au solvant par DXMS.

A : Cinétique d'échange deutérium-proton au sein de peptides internes ¹¹⁷.

B : Peptides (rouge) dont les protons ont été échangés à plus de 90 % après 1000 s d'incubation en tampon deutéré ¹⁷¹.

2.4.2 La spectrométrie de masse à radiolyse

Une technique voisine permettant de pallier à cette limitation est la spectrométrie de masse à radiolyse (radical probe mass spectrometry ou RPMS) ^{172,173}. Son principe consiste à irradier une solution contenant la protéine d'intérêt en présence d'eau deutérée. L'irradiation aux rayons X des molécules d'eau conduit à la formation de radicaux libres HO[•] en solution. Ces radicaux pourront alors oxyder les résidus qui leur sont accessibles, rendant ainsi possible leur détection ultérieure par spectrométrie de masse (*Figure 18*). Au contraire du marquage au deutérium utilisé dans le cas de la DXMS, l'oxydation radicalaire est irréversible. Le marquage des régions exposées au solvant est alors permanent et confère ainsi un avantage majeur à la RPMS en comparaison à la DXMS. La cinétique d'oxydation des chaînes latérales étant de l'ordre de la dizaine de milliseconde, la solvation des régions les moins accessibles ne pourra cependant pas être détectée. Cette limitation rend ainsi cette technique inadaptée à la détection de canaux d'eau dont la durée de vie est inférieure à la milliseconde. Un autre inconvénient de cette technique est que toutes les chaînes latérales d'acides aminés ne sont pas sensibles de manière équivalente à l'oxydation.

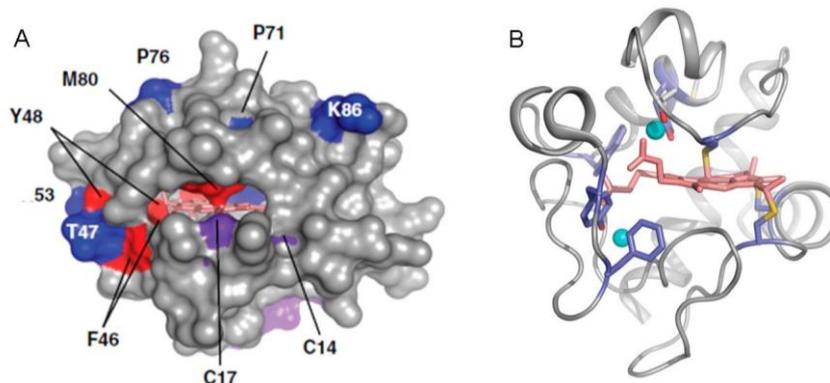


Figure 18. Mise en évidence de peptides solvatés de la cytochrome-c-oxydase par RPMS ¹⁷².

A : Degré d'oxydation des chaînes latérales augmentant du rouge au violet.

B : Localisation de deux molécules d'eau enfouies très conservées.

2.4.3 La spectroscopie infrarouge

Plus résolutive, la spectroscopie infrarouge (IR) a pu être utilisée pour caractériser la dynamique d'échange de protons au sein d'un canal d'eau interne à la structure de la bactériorhodopsine ¹⁵⁴. Cette méthode a notamment pu permettre de mettre en évidence un transfert de protons entre une molécule d'eau du canal et un résidu vicinal impliqué dans la fonction de la protéine. Ceci s'explique par le fait que lorsqu'un transfert de protons a lieu entre un résidu et une molécule d'eau, la fréquence de vibration du groupe accepteur de proton est modifiée et peut donc être détectée sur le spectre d'absorbance (*Figure 19*). La spectroscopie IR rend donc non seulement possible la localisation de l'eau interne au sein des protéines, mais permet en outre de caractériser sa dynamique. Une étude analogue utilisant la spectroscopie infrarouge bidimensionnelle (2D-IR) par ailleurs pu mettre en évidence la présence de molécules d'eau mobiles au sein de fibrilles de β -amyloïdes ¹⁷⁴.

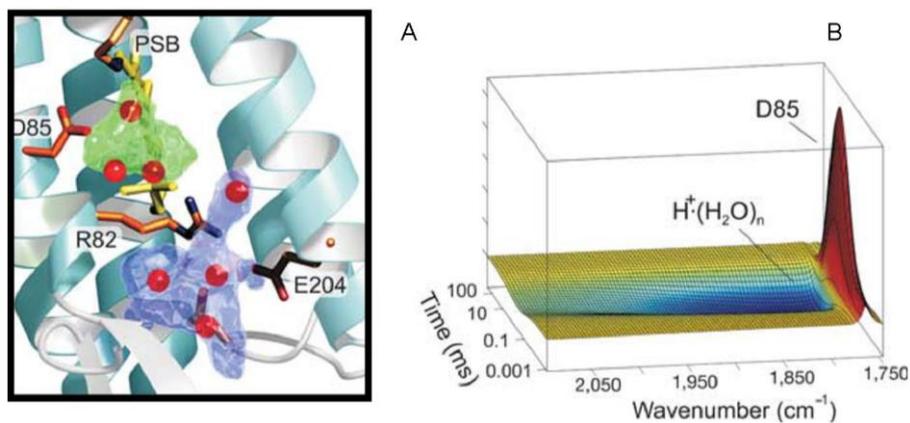


Figure 19. Mise en évidence d'un transfert de proton entre le résidu D85 de la bactériorhodopsine et une molécule d'eau par spectroscopie FTIR ¹⁵⁴.

A : Canal d'eau interne à la bactériorhodopsine.

B : Spectre FTIR 3D signalant un transfert de proton entre D85 et une molécule d'eau.

2.4.4 La Résonance Magnétique Nucléaire en micelle inverse

Enfin, une nouvelle méthode de RMN s'est également montrée efficace dans la caractérisation de la dynamique des molécules d'eau interne aux protéines. L'utilisation de la RMN traditionnelle pour cet objectif est en effet souvent limitée par l'anisotropie des molécules d'eau ainsi que leur faible temps de résidence. La RMN en micelle inverse ¹⁷⁵ a donc été développée en vue de surmonter cette difficulté. L'efficacité de cette technique repose sur l'encapsulation de la protéine d'intérêt au sein d'une gouttelette de solvant polaire, rendue possible par l'émulsion d'une solution contenant la protéine d'intérêt au sein d'un solvant organique apolaire. Le confinement d'une protéine unique au sein d'une goutte d'eau nanoscopique contribuerait à ralentir les échanges de solvant à l'intérieur de celle-ci. Par cet effet, il est alors possible de quantifier les phénomènes de relaxation de l'eau interne au sein une protéine donnée et donc de déterminer la position des molécules d'eau enfouies ainsi que leur temps de résidence respectif (*Figure 20*). La nécessité d'attribuer l'ensemble du spectre de la protéine est cependant la principale limitation de cette technique qui reste réservée aux protéines de tailles modestes.

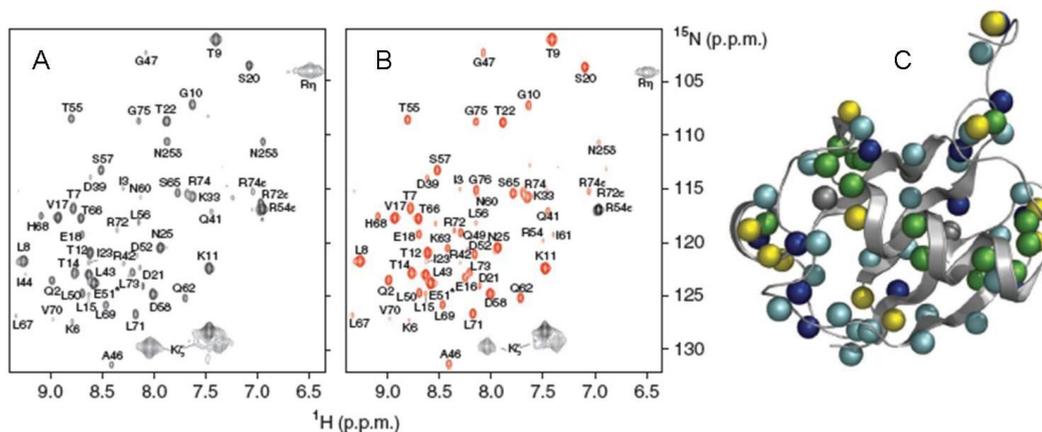


Figure 20. Caractérisation des sites d'hydratation internes à l'ubiquitine par RMN 3D en micelle inverse ¹⁷⁵.

A et B : Spectre de résonance ¹H-¹⁵N NOESY et ROESY montrant des interactions dipolaires entre protons des molécules d'eau et azote des amides du squelette peptidique.

C : Sites d'hydratation coloré par temps de résidence (bleu marine : ~ 10 ns).

Malgré l'intérêt que peuvent susciter ces technologies pour étudier la dynamique de l'eau interne aux protéines, le coût associé à l'achat des appareils et à leur maintenance ne permet pas toujours de les exploiter. En alternative, il est possible d'avoir recours à des méthodes de simulations informatiques permettant de prédire la trajectoire susceptible d'être explorée par l'eau interne au sein de la structure tridimensionnelle d'une protéine donnée. Ces méthodes théoriques, complémentaires aux méthodes expérimentales évoquées précédemment, sont couramment appelées « simulations de dynamique moléculaire ».

3 Modélisation de l'eau interne et simulation de sa dynamique

3.1 Principe de la mécanique moléculaire

Dans le but de pouvoir simuler le mouvement des molécules d'eau au sein d'une protéine d'intérêt, il est nécessaire d'adopter un modèle physique permettant de décrire l'ensemble des interactions avec celle-ci. En pratique, un ensemble fini de molécules d'eau est en général modélisé autour de la protéine en présence d'ions (Figure 21). L'ensemble des atomes modélisés constituent un système moléculaire. La mécanique moléculaire est un modèle physique très utilisée pour décrire les interactions intramoléculaires et intermoléculaires gouvernant la dynamique d'un système. Ce

modèle consiste à représenter les atomes sous la forme de sphères rigides et les liaisons covalentes sous la forme de ressorts. Les électrons ne sont donc pas traités de manière explicite. Les atomes ne sont alors décrits que par une masse et une charge ponctuelle. La modélisation de toute réaction chimique, nécessitant la rupture ou la formation de liaisons covalentes, est alors impossible.

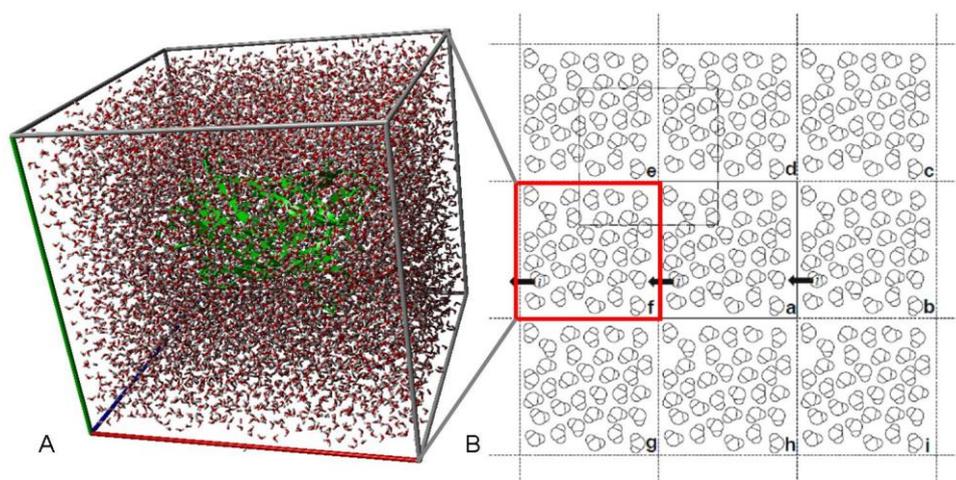


Figure 21. Définition d'un système « tout-atome » en simulation de dynamique moléculaire.

La protéine (verte) est immergée au sein d'une boîte de solvant (eau + contre ions) (A). Afin de s'affranchir des effets de bords lors du contact d'une particule sur l'une des faces de la boîte, des conditions périodiques sont utilisées. Le système est donc répliqué à l'infini dans les trois directions de l'espace (B). Les mouvements des atomes présents dans une boîte donnée sont répliqués dans toutes les autres boîtes. En conséquence, le départ d'une particule vers une boîte voisine entraîne alors son remplacement par une particule image d'une boîte adjacente du côté opposé. Le nombre de particules présent au sein d'un tel système (et donc d'une boîte) reste constant durant la simulation (Adapté d'après : http://ecole.modelisation.free.fr/cours_chipot.pdf).

3.2 Modèles théoriques de l'eau

Certes simpliste, cette description physique peut être complexifiée dans le cas de l'eau afin de mieux décrire les interactions entre atomes et de reproduire au mieux les propriétés physicochimiques de l'eau liquide. Il existe aujourd'hui trois catégories de modèles atomistiques d'eau : l'eau rigide, l'eau flexible, et l'eau polarisée. Le modèle d'eau rigide est décrit par deux longueurs de liaisons et un angle de valence de valeurs fixes. Contrairement au modèle d'eau flexible, ce modèle ne peut pas reproduire les

propriétés vibrationnelles de l'eau liquide. Le modèle physique de l'eau polarisée est caractérisé par un terme de polarisation explicite supplémentaire permettant de décrire plus finement les interactions électrostatiques entre molécules d'eau par rapport à une simple charge ponctuelle et ainsi de mieux décrire leur géométrie. Tendrant à moyennner les propriétés physiques de l'eau liquide, les modèles rigides permettent de réduire le temps nécessaire au calcul des interactions entre atomes et sont à ce jour les plus utilisés. Parmi les modèles les plus simples, quatre structures atomistiques sont les plus couramment employées. La structure d'eau la plus simple et la plus utilisée est une structure planaire à trois sites chargés (l'oxygène et les deux hydrogènes). Les modèles d'eau TIP3P¹⁷⁶ et SPC/E¹⁷⁷, qui sont les plus utilisés. Par l'ajout de sites chargés supplémentaires, les modèles d'eau à 4 (exemple : TIP4P¹⁷⁸) ou 5 points de charge (exemple : TIP5P¹⁷⁹) visent à mieux modéliser les orbitales électroniques non-liantes sur l'atome d'oxygène (Figure 22).

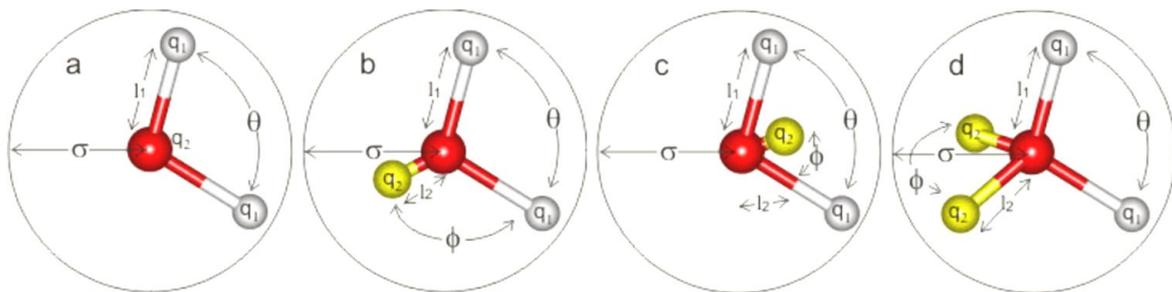


Figure 22. Structures de molécule d'eau utilisées en simulation de dynamique moléculaire.

Les paramètres communs à toutes les structures sont l'angle de valence θ , la longueur de liaison (l_1) et le rayon de Van der Waals σ .

A : Eau planaire à 3 sites (deux charges q_1 , une charge q_2).

B : Eau tétraédrique à 4 sites (la charge q_2 est partiellement délocalisée sur un 4^{ème} site).

C : Eau planaire à 4 sites (la charge q_2 est partiellement délocalisée sur un 4^{ème} site).

D : Eau tétraédrique à 5 sites (deux charges q_2 modélisent les orbitales non-liantes).

(Adapté d'après : http://www1.lsbu.ac.uk/water/water_models.html)

3.3 Interactions moléculaires et champs de force

Basées sur le modèle de la mécanique moléculaire, les simulations de dynamique moléculaire classiques peuvent être utilisées pour simuler les mouvements de vibration des atomes au sein d'un système moléculaire (ici protéine et solvant) donné ¹⁸⁰. Dans ce modèle théorique, la trajectoire d'un atome donné dépendra essentiellement de sa position dans l'espace par rapport aux autres atomes du système. Sa position déterminera ainsi sa capacité à interagir avec les autres atomes et lui confèrera une énergie potentielle d'interaction. La fonction d'énergie permettant de calculer l'énergie potentielle propre à chaque atome se décompose classiquement en une somme de quatre à cinq termes d'interactions.

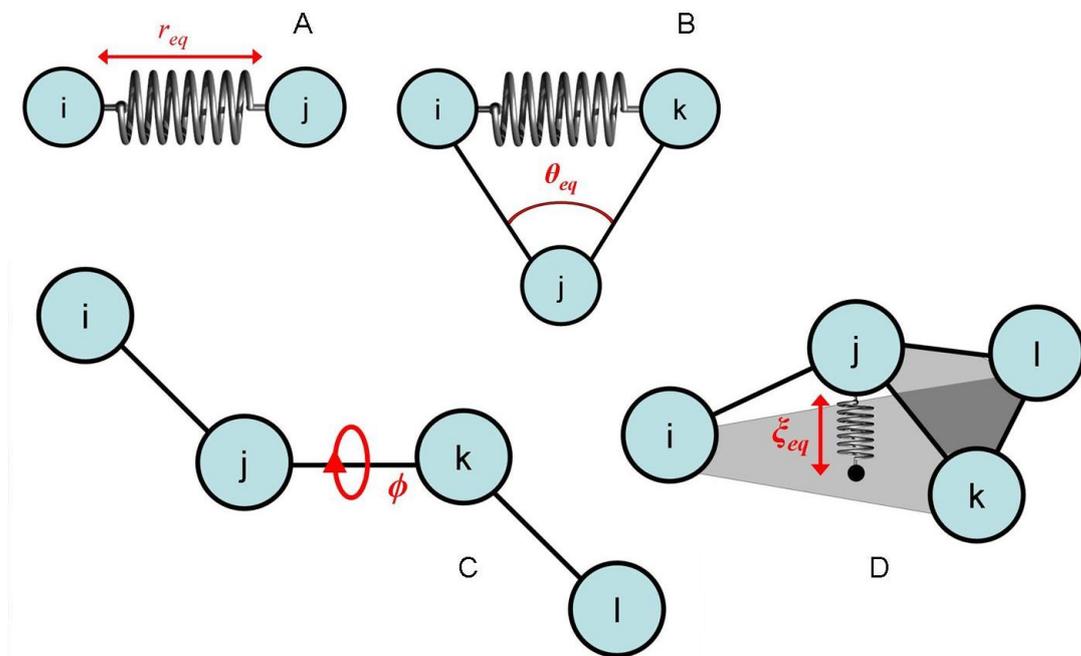


Figure 23. Ensemble des types d'interactions liées.

Les sphères *i*, *j*, *k*, *l* représentent les atomes en interactions. Les termes d'interactions A, B, C, D, correspondent respectivement à l'élongation de la liaison covalente, la déformation des angles de valence, la torsion d'un angle dièdre propre, et la contrainte angulaire d'un dièdre impropre. Dans le cas du dernier terme, on impose à l'atome *j* une contrainte visant à le maintenir à une certaine distance du plan défini par les atomes *i*, *k*, *l*. Les ressorts illustrent le caractère harmonique (oscillatoire) du mouvement des atomes en interactions et les flèches rouges schématisent la direction, le sens et l'amplitude du mouvement. Les termes r_{eq} , θ_{eq} , ξ_{eq} se réfèrent aux valeurs à

l'équilibre de longueur de liaison, d'angle de valence, et de distances de dièdre impropre. ϕ est l'angle de torsion du dièdre propre.

La description mathématique de chacun de ces termes d'interactions empiriques est déterminée sur la base de résultats expérimentaux. Chaque terme se rapporte à une valeur d'énergie propre à l'interaction qu'il définit et dont la valeur dépend de plusieurs paramètres. L'expression de la fonction d'énergie potentielle associée au jeu de paramètres empiriques décrivant chaque atome (masse, charge ponctuelle) ainsi que la topologie de leurs interactions (longueur de liaison, d'angle de valence, et de dièdres à l'équilibre) représente le champ de force. L'équation ci-dessous décrit la fonction d'énergie implémentée au sein du champ de force AMBER généralisé¹⁸¹:

$$E_{totale} = \sum_{liaison} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{torsions} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} + \frac{q_i q_j}{\epsilon r_{ij}} \right]$$

Les trois premiers termes de la fonction d'énergie se rapportent aux interactions entre toutes les paires d'atomes distants d'au plus trois liaisons covalentes (*Figure 23*). Ils quantifient les variations d'énergie engendrée par l'élongation des liaisons covalentes entre deux atomes liés, la déformation des angles de valences entre trois atomes liés, ainsi que la torsion des angles dièdres entre quatre atomes liés (dièdres propres). A ces termes d'interactions peuvent s'ajouter des termes correctifs de dièdres impropres. Les dièdres impropres décrivent une contrainte spatiale affectant un groupe de quatre atomes ne se suivant pas séquentiellement. L'énergie d'une telle contrainte s'exprime comme suit :

$$E_{impropres} = \sum_{distances} \frac{1}{2} K_\xi (\xi - \xi_{eq})^2$$

Cette contrainte est souvent appliquée pour maintenir la planarité de certains groupements comme la liaison peptidique ou les groupements carboxyles et amides. Ce terme est aussi employé pour maintenir la stéréochimie des centres anomériques tels les C α des aminoacides en empêchant toute inversion de configuration L en configuration D au cours d'une simulation.

A l'exception des angles dièdres propres, la dynamique de toutes les interactions liées se caractérise par des mouvements d'oscillations harmoniques entre deux atomes

ou entre un atome et un plan (dièdre impropre). En d'autres termes, ce formalisme indique que la variation d'énergie liée à chacune de ces interactions peut être modélisée sous la forme d'une fonction d'énergie de forme parabolique. Les termes $r - r_{eq}$, $\theta - \theta_{eq}$, $\xi - \xi_{eq}$ quantifient respectivement les variations de longueurs de liaison, d'angles de valences, de dièdres impropres par rapport à leurs valeurs respectives à la position d'équilibre. Les paramètres K_r , K_θ , K_ξ ou constantes de force, quantifient la fréquence des oscillations harmoniques associées à chaque type d'interaction. Les variations d'angles dièdres propres sont exprimées sous la forme d'une série de Fourier où n est la périodicité, γ l'angle de phase du dièdre ϕ et V_n la constante de torsion du dièdre. Les valeurs des constantes de forces sont en général dérivées d'analyses de molécules modèles par spectroscopie vibrationnelle ou obtenues à partir de calculs *ab-initio*.

Les termes d'interactions non-liés quantifient l'énergie des interactions entre atomes distants d'au moins quatre liaisons covalentes. Ils comprennent les interactions de Van der Waals et les interactions électrostatiques. Les interactions de Van der Waals entre deux atomes i et j distants l'un de l'autre d'une distance r_{ij} sont décrites par un potentiel de Lennard-Jones comportant deux termes d'interactions dits répulsif et attractif. Le terme répulsif $\frac{A_{ij}}{r_{ij}^{12}}$ rend compte du principe d'exclusion de Pauli stipulant que les nuages électroniques de deux atomes en interaction ne peuvent s'interpénétrer. Ce terme empêche donc deux atomes distants de se rapprocher trop l'un de l'autre. Le second terme attractif $\frac{B_{ij}}{r_{ij}^6}$ décrit l'attraction de deux atomes induit par les mouvements des électrons autour de leurs noyaux respectifs (forces de London). Les paramètres A_{ij} et B_{ij} sont des constantes propres à un couple d'atomes i, j donné.

$$E_{VanderWaas} = \sum_{i < j} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

L'énergie des interactions électrostatiques sont calculées d'après la loi de Coulomb et correspondent aux interactions entre charges partielles (q_i et q_j) portées par deux atomes i et j distants l'un de l'autre d'une distance r_{ij} . Les charges propres à chaque atome sont en général déterminées par le biais de calcul *ab-initio*. Le paramètre ϵ se

réfère à la constante diélectrique du milieu. Elle quantifie l'atténuation des interactions électrostatiques locales dues à l'environnement (solvant et protéine).

$$E_{electro} = \sum_{i < j} \left[\frac{q_i q_j}{\epsilon r_{ij}} \right]$$

Pour des raisons de temps de calcul, l'utilisation de l'équation décrite ci-dessus est cependant limitée aux interactions entre atomes espacés les uns des autres en deçà d'une certaine distance de troncature (souvent choisie entre 8 et 10 Å) associée au champ de force utilisé. L'utilisation d'un seuil de troncature engendre cependant une approximation importante dans la détermination du potentiel de Coulomb puisque les interactions électrostatiques à longues distances sont négligées. Afin de pallier à cette limitation, l'emploi de conditions périodiques s'avère donc indispensable. Dans le cas d'un système périodique, les interactions électrostatiques entre un atome donné et l'ensemble de ses images doivent également être prises en compte. Afin de calculer efficacement les interactions à longues distances, l'équation précédente peut donc être décomposée en deux termes. Le premier terme, aussi appelé somme directe, est utilisé pour le calcul de l'énergie liée aux interactions électrostatiques à courtes distances (en deçà d'un seuil de troncature défini). Le second terme, appelé somme d'Ewald, permet le calcul de l'énergie liée aux interactions à longues distances. En dépit de sa complexité, la résolution numérique de la somme d'Ewald peut être réalisée efficacement par le biais de transformées rapides de Fourier (Algorithme Particle Mesh Ewald, ou PME ¹⁸²).

3.4 La dynamique moléculaire

3.4.1 Principe

La dynamique moléculaire vise à décrire l'évolution des positions atomiques (leur trajectoire) d'un système au cours du temps. Exploitant la fonction d'énergie définie par la mécanique moléculaire, cette technique quantifie l'énergie potentielle caractérisant chaque atome du système à des petits intervalles de temps régulièrement espacés (de l'ordre de la femtoseconde). La valeur d'énergie potentielle propre à un atome à un instant donné de sa trajectoire reflète l'ensemble des interactions qui s'exercent sur celui-ci. L'énergie communiquée à un atome donné *via* ses interactions avec les atomes voisins (j, \dots, n) correspond elle-même à une force résultante exprimée selon la relation :

$$F_i = -\nabla_i E(r_j, \dots, r_n) = -\frac{dE(r_j, \dots, r_n)}{dr_i(t)}$$

La force s'appliquant à chaque atome i du système peut donc être directement déterminée par différenciation de la fonction d'énergie potentielle au voisinage de sa position $r_i(t)$. La valeur de cette force est calculée à chaque pas de temps. Elle détermine la variation de la vitesse de chaque atome du système (et donc son mouvement) d'un instant à l'autre au cours de sa trajectoire. Le mouvement des atomes au cours du temps est ainsi décrit par l'équation fondamentale de la dynamique dérivée de la seconde loi de Newton :

$$\vec{F} = m \times \vec{a}$$

\vec{F} étant le vecteur force s'appliquant sur l'atome donné, m sa masse et \vec{a} le vecteur accélération. On remarque ainsi que l'accélération d'un atome dépend de sa masse intrinsèque et de l'ensemble des forces qui s'exercent sur celui-ci à un instant donné de sa trajectoire. Connaissant l'énergie potentielle s'exerçant sur un atome donné et donc la force correspondante, il est donc possible d'en déduire son accélération. La vitesse propre à chaque atome \vec{v} peut ensuite être déterminée à partir de l'accélération selon la relation :

$$\vec{a}_i = \frac{d\vec{v}_i}{dt}$$

La position \vec{r} de l'atome peut finalement être obtenue à partir des vitesses atomiques selon l'expression :

$$\vec{v}_i = \frac{d\vec{r}_i}{dt}$$

Afin d'initier le mouvement des tous les atomes du système en début de simulation, il est nécessaire de leur communiquer une vitesse initiale. La vitesse initiale attribuée à un atome donné est en général choisie de manière aléatoire selon une distribution probabiliste (distribution de Maxwell-Boltzmann) centrée sur la température de simulation du système. En pratique, l'intégration des équations du mouvement est réalisée avec un pas de temps très court (ou pas d'intégration noté Δt , souvent compris entre 1 et 2 fs) afin de décrire le plus précisément possible l'évolution de l'énergie potentielle et donc des positions atomiques au cours du temps. L'intégration des

équations du mouvement peut être réalisée numériquement à l'aide de divers algorithmes.

3.4.2 Algorithmes d'intégration des équations du mouvement

L'algorithme du Leap Frog est souvent utilisé pour calculer le déplacement d'un atome donné au cours d'un pas d'intégration. Cet algorithme utilise la position $r(t)$ d'un atome donné à l'instant t , les forces $F(t)$ s'appliquant sur cet atome, ainsi que sa vitesse v à l'instant $t - \frac{1}{2}\Delta t$ pour déterminer sa position à l'instant $t + \Delta t$ selon l'équation :

$$r(t + \Delta t) = r(t) + \Delta t v(t + \frac{1}{2}\Delta t) + o(\Delta t)$$

La vitesse étant exprimée sous la forme (où m est la masse de l'atome):

$$v(t + \frac{1}{2}\Delta t) = v(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m} F(t)$$

L'intégration successive de ces équations pour tout atome du système sur un ensemble d'intervalles de temps successifs détermine la trajectoire de la dynamique moléculaire. En alternative à la méthode précédente, l'algorithme de Verlet peut également être utilisé. Cet algorithme calcule la position d'un atome donné à l'instant $t + \Delta t$ via l'équation suivante :

$$r(t + \Delta t) = 2r(t) - r(t - \Delta t) + \left(\frac{F}{m}\right)\Delta t^2 + o(\Delta t^4)$$

Avec :

$$r(t - \Delta t) = r(t) - v\Delta t + \frac{1}{2}\left(\frac{F}{m}\right)\Delta t^2 - o(\Delta t^3)$$

La vitesse de l'atome à l'instant t est calculée par la relation :

$$v(t) = \frac{r(t + \Delta t) - r(t - \Delta t)}{2\Delta t} + o(\Delta t^2)$$

Un des inconvénients majeurs de la méthode de Verlet reste son imprécision liée en partie à la troncature réalisée lors du calcul des vitesses. Afin de pallier à cette limitation, d'autres algorithmes plus précis tels que l'algorithme de Beeman ont également été développés.

3.4.3 Ensembles thermodynamiques

Les algorithmes d'intégration des équations du mouvement présentés plus haut s'appliquent à l'étude de systèmes moléculaires à l'équilibre thermodynamique. Dans cet état, l'énergie interne totale du système est conservée. On dit alors que le système moléculaire simulé évolue dans l'ensemble thermodynamique microcanonique (ou NVE, Nombre de particules, Volume, et Energie interne constants).

La température et la pression d'un système moléculaire simulé dans l'ensemble NVE peuvent cependant évoluer. Ainsi, afin de pouvoir comparer une simulation avec des résultats expérimentaux, il est souvent d'usage de coupler le système à un thermostat et à un barostat dans le but de maintenir sa température et sa pression proches des valeurs utilisées au laboratoire.

On dit alors que le système évolue dans un ensemble thermodynamique à température et pression constante ou ensemble NPT. Dans ce cas, d'autres algorithmes sont couplés aux algorithmes classiques d'intégration des équations du mouvement dans le but de moduler la dynamique du système pour compenser les variations la température et la pression au cours de la simulation.

3.4.4 Ergodicité

Lors d'une simulation de dynamique moléculaire, l'enregistrement des coordonnées atomiques du système à des intervalles de temps réguliers en cours de simulation permet d'obtenir une série de clichés permettant notamment de reconstituer la trajectoire des molécules d'eau internes à la protéine d'intérêt.

Afin d'obtenir un échantillonnage exhaustif de la trajectoire des atomes constituant un système moléculaire donné, il est d'usage de poursuivre une simulation de dynamique moléculaire sur un intervalle de temps « suffisamment long ». Puisque la dynamique d'un système moléculaire peut être décrite par un ensemble d'états conformationnels distincts, une simulation suffisamment longue devrait en théorie laisser au système le temps d'adopter toutes les conformations possibles.

Le choix de l'intervalle de temps est déterminé par l'utilisateur et dépend du phénomène physique (changement conformationnel, mouvement de chaînes latérales, diffusion de molécules d'eau...) que l'on souhaite observer durant la simulation. En pratique, il est intéressant d'effectuer une simulation au moins dix fois plus longue que le temps minimum requis pour observer un phénomène physique donné.

L'intérêt de réaliser une simulation sur un temps long repose sur le principe d'ergodicité. En supposant que la dynamique d'un système moléculaire soit décrite par une observable donnée, il est possible d'estimer le comportement dynamique moyen de ce système en calculant la moyenne de cette observable sur un intervalle de temps « suffisamment long ». Cela revient également à considérer que le comportement moyen d'un système simulé sur une longue trajectoire est identique à celui d'un ensemble infini du même système simulé sur une trajectoire courte.

- Chapitre II -

Matériel et méthodes

4 Matériel

4.1 Liste des milieux, tampons, et réactifs utilisés

- **Cultures bactériennes**

LB : Lysogeny Broth. Peptone 10 g, extrait de levure 5 g, NaCl 10 g, eau distillée qsp 1 L

LB-Amp : milieu LB supplémenté en ampicilline (100 µg/mL)

ZYP5052 (medium 26): milieu d'auto-induction pour la production d'AgaD

- **Réaction de polymérisation en chaîne (PCR) mutagène**

dNTP :

. Ttβgly : dATP et dTTP 3 mM, dGTP et dCTP 7 mM dans de l'eau distillée.

. AgaD/TcTS : dATP, dTTP, dGTP, dCTP 2 mM dans de l'eau distillée.

- **Préparation de cellules compétentes**

TSS : LB avec 10 % de PEG 8000, 5 % de DMSO, 30mM de MgCl₂, pH 6

- **Purification de protéines**

Tampon de lyse :

. Ttβgly/TcTS : tampon phosphate 60mM pH 8, imidazole 10mM, NaCl 0,2M

. AgaD : Tris-HCl 50 mM pH 7,5, NaCl 300 mM, MgCl₂ 1 mM, CaCl₂ 1 mM,
Imidazole 50 mM.

Tampon de lavage :

. Ttβgly/TcTS : tampon phosphate 60mM pH 8, imidazole 25mM, NaCl 0,2M

. AgaD : Tris-HCl 50 mM pH 7,5, NaCl 300 mM, MgCl₂ 1 mM, CaCl₂ 1 mM,
Imidazole 70 mM.

Tampon d'élution:

. Ttβgly/TcTS : tampon phosphate 60mM pH 8, imidazole 250mM, NaCl 0,2M

. AgaD : Tris-HCl 50 mM pH 7,5, NaCl 300 mM, MgCl₂ 1 mM, CaCl₂ 1 mM,
Imidazole 200 mM.

IPTG : Isopropyl β-D-thiogalactoside

PBS : NaCl 137mM; KCl 2,7mM; Na₂HPO₄ 10mM; KH₂PO₄ 1,76mM; pH 6,4

- **Contrôle de pureté par SDS-PAGE**

Tampon de charge : Lane-Marker Non Reducing Sample Buffer 5x (Thermo scientific)

- **Cinétique enzymatique**

Tampons d'activité :

. Ttβgly : (voir le tampon PBS)

. AgaD : MOPS 500 mM, NaCl 3 M, pH 7,5

. TcTS : tampon phosphate 100 mM, pH 6

Borax : Tétraborate de sodium ($\text{Na}_2\text{B}_4\text{O}_7 \cdot 10\text{H}_2\text{O}$)

Tampon de migration (électrophorèse capillaire) : Borax 50 mM; pH 10,4

- **Caractérisation d'activité par chromatographie sur couche mince (CCM)**

Révéléateur orcinol : 0,3 g d'orcinol, 30mL acide sulfurique, 270 ml éthanol absolu

- **Liste des principaux réactifs utilisés et leurs fournisseurs :**

Thermo Scientific : Bleu de Coomassie, Pfu DNA polymerase, Pfu buffer+MgSO₄, dNTP

Thermo Fisher : PageRuler™ Prestained Plus

VWR : KCl, NaCl, IPTG, Imidazole, Ac. Sulfurique >99 %, Ac. Acétique, Ethanol, Butanol

Finnzyme : *DpnI*

Merck : Na₂HPO₄, KH₂PO₄

Sigma-Aldrich : ONPG, Borax, Orcinol, Borax, NaBH₃CN, MOPS, sialyl-PNP, 3'SL

Applichem-Panreac : Imidazole 99 %

Eurofins : Amorces PCR

4.1.1 Appareillage utilisé

- **Mutagenèse dirigée**

Thermocycleur: MJ Mini™ (Biorad)

Incubateur: Incu-Shaker Mini® (Benchmark)

- **Cultures bactériennes**

Incubateur/Agitateur: Minitron® (Infors HT)

Spectrophotomètre: Genesys™ 20 (Thermo Scientific)

- **Purification de protéines**

Nanodrop® 1000 (Thermo Scientific)

Centrifugation pots 1L: SORVALL® RC6 plus (Thermo electron)

Centrifugation tubes Falcon® 50 mL: Eppendorf centrifuge 5810R

Centrifugation tubes eppendorf 1,5 mL: Eppendorf centrifuge 5424

Centrifugation Vivaspin® : Jouan centrifuge Cr3i (Thermo electron)

Migration des gels SDS-PAGE: Mini PROTEAN® tetra cell (Biorad)

pH-mètre: Mettler-Toledo SevenEasy™

Sonicateur: Vibra-Cell™ (Thermo Scientific)

- **Cinétiques enzymatiques**

Lecteurs de plaque : TECAN Infinite M1000 ; Labsystem integrated EIA

Electrophorèse capillaire : Beckman P/ACE System 5000

HPEAC-PAD : Dionex™ ICS5000 équipé d'une colonne Carbopac™ PA100 (4 x 250 mm)

4.1.2 Souches bactériennes utilisées

Souche	Génotype	Résistance	Fournisseur
XL1-Blue MRF' (<i>Escherichia Coli</i>)	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F'proABlacIqZΔM15Tn10]	tétracycline	Stratagene
BL21 (DE3) (<i>Escherichia Coli</i>)	F- ompT hsdSB(rB-mB-) gal dcm (DE3)	-	Invitrogen

4.1.3 Plasmides

Le plasmide d'expression utilisé pour la transformation du gène de la Ttβgly est la pBTac2. Ce plasmide contient le gène de la forme sauvage de la Ttβgly (identifiant GenBank: Y16753) cloné entre les sites de restriction *EcoRI* et *PstI*. Il possède un opérateur Lac ainsi qu'un promoteur de transcription Ptac permettant son expression au sein de la souche utilisée. En vue de la purification ultérieure de la protéine correspondante par chromatographie d'affinité, le gène de la Ttβgly a été préalablement fusionné à une étiquette multi-histidine (6xHis-Tag) en C-terminal (*Figure 24*).

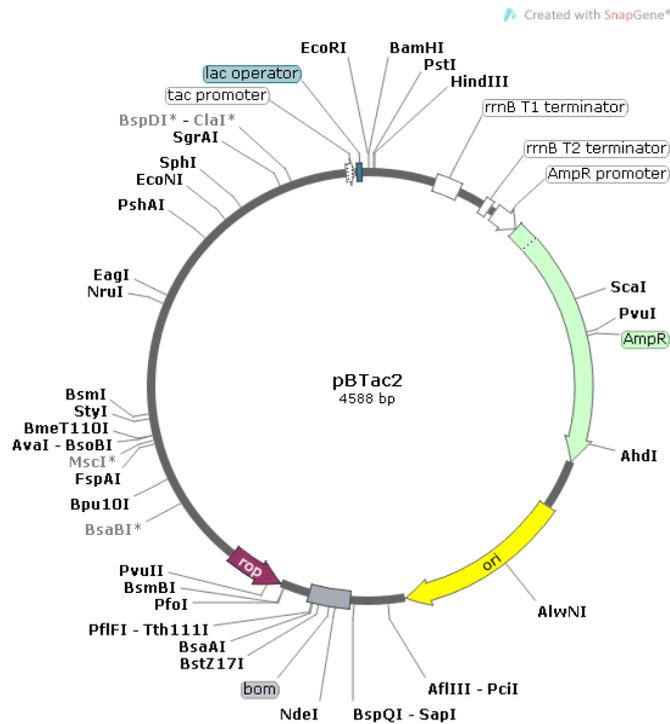


Figure 24. Carte du plasmide pBTac2 utilisé pour l'expression de la *Ttβgly* à titre d'exemple. La figure a été créée à l'aide du logiciel SnapGene Viewer v3.2.

Le plasmide d'expression utilisé pour la transformation du gène de TcTS est le pTrcHisB. Le gène de TcTS a été inséré entre les sites de restriction *NheI* et *EcoRI* et est a été préalablement fusionné à une étiquette hexa-histidine en N-terminal. Le plasmide d'expression utilisé pour la transformation du gène d'AgaD est le pFO4¹⁸⁷. L'ensemble des plasmides utilisés possèdent un gène de résistance à l'ampicilline permettant la sélection ultérieure des clones recombinants sur milieu gélosé sélectif LB-Amp,

5 Méthodes

5.1 Méthodes expérimentales

Plusieurs étapes (*Figure 25*) ont été nécessaires pour la production et à la caractérisation des enzymes étudiées dans ce projet. Ces étapes seront décrites chronologiquement.

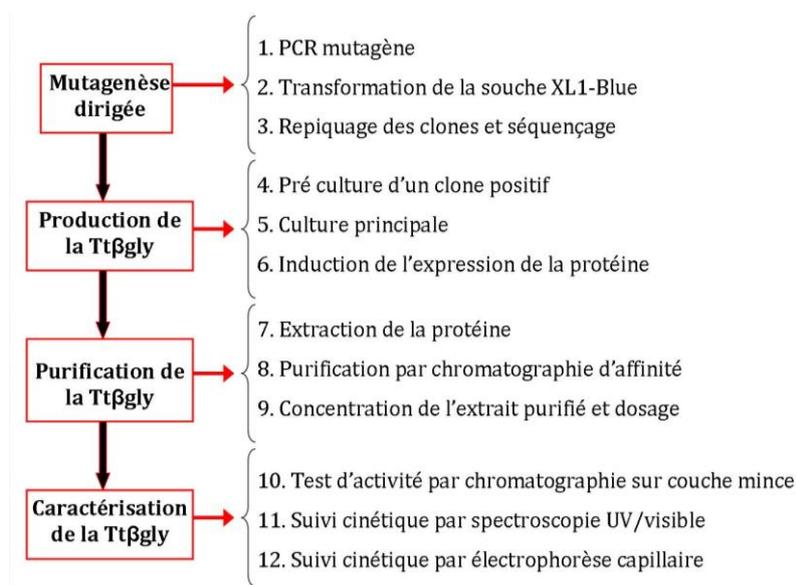


Figure 25. Schéma récapitulatif de l'ensemble des étapes expérimentales appliquées pour la production et la caractérisation de Ttβgly à titre d'exemple.

5.1.1 La PCR mutagène

La PCR mutagène a été utilisée afin d'introduire une mutation ponctuelle sur un codon spécifique du gène des enzymes étudiées dans le cadre de ce projet de thèse (Ttβgly, AgaD, TcTS). Cette méthode permet ainsi de substituer un résidu spécifique dans la séquence de la protéine par un autre préalablement choisi. Son principe repose sur l'hybridation complémentaire d'un couple d'oligonucléotides mutagènes à la région du gène d'intérêt dans laquelle se trouve le codon à muter. L'incorporation du codon muté au sein des deux brins complémentaires du gène d'intérêt est ensuite réalisée par une polymérase lors d'un cycle de PCR classique.

La préparation du milieu de réaction pour la PCR consiste à mélanger de l'ADN plasmidique en présence des amorces mutagènes sens et anti-sens, et d'un mélange de désoxyribonucléotides (dATP, dTTP, dCTP, dGTP). 1 µL de Pfu ADN polymérase est finalement ajouté au milieu en présence de son tampon d'activité. Un total de 18 cycles de PCR est ensuite réalisé selon le protocole suivant :

Composition du milieu réactionnel :

ADN plasmidique 5-50 ng

Programme PCR :

95°C 30 sec.

Amorce sens (10 mM)	1 µL	95°C	30 sec.	} x 18
Amorce anti-sens (10 mM)	1 µL	55 °C	1 min.	
Mélange dNTP*	1 µL	68°C	7 min.	
Tampon Pfu polymerase 10x	5 µL	68°C	10 min.	
Pfu ADN polymerase (2,5 U/µL)	1 µL	14°C	∞	
Eau milliQ	qsp 50 µL			

Le gène de la Ttβgly étant issu d'un organisme thermophile, celui-ci est particulièrement riche en nucléotides C et G (70% de sa séquence totale). Afin de respecter ce biais de composition lors de l'amplification par PCR, des proportions identiques de nucléotides ont été introduits au sein des milieux de réaction (30% dATP et dTTP ; 70% dCTP et dGTP). Le protocole de PCR mutagène a été réalisé *via* le kit Quick Change® Mutagenesis (Stratagene) selon les indications décrites par le fournisseur.

- **Liste des amorces mutagènes (5' ->3') utilisées dans ce projet :**

- i. Mutagenèse de Ttβgly :

CTCTACCACTTTGACCTGCCTTTGGCC	W120F
CCCTCTACCACATTGACCTGCCTTTGGCC	W120I
GGGCACTGGACTGGAGAACTTGCACCCGGCCTCAGGAACC	H178L
ACGGGGGAAATCGCACCCGGCCTC	H178I
ACGGGGGAATTTGCACCCGGCCTC	H178F
GGGCACTGGACTGGAGAACTTGCACCCGGCCTCAGGAACC	A179P
TACACCCGGCGCTTTGGCCTCTACTACG	S401F
CGAGCCCTGGATGTCTGCCTTCCTCG	C167M
GAGCCCTGGATCTCTGCCTTCCTCG	C167I
CGAGCCCTGGGTCTCTGCCTTCCTCG	C167V
CCCTGGTGCATTGCCTTCCTCG	S168V
CCCTGGTGCATTGCCTTCCTCG	S168I
GCCTTCCTCGGGTGGTGGACGGGGAAACACG	H173W
GAACACGCCCCCGGTTCGCAGGAACCTGGAAGC	L182R

ii. Mutagenèse d'AgaD :

gatggatatcatcattaatgtggaattacaaagttggcaagccgaagcctatc	D341L
caagccgaagcctatcgag <u>cc</u> ccaatgct	S351A
ggatatcatcattaatgtggaattac <u>at</u> agttggcaagccga	Q342H
gcaagccgaagcctatcgat <u>tt</u> ccaatgctgcg	S351F

iii. Mutagenèse de TcTS :

GTCCTGGTTGGAAGCTACAACAAATCGAGGAGCTACTGGACGTCGCATGG	S115K
GCTACAACAGTTCGAGGAACTACTGGACGTCGCATGGTG	S118N
AAGGGAAGGTGGATGCGCGACCGACTGAACC	L313M

5.1.2 Transformation des souches bactériennes XL1-Blue et BL21

Après PCR, un mélange de plasmides, dont certains contiennent le gène parental et d'autres le gène muté, est obtenu. Afin d'éliminer les plasmides parentaux n'ayant pas incorporé la mutation, 1 µL d'enzyme de restriction *DpnI* est ajouté au mix PCR. Le tout est ensuite incubé pendant 2 h à 37°C. La transformation du mélange résultant au sein d'un aliquote de cellules compétentes XL1-Blue est finalement réalisée par choc thermique. 5 µL du mélange PCR (non purifié) est introduit dans un volume de 80 µL de suspension cellulaire préalablement décongelée dans la glace. Après une incubation de 10 min, le mélange est porté à 42°C pendant 45 s et réintroduit pendant 2 min dans la glace. 800 µL de milieu LB est par la suite rajouté au mélange afin de permettre aux cellules de se régénérer suite au choc thermique. Le tout est finalement incubé à 37°C pendant 1 à 2 h. Une centrifugation du mélange est ensuite réalisée pendant 5 min à 1500 g à température ambiante afin de précipiter les cellules. Après élimination de 900 µL de surnageant, les cellules sont resuspendues dans le volume restant avant d'être isolé sur une gélose de LB supplémentée d'ampicilline. L'incubation de la gélose à 37°C est réalisée sur la nuit pour permettre le développement de colonies.

5.1.3 Repiquage des clones et séquençage

Un ensemble de colonies suspectées de contenir le plasmide d'intérêt est sélectionné aléatoirement sur la surface de la gélose. Chaque clone est repiqué dans 5 à 10 mL de milieu LB-Amp et incubé sur la nuit à 37°C. Après culture, 500 µL de chaque suspension bactérienne obtenue est conservé en présence de 250 µL de glycérol 80% à -80°C en attendant les résultats du séquençage. Dans le but de déterminer la séquence du gène muté, une préparation d'ADN plasmidique est réalisée sur 5 mL de suspension bactérienne par l'intermédiaire du kit QIAprep® Spin Miniprep Kit (Qiagen). L'élution de l'ADN purifié est effectuée par ajout de 45 µL d'eau distillée chauffée à 70°C. La concentration d'ADN extrait est ensuite déterminée par mesure d'absorbance à 260 nm (Nanodrop®). Entre 100 et 500 ng d'ADN est finalement séquencé. Pour chaque clone, la réaction de séquençage est réalisée en présence de 2,5 µL d'amorce sens et anti-sens à 10 µM et ajustées à 10 µL avec de l'eau milliQ. Le séquençage est effectué par la société GATC biotech. La qualité des séquences a été évaluée par le logiciel Snapgene® Viewer. Un alignement de la séquence sauvage de l'enzyme étudiée avec celle issue du séquençage est finalement réalisé par l'algorithme de Smith et Waterman¹⁸⁸ issu de la suite logicielle EMBOSS (https://www.ebi.ac.uk/Tools/psa/emboss_water/) afin de localiser les mutations introduites. Parmi l'ensemble des clones conservés à -80°C, seuls les clones positifs (ayant incorporé la mutation désirée) sont conservés.

5.1.4 Préculture d'un clone positif

Afin de préparer une production d'enzyme, un extrait de clone préalablement conservé à -80°C est introduit dans 10 mL de milieu LB-Amp et incubé à 37°C sur la nuit. Cette préculture va permettre de raviver la souche congelée et d'obtenir une densité cellulaire suffisante pour l'ensemencement ultérieur d'un volume de culture plus important.

5.1.5 Culture principale et induction de l'expression de la protéine

Dans le cas d'une production de Ttβgly et de TcTS, 1L de milieu LB supplémenté avec 100 µg/mL d'ampicilline est ensemencé le jour suivant avec la préculture de la veille et est incubé dans un premier temps à 37°C. Un suivi régulier de la densité optique permet

d'induire l'expression du gène de la protéine lorsque la culture atteint sa phase exponentielle de croissance. 0,5 mM d'inducteur d'expression (IPTG) sont ajoutés au milieu lorsqu'une densité optique de 0,6 est atteinte. Celui-ci est finalement incubé à 30°C sous agitation et sur la nuit. Dans le cas d'une production d'AgaD, les cellules issues d'une première préculture sont mises en culture continue pendant trois jours à 20°C dans 1L de milieu ZYP5052 supplémenté avec 100 µg/mL d'ampicilline.

5.1.6 Lyse des cellules

Après production, l'enzyme est extraite en vue de sa purification. Après culture, une centrifugation des cellules (5000 rpm, 20 min, 4°C) est effectuée. Le culot cellulaire est ensuite repris dans un volume de tampon de lyse (à raison d'environ 1 mL de tampon pour 5 mg de culot) en présence d'un inhibiteur de protéases (PMSF, 100 mM final). La suspension obtenue est ensuite soumise à 10 cycles de sonication (5 pulses de 10 s chaque, amplitude fixée à 30 % de la valeur maximale) dans la glace afin de lyser les cellules. Dans le cas de Ttβgly uniquement, le surnageant contenant les protéines est ensuite récupéré et incubé à 75°C pendant 30 min. Ttβgly étant thermostable, cette étape vise à dénaturer tout autre contaminant protéique en solution et permet ainsi d'augmenter le rendement de purification. Une centrifugation finale (10000 rpm, 30 min, 4°C) est finalement effectuée sur l'extrait de lyse afin de culoter les débris cellulaires et autres corps d'inclusion (la protéine d'intérêt est récupérée dans le surnageant).

5.1.7 Purification des protéines par chromatographie d'affinité Ni-NTA

La protéine possédant une étiquette multi-histidine en position C-terminale, il est donc possible de la purifier par chromatographie d'affinité sur nickel immobilisé (IMAC). Les groupements imidazole des histidines possèdent en effet une forte affinité avec les ions nickel. Un volume d'environ 300 µL de résine porteuse de groupements nitrilo-acétate complexés aux ions nickel (résine Ni-NTA agarose, Qiagen) est utilisé pour une colonne de purification. La résine est équilibrée avec 5 mL de tampon de lavage. L'extrait brut y est ensuite déposé à deux reprises afin de favoriser la fixation de la protéine sur la résine. La colonne est ensuite lavée avec 10mL de tampon de lavage afin d'éliminer les

contaminants protéiques résiduels faiblement affins pour le nickel. La protéine immobilisée est finalement éluée sous la forme de fractions de 1 à 2 mL chacune par l'ajout de 1 à 6 mL de tampon d'éluion. Les fractions les plus concentrées sont ensuite analysées par SDS-PAGE. Dans les cas de Ttβgly et de TcTS, la détermination de la concentration en protéine est réalisée sur 2 μL de fraction d'éluion par spectrophotométrie d'absorption à 280 nm (Nanodrop®). La mesure utilise la valeur du coefficient d'extinction molaire ainsi que la masse (en kDa) de la protéine d'intérêt. Ces paramètres ont été préalablement calculés *via* le programme ProtParam (<http://web.expasy.org/protparam/>) en utilisant la séquence de la forme sauvage de chaque protéine. La concentration des éluats d'AgaD est en revanche déterminée par électrophorèse capillaire (Agilent 2100 Bioanalyser) et confirmée par dosage au Bradford en utilisant l'albumine bovine sérique comme étalon.

5.1.8 Analyse de la pureté des fractions d'éluion par SDS-PAGE

En vue d'estimer la pureté des fractions éluées, celles-ci sont analysées par électrophorèse sur gel de polyacrylamide (10%) en présence de sodium dodécyl sulfate (SDS-PAGE¹⁸⁹). Avant toute séparation, un extrait de chaque échantillon purifié (soit 2 à 3 μg de protéines) est dénaturé à 95°C en présence de tampon de charge. Chaque fraction dénaturée est centrifugée puis déposée sur le gel. La migration est effectuée à 120 V au sein d'un tampon de migration tris-glycine. Après séparation, les fractions sont révélées par coloration au bleu de Coomassie (*Figure 26*). La fraction contenant la protéine d'intérêt peut être identifiée par l'utilisation d'un marqueur de taille (PageRuler™ Prestained Plus) ayant préalablement été séparé dans les mêmes conditions. Les échantillons possédant les fractions les plus denses sont alors regroupés en un pool commun.

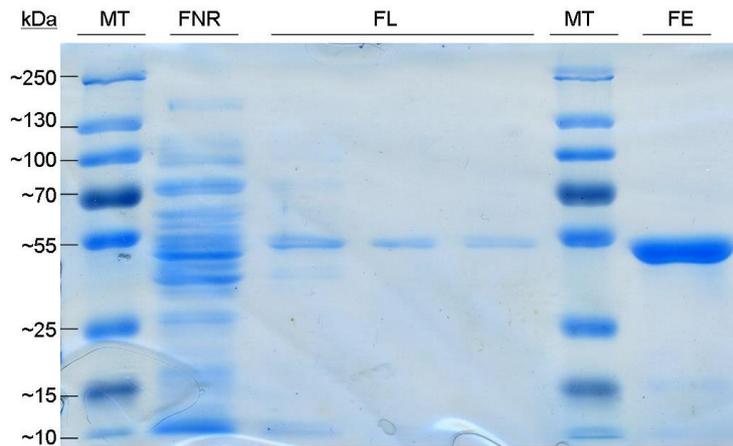


Figure 26. Exemple de gel SDS-PAGE obtenu pour le contrôle de la purification de la Tt β gly.

La piste MT se réfère à la séparation des marqueurs de taille Protéiques (mélange de protéines témoins de taille connues) servant à localiser la protéine d'intérêt basé sur sa taille. Les masses de chacune des protéines témoins sont indiquées à gauche en kDa. La piste FNR se réfère à la séparation des protéines contenues au sein de la fraction non-retenue sur colonne d'affinité. Les trois fractions FL se réfère à la séparation de trois fractions de lavage distinctes. La fraction FE se réfère à la fraction éluée en fin de purification et contenant la protéine d'intérêt (Tt β gly). Le colorant utilisé est le bleu de Coomassie.

5.1.9 Echange de tampon et concentration des protéines purifiées

Cette étape n'a été réalisée que dans le cadre d'une production de Tt β gly et de TcTS. Afin de pouvoir conditionner l'enzyme dans un tampon favorisant son activité, le tampon d'élution doit être remplacé par un tampon d'activité (ici le PBS). Une unité d'ultrafiltration (6 mL, 30 000 MWCO, Vivaspin®) est utilisé. Une étape de lavage de la membrane de filtration est tout d'abord réalisée afin de se débarrasser des additifs présents à sa surface et pouvant ultérieurement interférer avec les tests d'activité enzymatiques. Le lavage consiste à centrifuger (4000 rpm, 4 min, 4°C) à 4 reprises l'unité en présence de 6 mL d'eau distillée. Cette opération est répétée avec un volume équivalent de PBS afin d'équilibrer la membrane. Le pool de protéines purifiées y est finalement déposé. Une première centrifugation à 4000 rpm (~ 6-7 min, 4°C) permet de concentrer l'échantillon. Le volume résiduel non filtré est ensuite ajusté à environ 6 mL par ajout de PBS et à nouveau centrifugé dans les mêmes conditions que la première centrifugation. L'étape précédente est ensuite à nouveau répétée à 6 reprises afin d'échanger le tampon d'élution contre le PBS. Une dernière centrifugation est finalement

réalisée pour concentrer l'échantillon. Le volume final après centrifugation est ajusté à 1 mL avec du PBS. Ce volume final est récupéré après resuspension de la protéine adsorbée sur la membrane. La concentration de protéine resuspendue est déterminée par mesure au Nanodrop.

5.1.10 Test d'activité enzymatique par chromatographie sur couche mince

Ce test nous permet d'avoir un aperçu qualitatif de l'activité enzymatique par l'estimation des proportions de produits d'hydrolyse et de transglycosylation obtenus après réaction. Il est utilisé ici à des fins de criblage rapide nous permettant ainsi d'orienter l'analyse vers des tests quantitatifs dès lors qu'une activité de transglycosylation augmentée a été mise en évidence chez un mutant. L'activité de la Tt β gly est évaluée sur l'ONPG qui joue ici à la fois le rôle de composé donneur et accepteur de glycosyle. La réaction enzymatique est déclenchée en ajoutant 6,6 μ L d'enzyme à 0,1 mg/mL à 200 μ L d'ONPG à 25 mM en tampon PBS. Le milieu réactionnel est ensuite incubé à 40°C pendant 3h. Un volume de 1 μ L est prélevé toutes 30 à 60 min et déposé sur plaque de silice.

Afin d'estimer l'activité de transglycosylation d'AgaD, 12 μ L d'enzyme purifiée sont mélangés à 3 μ L d'agarodécaose (LA-G)₅ à 10 mg/mL auxquels sont ajoutés 3 μ L de tampon de réaction (MOPS 500 mM, NaCl 3 M, pH 7,5) et 12 L d'eau distillée. Le milieu réactionnel est finalement incubé à 25°C sous agitation et plusieurs prélèvements sont ensuite réalisés à 2, 4, 6, 8, et 24 h. L'agarodécaose a été préparé à partir de l'hydrolyse par AgaD d'une solution d'agarose commercial à 0,5 g/L.

L'activité hydrolytique et de transglycosylation de TcTS a été évaluée *via* deux tests (test d'hydrolyse et test de transglycosylation) effectués séparément. Le test de transglycosylation a été réalisé en incubant 0,6 μ g d'enzyme avec 2 μ L de sucre donneur 3'SL et 2 μ L de sucre accepteur (LNT ou lactose) à 200 mM dans du tampon NaH₂PO₄ à pH 6, qsp 12 μ L final. Le test d'hydrolyse a été effectué dans les mêmes conditions à l'exception que le sucre accepteur est absent du milieu réactionnel. Chaque test d'activité a été effectué avec diverses concentrations en enzyme (dilutions 1/128, 1/64 et 1/32). Le but est d'employer plusieurs dilutions d'enzyme est de mimer un suivi cinétique de la réaction enzymatique. Dans un suivi

cinétique classique, on arrête la réaction à différents temps. Dans le but de s'affranchir de l'inconvenance de réaliser un grand nombre de prélèvements au cours du temps au sein d'un même milieu réactionnel, on choisit de déclencher plusieurs réactions avec différentes concentrations finales en enzyme et d'arrêter l'ensemble de ces réactions après un temps d'incubation défini. Un total de 4 prélèvements espacés de 20 min. entre chaque a néanmoins été effectué sur chaque milieu réactionnel pour chacune des dilutions d'enzyme réalisée afin d'augmenter le nombre de points cinétiques. La séparation de chaque milieu réactionnel montre un degré d'avancement donné de la réaction qui est fonction de l'importance de la dilution initiale de l'enzyme. L'arrêt de chaque réaction a été réalisé par chauffage à 70°C pendant 10 min.

Après séchage, les composés formés par Ttβgly et AgaD ont été élués en présence du mélange butanol/éthanol/eau (proportions 5/5/3) et ceux formés par TcTS par un mélange butanol/acide acétique/eau (proportions 2/5/3). Après migration et séchage, la plaque a été incubée 1 à 2 s dans un bain d'orcinol sulfurique puis chauffée à 100°C jusqu'à révélation des composés (*Figure 27*).

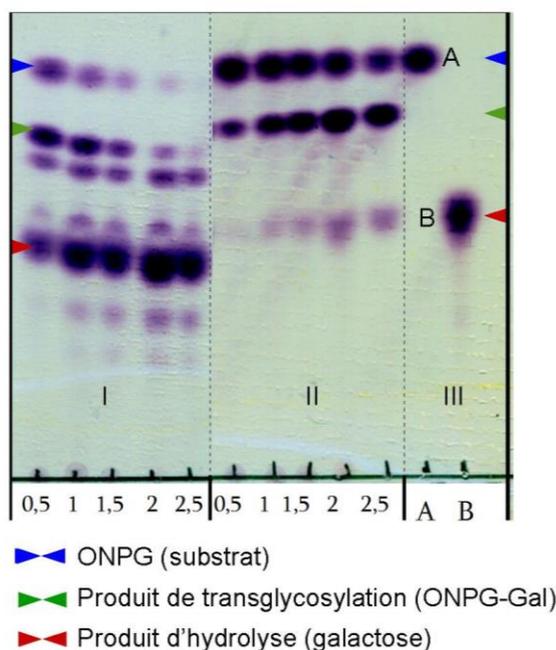


Figure 27. Exemple de chromatogramme obtenu par chromatographie sur couche mince.

Les plages I et II montrent le suivi cinétique de la Ttβgly sauvage et du mutant W120FC167I respectivement en présence d'ONPG à 25 mM en tampon PBS à 40°C. Les nombres indiqués sur la ligne de dépôt se réfèrent au temps de prélèvement du milieu réactionnel (en h). La plage III

montre la migration des composés témoins servant à identifier les espèces moléculaires après migration (A : ONPG, B : Galactose, 25 mM).

5.1.11 Suivi cinétique par spectrophotométrie d'absorption

Cette technique est utilisée en vue de déterminer les constantes cinétiques de la réaction enzymatique. L'activité enzymatique de Tt β gly a été évaluée pour différentes concentrations initiales en ONPG : 0,5 ; 1 ; 1,5 ; 2 ; 2,5 ; 3 ; 3,5 ; 4 ; 4,5 ; 5 ; 6 ; 7 ; 8 ; 9 ; 10 ; 12 ; 14 ; 16 ; 18 ; 20 ; 22 ; 24 ; 26 et 28 mM (dilué dans du PBS). Chaque milieu réactionnel est introduit au sein d'un puits d'une microplaque 96 puits et pré-incubé à 40°C. Chaque réaction est déclenchée par ajout de 0,66 μ g d'enzyme dans 200 μ L d'ONPG à une concentration donnée. La cinétique de libération de l'orthonitrophénol (ONP) est suivie par mesure de la densité optique à 405 nm. Les mesures sont réalisées toutes les 20 s pendant 15 min à 40°C. Pour chaque variant, les mesures sont répétées sur trois échantillons issus de trois purifications distinctes.

Le substrat utilisé pour le suivi cinétique de TcTS est le sialyl-PNP. La libération du PNP (groupe partant) lors de la première étape du mécanisme catalytique constitue en l'absence d'accepteur potentiel l'unique réaction (hydrolyse) catalysée en solution. La déglycosylation du sialyl-PNP peut être suivie au cours du temps par spectrophotométrie d'absorption à 450 nm. L'activité enzymatique de TcTS a été évaluée pour dix concentrations initiales en sialyl-PNP allant de 1 à 10 mM. La réaction a été déclenchée *via* 1,5 μ g d'enzyme purifiée dans du tampon phosphate 100 mM, pH 6 qsp 150 μ L final et suivie à 37°C. Une mesure de l'absorbance dans chacun des 10 puits a été effectuée toutes les 20 s pendant 15 min.

A partir de chaque courbe cinétique (obtenue par tracé des valeurs d'absorbance mesurées en fonction du temps pour un puits et donc une concentration en substrat donnée), les vitesses initiales sont déterminées en effectuant une régression linéaire sur au minimum 20 points du tracé. Seules les régressions présentant un coefficient de corrélation R^2 supérieur à 0,99 sont prises en compte. Le tracé des valeurs des vitesses initiales pour chaque concentration de départ en substrat nous permet de déterminer les constantes cinétiques de la réaction. Lorsque l'allure du tracé est hyperbolique, alors

la cinétique peut être modélisée par une régression hyperbolique selon l'équation de Michaëlis-Menten :

$$v = \frac{V_m S_0}{K_m + S_0} \text{ avec : } V_m = k_{cat} E_0$$

Les constantes cinétiques V_m (vitesse maximale de la réaction), K_m (constante de Michaëlis) et donc k_{cat}/K_m peuvent ainsi être déterminées à partir des paramètres de la régression.

Lorsque l'allure du tracé cinétique n'est pas hyperbolique, on ne peut alors accéder aux paramètres V_m et K_m . Il reste cependant possible d'estimer la performance catalytique de l'enzyme (soit le rapport k_{cat}/K_m) en réalisant une régression linéaire sur les premiers points de la cinétique (aux basses concentrations en substrat) pour lesquels une relation linéaire entre la vitesse initiale et la concentration initiale en substrat est observé. Les régressions ont été effectuées en utilisant le logiciel Excel 2003.

5.1.12 Suivi cinétique par électrophorèse capillaire

i. Suivi cinétique de Tt β gly

La quantification des rendements de transglycosylation a été réalisée par électrophorèse capillaire. La réaction enzymatique de Tt β gly est déclenchée par l'ajout de 0,66 μ g d'enzyme dans 200 μ L de solution d'ONPG à 5 mM préparée dans du PBS et pré-incubée à 40°C. Le suivi de la cinétique de formation des produits de transglycosylation s'effectue à 40°C pendant environ 10 h. L'utilisation d'une concentration faible en ONPG a pour effet de défavoriser la transglycosylation en minimisant la concentration en accepteur dans le milieu réactionnel. Ceci permettra ainsi de mieux discriminer les mutants présentant une activité de transglycosylation augmentée par rapport à l'enzyme sauvage. Un standard interne (imidazole à 10 mM final) est ajouté au milieu réactionnel afin de corriger les variations de volumes prélevés par l'automate d'une séparation à l'autre.

Environ 5 nL de mélange réactionnel sont ensuite prélevés toutes les 20 min et soumis à une séparation électrophorétique de l'anode vers la cathode le long d'un capillaire de silice de 47 cm, thermostaté à 18°C. Une solution de tétra-borate de sodium 50 mM, pH 10,4 est utilisée comme tampon de migration. La séparation se déroule à 20 kV et la détection des composés présents en solution est réalisée par mesure de l'absorbance à 214 nm. Porteurs d'un groupement ONP absorbant à cette longueur d'onde, seuls le substrat résiduel et le produit de transglycosylation formé pourront être détectés.

ii. Suivi cinétique de TcTS

Conformément à de nombreux travaux décrits dans la littérature, nous avons préféré tester les mutants de TcTS sur des substrats naturels par opposition à Ttβgly qui n'a été testée que sur un substrat synthétique (ONPG). Le 3'SL ainsi que le LNT ont été donc utilisés comme substrats donneur et accepteur respectivement. Contrairement au sialyl-PNP (substrat synthétique) et au LNT, le 3'SL n'absorbe pas dans le visible et n'est donc pas détectable comme tel par spectrophotométrie d'absorption. Afin de pouvoir suivre l'évolution de la concentration de ce substrat au cours du temps, une réaction de dérivation de l'ensemble des composés du milieu réactionnel avec un chromophore (le 4-aminobenzonitrile) a été réalisée sur la nuit à 40°C sur l'ensemble des échantillons à séparer (prélèvements et témoins).

La réaction a été déclenchée par ajout de 1,5 µg d'enzyme purifiée dans 125 µL de tampon phosphate 100 mM, pH 6 pré-incubé à 37°C en présence de 7,5 µL de sucre donneur de groupement sialyl (3'SL) et de 7,5 µL de sucre accepteur (LNT) à 200 mM chaque. Afin de définir les points cinétiques à analyser, 10 µL de ce milieu réactionnel ont été prélevés toutes les 30-40 min (soit ~ 14 prélèvements en tout) et la réaction a été arrêtée par incubation du prélèvement au thermocycler à 75°C et ajout du réactif de dérivation. Afin de pouvoir quantifier les différentes espèces en solution à chaque temps d'arrêt, la séparation de chaque prélèvement (après centrifugation) a été réalisée dans un deuxième temps à température ambiante. La séparation a été effectuée à 17 kV en tampon Borax 50 mM, pH 10,4 et la détection a été réalisée à 214 nm.

iii. Suivi cinétique d'AgaD

La réaction enzymatique a été déclenchée dans 10 μL de tampon de réaction (MOPS 500 mM, NaCl 3 M, pH 7,5) en présence de 10 μL d'agarodécaose à 100 mg/mL (qsp 100 μL d'eau distillée) par ajout de 4 à 80 μL d'enzyme sauvage purifiée (soit 0,21 μg) ou de 0,44 μg pour le mutant Q342H, de 1,9 μg pour le mutant D341L/S351A, de 4,2 μg pour le mutant D341L/S351F, de 2,4 μg pour le mutant D341L/Q342H/S351A et de 0,9 μg pour le mutant D341L/Q342H/S351F. Le milieu réactionnel a ensuite été incubé à 25°C sous agitation. Un aliquote de 4 μL de ce milieu a ensuite été prélevé toutes les heures et mélangé à 4 μL d'une solution de maltotriose à 5 mg/mL servant de standard interne. L'ajout à ce mélange de 30 μL d'une solution d'ANTS à 0,15 M conditionnée dans 15 % d'acide acétique (v/v) et de 30 μL de NaBH_3CN à 1 M en DMSO va permettre d'effectuer la réaction de dérivation indispensable à la détection des composés du milieu réactionnel (substrat et produits de réaction). Pour accélérer la réaction de dérivation, les mélanges finaux sont incubés à 55°C pendant 7 h avant d'être séparés par électrophorèse. Contrairement à Tt β gly et TcTS, la séparation a été réalisée au sein d'un capillaire de silice de 54 cm de long et de 120 μm de diamètre. Les échantillons ont été introduits dans le capillaire par injection hydrodynamique à 40 mbar pendant 6 s. Une solution d'acide phosphorique à 50 mM, ajustée à pH 2,5 par ajout de triéthylamine, a été utilisée comme tampon de migration. La séparation a été effectuée de la cathode vers l'anode à -25 kV, 30°C et la détection a été réalisée à 214 nm.

iv. Généralités et calcul du rendement de transglycosylation

Pour tous les suivis cinétiques réalisés par cette technique, un nettoyage sous pression (injection d'eau pendant 3 min, puis de NaOH à 0,1 M pendant 2 min, puis à nouveau d'eau pendant 2 min) suivi d'un reconditionnement de la phase stationnaire du capillaire dans le tampon de migration est réalisé avant chaque passage d'échantillons. Un électrophérogramme est ainsi obtenu après chaque séparation. Chaque espèce moléculaire présente en solution (substrat résiduel, produit de transglycosylation, standard interne) est identifiée par un pic centré sur son temps de rétention au sein du capillaire. L'intégration de chaque pic (réalisée par le logiciel Beckman Coulter's 32 Karat™ Software 8.0) permet de déterminer son aire relative (*Figure 28*).

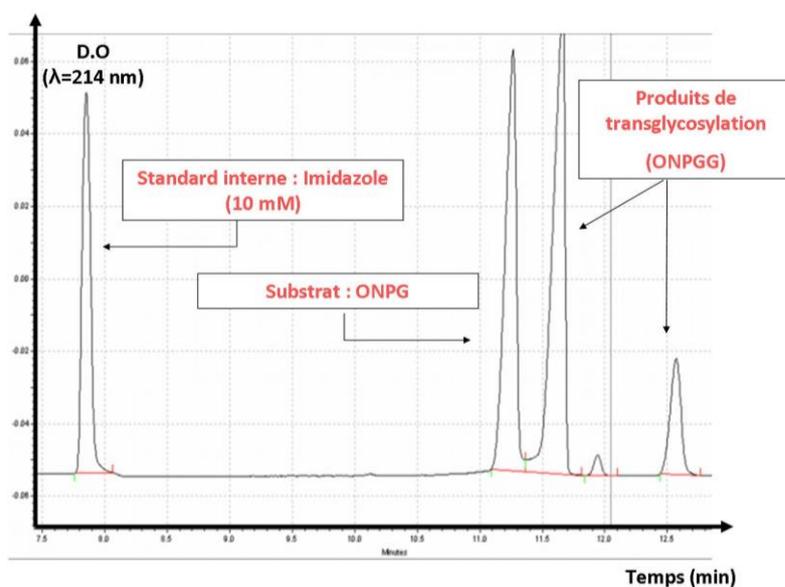


Figure 28. Profil type d'électrophorèse capillaire dans le cas d'un suivi cinétique de la Ttβgly sauvage à titre d'exemple.

Les deux pics se référant aux produits de transglycosylation indiquent la formation de deux régioisomères distincts.

L'aire des pics correspondant au substrat résiduel ainsi qu'au produit de transglycosylation est ensuite normalisée par celle du standard interne. Ces aires normalisées sont ensuite utilisées pour déterminer la concentration de chaque espèce en solution. La connaissance de la concentration initiale en substrat (5 mM) et de l'aire normalisée du pic du substrat au temps initial nous permet de déterminer l'évolution de la concentration du substrat sur l'ensemble de la cinétique. Le calcul de la concentration du produit de transglycosylation C_{iX}^{trans} à partir de son aire relative A_{iX}^{trans} est réalisé selon la relation suivante où C_{i0}^{ONPG} et A_{i0}^{ONPG} correspondent respectivement à la concentration et à l'aire normalisée du substrat au début de la cinétique :

$$C_{iX}^{trans} = \frac{A_{iX}^{trans} C_{i0}^{ONPG}}{A_{i0}^{ONPG}}$$

Il est alors possible de déduire la concentration C_{iX}^{Gal} du produit d'hydrolyse à chaque instant de la cinétique par la relation :

$$C_{iX}^{Gal} = C_{iX}^{ONPG} - C_{iX}^{trans}$$

Sachant que deux molécules de substrat (ONPG) sont utilisées pour former une molécule de produit de transglycosylation (ONPG-Gal), le rendement maximal de transglycosylation peut être obtenu *via* la relation :

$$R_{\max}^{\text{trans}} = \frac{2 \times C_{t_{\max}}^{\text{trans}}}{(C_{t_0}^{\text{ONPG}} - C_{t_{\max}}^{\text{ONPG}})}$$

Où $C_{t_{\max}}^{\text{trans}}$ est la concentration maximale en produit de transglycosylation formé au cours de la cinétique (à un instant noté t_{\max}), $C_{t_0}^{\text{ONPG}}$ la concentration initiale en substrat, et $C_{t_{\max}}^{\text{ONPG}}$ la concentration en substrat résiduel à l'instant t_{\max} .

5.1.13 Détection d'oligosaccharides par analyse HPAEC-PAD

Afin de détecter les oligosaccharides produits par AgaD et ses mutants, les milieux réactionnels précédemment préparés (section 2.2.12) ont été séparés par chromatographie ionique HPAEC-PAD. Le système a été préalablement équilibré avec une solution de NaOH à 150 mM. Les échantillons introduits en colonne ont été élués en 30 min à un débit de 0,5 mL/min par ajout d'une solution de NaOH à 150 mM et d'un gradient linéaire d'acétate de sodium allant de 0 à 0,5 M. Les oligosaccharides ont été détectés *via* un détecteur électrochimique (électrode en or) en utilisant les paramètres suivants : 0,00 s (+0,1 V) ; 0,20 s (+0,1 V; Start data acquisition) ; 0,40 s (+0,1 V; Stop data acquisition) ; 0,41 s (-2,0 V) ; 0,42 s (-2,0 V) ; 0,43 s (+0,6 V) ; 0,44 s (-0,1 V) ; 0,50 s (-0,1 V).

5.2 Méthodes de bioinformatique et de modélisation moléculaire

5.2.1 Alignement multiple de séquences de GH1

L'alignement des séquences protéiques des GH bactériennes et eucaryotes de la famille 1 a été réalisé en vue de déterminer les mutations à réaliser sur les positions bordant le canal principal mis en évidence chez Ttβgly. Il a été réalisé grâce au logiciel MUSCLE¹⁹⁰. Un alignement de séquences de GH1 bactériennes (hydrolases strictes) a préalablement été construit par Yves Henri Sanejouand (UFIP, Equipe de conception de protéines *in-silico*). Cet alignement regroupe l'ensemble des séquences de GH1 bactériennes répertoriées dans UniprotKB qui présentent entre 20 et 90% d'identité de

séquence avec celle de la Tt β gly. Afin de déterminer le type de mutations à réaliser sur Tt β gly, un total de 21 séquences de GH1 eucaryotes (considérées comme transglycosylases) ont été ajoutées à l'alignement existant. Quatre séquences, incluant entre autre quatre séquences de GH1 de plantes (aa5gt⁶⁷, aa7gt⁶⁷, os9bglu31⁶⁸, et SFR2⁶⁶) et une séquence de GH1 d'insecte (PDB : 3VIK), ont été caractérisées comme transglycosylases strictes.

La séquence d'une β -glycosidase BGlu1¹⁹¹ de riz (PDB : 2RGM) connue pour son activité de transglycosylation a été utilisée comme sonde afin de consolider ce jeu de données. Un psi-blast¹⁹² de cette séquence contre l'ensemble des séquences protéiques de plantes répertoriées dans UniprotKB a tout d'abord été effectué dans le but de rechercher d'autres séquences de transglycosylases putatives. Un alignement des l'ensemble des séquences identifiées par blast en présence des quatre séquences de transglycosylases connues (aa5gt, aa7gt, os9bglu31, et SFR2) a ensuite été réalisé *via* MUSCLE. Un arbre phylogénétique, construit *via* ClustalW2^{193,194} à partir de l'alignement précédent, a finalement permis d'identifier 15 séquences évolutivement proches des transglycosylases caractérisées. Un alignement multiple de l'ensemble des séquences de GH1 eucaryotes identifiées a ensuite été construit. Cet alignement multiple a finalement été aligné à celui des séquences de GH1 bactériennes (méthode d'alignement profil à profil) afin de construire l'alignement final.

5.2.2 Analyse des séquences de GH33 trypanosomales

i. Consolidation d'un jeu de séquences à analyser

Une collecte de l'ensemble des séquences de domaines catalytique de (trans)sialidases de trypanosomes complètes répertoriées dans les banques de séquences UniprotKB et InterPro a tout d'abord été réalisée. Une comparaison de toutes les paires de séquences possibles à ensuite été effectué afin d'éliminer les séquences redondantes. Une distribution statistique de la taille des séquences restantes au sein du jeu initial a ensuite été établie afin de déterminer la longueur de séquence la plus couramment observée. Cette dernière étape a été réalisée dans le but de définir la longueur de

séquence la plus probable pour le domaine catalytique des GH33 trypanosomales et ainsi d'éliminer toute séquence présentant une taille inférieure à ce seuil.

ii. Recherche de séquences caractérisées dans la littérature

Le jeu de séquences résultant a également été supplémenté par l'ensemble des séquences de domaines catalytiques des trans-sialidases des trypanosomes *T. brucei*^{195, 196}, *T. congolense*^{197, 198, 199, 200, 201}, et *T. vivax*²⁰² caractérisées biochimiquement à ce jour et répertoriées dans la littérature (*Tableau 2*). Ces séquences ont donc été rajoutées manuellement dans le jeu de données constitué précédemment car elles n'étaient pas répertoriées dans UniprotKB ou InterPro et n'ont donc pas pu être collectées lors du criblage initial.

Espèces	Identifiants	Références
<i>T. cruzi</i>	Q26850 Q26964 Q26966 P18269 Q26969 Q9BHJ5	CAZy database
<i>T. rangeli</i>	O44049 Q08672 Q27064	CAZy database
<i>T. brucei sp. brucei</i>	Q57YT7	Montagna et coll. ^{195, 196}
<i>T. brucei</i>	A0MH09	
<i>T. vivax (Y486)</i>	F9WU40	Guegan et coll. ²⁰²
<i>T. congolense</i>	TcoTS1a TcoTS1b TcoTS1c TcoTS1d	Tiralongo et coll. ^{197, 198} Joachim Schmitt et coll. ¹⁹⁹ Koliwer-Brandl et coll. ²⁰⁰ Coustou et coll. ²⁰¹

Tableau 2. Séquences de GH33 de trypanosomes collectées dans la base de données CAZy et issues de la littérature.

iii. Alignement multiple des séquences sélectionnées

A l'exception des séquences portant les identifiants F9WU40 (*T. vivax*) et P18269 (*T. cruzi*), l'ensemble des autres séquences collectées dans la littérature (et biochimiquement caractérisées) a été aligné afin de constituer un alignement de référence (alignement profil de référence). Les séquences collectées lors du premier criblage ont été groupées par espèces, puis alignées entre elles au sein de chaque groupe d'espèces. L'intérêt de construire ces alignements intermédiaires est de maximiser la robustesse de l'alignement final en alignant d'abord entre elles l'ensemble des séquences apparentées (regroupées ici sous la même espèce). Les alignements multiples ainsi constitués ont été finalement alignés un à un au profil de référence (méthode d'alignement « profil à profil ») pour construire l'alignement multiple final. A ce stade, un filtre a été appliqué en vue d'éliminer les séquences trop divergentes au sein de chaque alignement intermédiaire. Ces séquences regroupent l'ensemble des séquences ne possédant pas les résidus retrouvés dans la machinerie catalytique de TrSA ou TcTS ainsi que celles présentant des insertions trop fréquentes au sein des régions correspondant aux éléments de structure secondaire retrouvées chez TrSA et TcTS.

iv. Construction d'un arbre phylogénétique

A partir de l'alignement multiple final (MSA), un arbre phylogénétique a été construit afin de caractériser les relations évolutives entre les séquences de (trans)sialidases sélectionnées. Avant de construire l'arbre, une séquence de sialidase non-trypanosomale de *Trichomonas vaginalis* (identifiant Uniprot : A2EIP2) a préalablement été alignée à l'ensemble des séquences du MSA (alignement profil à profil). L'inclusion de cette séquence représentant un « extra-groupe » a été utilisée pour renforcer la robustesse de l'arbre. Le choix de la séquence de l'extra-groupe a été fait en tenant compte du critère de sélection précédemment appliqué sur le jeu de séquence initial (limite inférieure de taille à 300 résidus) tout en considérant le choix d'une séquence suffisamment proche du groupe des Trypanosomes sur le plan phylogénétique.

5.2.3 Simulation de dynamique moléculaire de la Ttβgly

La structure cristallographique de la Ttβgly sauvage (PDB : 1UG6) a été téléchargée depuis la PDB (<http://www.rcsb.org/pdb/home/home.do>). L'état de protonation des résidus titrables a été prédit par le programme *pdb2gmx* du logiciel GROMACS^{203,204}. En accord avec le modèle mécanistique des GH à rétention de configuration, le carboxylate du résidu acide-base (E164) a été maintenu à l'état déprotoné. La partie glucosyle liée au nucléophile (E338) a été paramétrisé pour le champ de force Glycam06²⁰⁵ via l'interface web Glycam (<http://glycam.org/>). Le champ de force AMBER99SB-ildn²⁰⁶ a été utilisé pour la simulation de l'ensemble du système en présence de conditions périodiques. Le modèle d'eau rigide TIP3P¹⁷⁶ a été choisi. L'ensemble des calculs a été réalisé avec la version 4.6 du logiciel GROMACS en utilisant les ressources du supercalculateur Occigen du CINES. Pour les simulations des mutants, les substitutions ont été introduites via le programme *fixbb* de la suite Rosetta Design²⁰⁷ (version 2015.38158). Une optimisation du positionnement (« repacking ») des chaînes latérales des résidus mutés a également systématiquement été effectuée.

Une boîte cubique de dimension 8,88 nm³ a été construite autour de la protéine afin d'y accueillir le solvant. La neutralisation de sa charge a été réalisée par l'ajout de contre-ions (NaCl à 115 mM). Une minimisation d'énergie a ensuite été effectuée en trois étapes (algorithme du *Steepest Descent*) afin d'optimiser le positionnement et l'orientation des atomes les uns par rapport aux autres et d'éliminer les contraintes stériques liés à une trop grande proximité de certains atomes dans l'espace. Une première minimisation du solvant a été réalisée en 1000 étapes en maintenant la protéine rigide par le biais de contraintes harmoniques. La suppression des contraintes sur les chaînes latérales a ensuite permis leur minimisation conjointement au solvant en 1000 étapes. Les contraintes sur le squelette peptidique ont finalement été relâchées afin de permettre la minimisation de l'intégralité du système jusqu'à convergence de l'algorithme (seuil de tolérance fixé à 10 kJ/mol). La température du système a ensuite été équilibrée à 350 K (Ttβgly étant issue d'un organisme thermophile) pendant 500 ps. Cette simulation a été suivie d'une équilibration de la pression du système à température constante. Durant toutes les étapes d'équilibration du solvant, des contraintes harmoniques de 100 kJ/mol/nm² ont été imposées sur l'ensemble de la

protéine. La production de la trajectoire a finalement été réalisée pendant 500 ns en l'absence de toutes autres contraintes que celles imposées sur la partie glucosyle liée au résidu nucléophile.

Le maintien d'une température constante est assuré par le recalibrage des vitesses (Velocity rescaling²⁰⁸) avec une constante de couplage de 0,1 ps. La pression du système (1.0 bar) est maintenue constante par le barostat de Parrinello-Rahman²⁰⁹ avec une constante de couplage de 1,0 ps. L'introduction de contraintes holonomiques sur les liaisons covalentes *via* l'algorithme LINCS²¹⁰ a permis d'utiliser un pas d'intégration de 2 fs. Pour chaque atome, la distance de troncature (*cutoff*) de la collecte de ses partenaires d'interactions ainsi que des interactions de type Van der Waals et coulombiennes est de 12 Å. La méthode PME¹⁸² (*Particle Mesh Ewald*) a été utilisée pour le calcul des interactions électrostatiques à longue distance. Les coordonnées atomiques du système ont été enregistrées toutes les 100 ps.

5.2.4 Simulation de dynamique moléculaire de l'AgaD

La structure cristallographique de la forme libre de l'AgaD sauvage (PDB : 4ASM¹⁸⁶) a été téléchargée depuis la PDB. La préparation du système a été réalisée dans les mêmes conditions que celles décrites ci-dessus à quelques exceptions près : seule l'apoprotéine a été simulée, sa solvatation a été effectuée au sein d'une boîte cubique de 8,66 nm³ et la température de simulation était de 300 K (organisme mésophile).

5.2.5 Simulations de dynamique moléculaire de TcTS et TrSA

Les structures cristallographiques des formes sauvages de TcTS (PDB : 1S0I²¹¹) et de TrSA (PDB : 1N1S²¹²) ont été téléchargées depuis la PDB. Les structures cristallographiques de TcTS et de TrSA présentent toutes deux une boucle non-résolue allant du résidu A399 au G409 chez TcTS et du résidu A406 au G413 chez TrSA (numérotation du fichier PDB). Cette boucle a donc été reconstruite *via* le logiciel MODELLER v9.11²¹³ et sa géométrie a été optimisée²¹⁴ (l'algorithme « refine.slow » a été employé). La séquence de la structure cristallographique de TcTS présente huit résidus mutés incluant le résidu acide-base catalytique et sept autres localisés en surface

de la protéine. Le logiciel MODELLER a donc été employé afin de les remplacer par les résidus originaux présents dans la séquence de la forme sauvage de l'enzyme. Le sialyl-lactose étant absent de la structure cristallographique de TrSA, celui présent dans la structure de TcTS a donc été amarré dans sa conformation cristallographique originale dans le site catalytique de TrSA par analogie.

Le complexe de Michaëlis (ou complexe enzyme-substrat) de la protéine avec son substrat naturel (le sialyl-lactose) a été simulé et sa solvation a été effectuée au sein d'une boîte octaédrique. La distance entre l'atome de la protéine le plus éloigné de son centre de gravité et le bord de la boîte a été fixé à 1 nm. La température de simulation était de 300 K. Conformément au mécanisme catalytique des GH à rétention de configuration, le résidu d'acide aspartique jouant le rôle d'acide-base dans chacune des enzymes a été maintenu à l'état protoné.

Le sialyl-lactose a été paramétré dans le champ de force Glycam06²⁰⁵ *via* le module de construction « Carbohydrate builder » du serveur web : <http://glycam.org/> en spécifiant la séquence suivante : a-D-Neu5Ac-(2-3)-b-D-Galp-(1-4)-b-D-Glcp-OH (ou a et b indiquent l'anomérie α et β). Afin de pouvoir être utilisé par le logiciel GROMACS, le fichier spécifiant la topologie du sialyl-lactose initialement au format AMBER (.prm) a ensuite été converti au format .top *via* le script acpype.py²¹⁵. La section suivante, spécifiant les paramètres de calcul des interactions non-liées, a été introduite au sein du fichier topologie généré par *pdb2gmx* :

```
[ atomtypes ]
;name bond_type mass charge ptype sigma epsilon Amb
HO HO 0.00000 0.00000 A 0.00000e+00 0.00000e+00;0.00 0.0000
OH OH 0.00000 0.00000 A 3.06647e-01 8.80314e-01;1.72 0.2104
CG CG 0.00000 0.00000 A 3.39967e-01 4.57730e-01;1.91 0.1094
H2 H2 0.00000 0.00000 A 2.29317e-01 6.56888e-02;1.29 0.0157
OS OS 0.00000 0.00000 A 3.00001e-01 7.11280e-01;1.68 0.1700
H1 H1 0.00000 0.00000 A 2.47135e-01 6.56888e-02;1.39 0.0157
CY CY 0.00000 0.00000 A 3.39967e-01 4.57730e-01;1.91 0.1094
C C 0.00000 0.00000 A 3.39967e-01 3.59824e-01;1.91 0.0860
O2 O2 0.00000 0.00000 A 2.95992e-01 8.78640e-01;1.66 0.2100
HC HC 0.00000 0.00000 A 2.64953e-01 6.56888e-02;1.49 0.0157
N N 0.00000 0.00000 A 3.25000e-01 7.11280e-01;1.82 0.1700
H H 0.00000 0.00000 A 1.06908e-01 6.56888e-02;0.60 0.0157
O O 0.00000 0.00000 A 2.95992e-01 8.78640e-01;1.66 0.2100
OY OY 0.00000 0.00000 A 3.00001e-01 7.11280e-01;1.68 0.1700
```

Le reste de la préparation des deux systèmes a été réalisée dans les mêmes conditions que ceux décrites dans la simulation d'AgaD.

5.2.6 Paramétrage de la liaison glycosyl-enzyme de Ttβgly

La seule structure d'intermédiaire covalent répertoriée dans la PDB parmi l'ensemble des GH1 homologues à 1UG6 est celle de la β-glycosidase de *Thermotoga maritima* (Tmaβgly, PDB : 1OIN²¹⁶) qui possède un glucosyle lié à son glutamate nucléophile. La connaissance du positionnement de l'intermédiaire covalent étant cruciale, et en dépit du fait que les tests d'activité de la Ttβgly et ses mutants aient été réalisés avec l'ONPG (et donc le galactose comme substrat donneur), le glucosyl-glutamate a été choisi comme intermédiaire covalent. Ce résidu n'étant pas décrit dans le champ de force choisi lors de ce travail, il était nécessaire de le paramétrer. Si les paramètres de nombreux sucres liés aux aminoacides de type alcools et amides peuvent en effet être obtenus à partir du champ de force Glycam06, celui-ci n'est cependant pas paramétré pour simuler des sucres liés à des fonctions carboxyles.

Essentiels au calcul des potentiels d'interactions, ces paramètres regroupent les charges partielles associées aux atomes, les longueurs de liaisons et valeurs d'angles à l'équilibre, ainsi que les constantes de force associées aux interactions. Ne disposant cependant pas de l'expertise nécessaire pour le calcul *ab-initio* des constantes de force, seules les charges partielles associées à chacun des atomes du glucosyl-glutamate ont été calculées *ab-initio* par Philippe Arnaud (UFIP, Equipe de conception de protéines *in-silico*) Les valeurs des longueurs de liaisons et d'angles à l'équilibre ainsi que les constantes de force associées à la partie glucosyle ne pouvant être calculées, des valeurs pour ces paramètres ont donc été choisies par analogie afin de pouvoir intégrer ce nouveau résidu dans le champ de force AMBER. L'utilisation des coordonnées atomiques de la conformation cristallographique du glucosyl-glutamate chez Tmaβgly permet de résoudre le problème du choix des valeurs de longueurs de liaison et d'angles à l'équilibre.

En ce qui concerne les valeurs des constantes de force associées, les paramètres de l'analogie glucosyl-thréonine obtenus dans le champ de force Glycam06 peuvent être transposés au glucosyl-glutamate pour modéliser la liaison covalente entre la protéine et le sucre. Ces paramètres n'étant pas corrects, la simulation de la dynamique du glucosyl-

5.2.7 Modification du champ de force

Le champ de force AMBER comporte cinq fichiers permettant de décrire l'ensemble des paramètres et des interactions affectant les atomes d'un système :

- ⇒ aminoacids.rtp
- ⇒ atomtypes.atp
- ⇒ ffbonded.itp
- ⇒ ffnonbonded.itp
- ⇒ residuetypes.dat

Ces fichiers doivent être modifiés en vue d'intégrer les paramètres du glucosyl-glutamate au champ de force AMBER99SB-ildn. Les modifications réalisées sur ces fichiers sont détaillées dans la partie Annexe.

5.2.8 Construction du glucosyl-enzyme de Tt β gly

La construction de l'intermédiaire covalent de Tt β gly a été réalisée *via* l'interface graphique du logiciel PyMOL (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.) et fut faite en substituant le glutamate nucléophile original par un glucosyl-glutamate pris dans la conformation cristalline de son homologue structural, Tma β gly (1OIN). La présence d'un atome de fluor en position 2 du cycle osidique de l'intermédiaire covalent de Tma β gly n'a pas permis de l'intégrer directement au sein de Tt β gly. Le 2-fluoro-glucosyle de Tma β gly a donc été préalablement remplacé par la partie glucosyle de la structure optimisée du glucosyl-glutamate. Les atomes du nouveau résidu ont finalement été renumérotés en accord avec la numérotation spécifiée dans le fichier topologie dérivé de Glycam06.

5.2.9 Analyse des trajectoires de simulations

5.2.9.1 Calcul du RMSD

Le RMSD (Root Mean Square Deviation, ou écart quadratique moyen) est un paramètre permettant d'estimer la déviation des coordonnées atomiques d'une conformation moléculaire donnée par rapport à une conformation de référence. Dans le cadre des simulations réalisées, ce paramètre permet d'estimer la stabilité conformationnelle de la protéine au cours du temps. En pratique, la conformation de la protéine prise à chaque cliché de la trajectoire (C_t) est superposée à sa conformation initiale (C_0) en début de simulation. Les superpositions sont réalisées sur le squelette peptidique des deux conformères. Après superposition, les écarts quadratiques entre toutes les paires d'atomes identiques entre les deux conformations C_t et C_0 sont calculés, puis moyennés.

$$RMSD(C_0, C_1) = \sqrt{\frac{1}{n} \sum_{i=1}^n |r_{i,1} - r_{i,0}|^2}$$

Où $r_{i,0}$ et $r_{i,1}$ représentent respectivement le vecteur position d'un atome i au sein de la conformation prise au cliché initial (0) et au cliché suivant. n correspond au nombre total d'atomes de la protéine. Le calcul du RMSD sur l'ensemble de la trajectoire a été effectué *via* le module *g_rms* du logiciel GROMACS.

5.2.9.2 Calcul du RMSF

Le RMSF (Root Mean Square Fluctuation, ou fluctuation quadratique moyenne) a été utilisé pour évaluer la flexibilité moyenne de chaque résidu au sein de Ttβgly et d'AgaD. Le module *g_rmsf* du logiciel GROMACS a été utilisé. Pour chaque conformation d'un résidu donné, l'écart de ses coordonnées atomiques par rapport à sa conformation moyennée sur l'ensemble de la trajectoire est calculé, puis moyenné. Le calcul du RMSF de chaque résidu au sein d'une protéine permet ainsi de localiser les régions les plus flexibles au cours de la dynamique.

$$RMSF(i) = \sqrt{\frac{1}{N} \sum_{j=1}^N |r_{i,j} - \langle r_i \rangle_N|^2}$$

Où r_i représente le vecteur position de l'atome i d'un résidu donné pris dans une conformation j . $\langle r_i \rangle_N$ correspond à la position moyenne de l'atome i et N au nombre total de clichés au sein de la trajectoire.

Le RMSF peut être relié au facteur B de température par la relation :

$$B = \frac{8\pi^2}{3} (\text{RMSF})^2$$

5.2.9.3 Liaisons hydrogène

La liaison hydrogène (H) est une interaction électrostatique qui a lieu lorsqu'un atome d'hydrogène covalentement lié à un atome (dit « donneur ») interagit à distance avec un atome fortement électronégatif ou polarisé négativement (dit « accepteur »). La capacité de deux atomes à former une liaison H dépend d'au moins 2 critères géométriques définis : la distance r_{max} maximale entre atomes donneur et accepteur ainsi que l'angle θ_{max} formé entre l'hydrogène et l'atome accepteur.

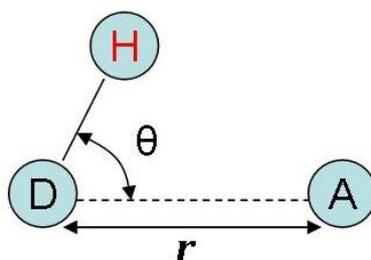


Figure 29. Représentation schématique d'une liaison hydrogène.

Les atomes D et A correspondent respectivement aux atomes donneur et accepteur de liaison H.

Le module g_hbond du logiciel GROMACS a été utilisé pour évaluer le nombre de liaisons hydrogène établies entre un atome donneur et un atome accepteur au sein des protéines simulées. En utilisant les paramètres par défaut du programme, r_{max} et θ_{max} doivent être respectivement inférieurs ou égaux à 3,5 Å et 30° pour qu'une liaison H puisse être établie (Figure 29).

5.2.9.4 Flexibilité conformationnelle des chaînes latérales

Le calcul des angles dièdres propres (notés χ) des chaînes latérales de résidus spécifiques (*Figure 30*) a été employé chez Tt β gly et AgaD pour caractériser leur flexibilité conformationnelle au cours du temps. Un script écrit en python a été développé à cet effet.

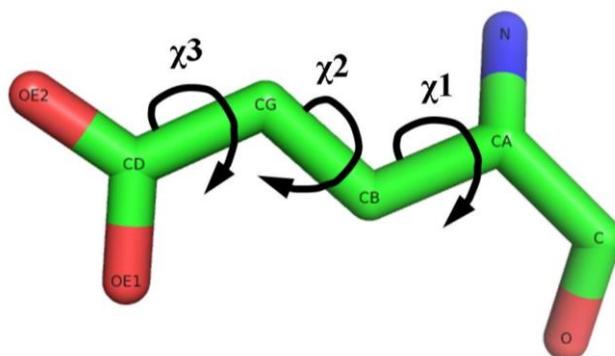


Figure 30. Angles de torsion (χ^1, χ^2, χ^3) de la chaîne latérale d'un résidu aspartate.

5.2.9.5 Reconstruction des canaux d'eau potentiels

L'identification des canaux d'eau potentiels sur l'ensemble des trajectoires générées est basée sur le suivi du déplacement des molécules d'eau internes selon un protocole original. L'isolement de l'eau interne aux structures protéiques sur l'intégralité de la trajectoire a été réalisé par le biais du programme PDBNETVOIS précédemment développé par Yves Henri Sanejouand. L'eau interne (ou enfouie) se définit comme l'ensemble des molécules d'eau qui n'établissent pas de liaisons hydrogène avec le solvant extérieur (ou milieu). A partir de l'ensemble des molécules d'eau internes isolées sur chaque cliché instantané de trajectoire (enregistré toutes les 100 ps) pour une simulation donnée, une recherche des sous-réseaux de molécules d'eau internes a été entreprise (*Figure 31*). Un sous-réseau d'eau interne a été défini comme un ensemble de molécules d'eau internes interconnectées, dont les oxygènes sont distants d'au plus 3,2 Å (longueur moyenne d'une liaison hydrogène) les uns des autres.

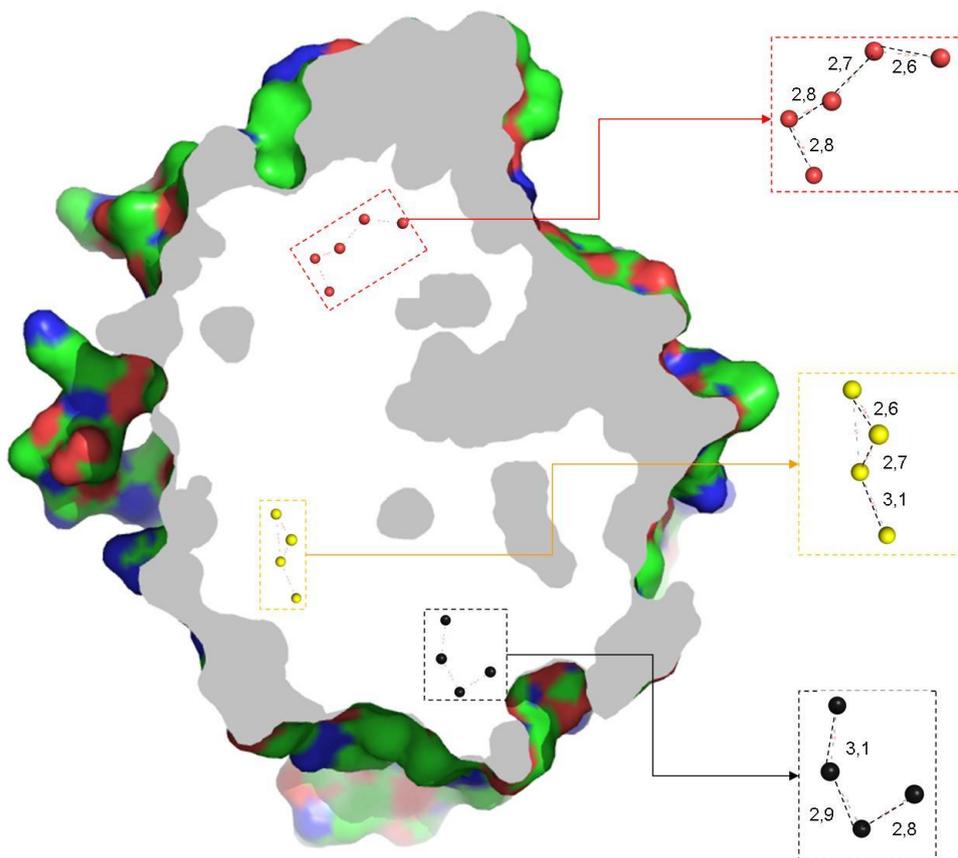


Figure 31. Identification des sous-réseaux d'eau interne sur un cliché de trajectoire

Dans cet exemple, trois sous-réseaux d'eau interne (rouge, jaune, et noir) sont identifiés sur un cliché de trajectoire. Les sphères représentent les oxygènes des molécules d'eau et les distances inter-oxygènes sont illustrées par des pointillés.

Identifiés sur chaque cliché de trajectoire, ces sous-réseaux servent de point de départ pour l'identification de canaux d'eau potentiels. Sur l'ensemble d'une trajectoire donnée, les clichés de trajectoire contigus dans le temps de simulation (i.e. deux clichés de trajectoire successifs à 100 ps d'intervalle) ont ensuite été superposés deux à deux en vue d'identifier les transferts de molécules d'eau entre sous-réseaux proches (i.e. distants d'au plus $3,2 \text{ \AA}$) d'un cliché à l'autre. La mise en évidence d'un transfert d'au moins une molécule d'eau entre deux sous-réseaux d'eau interne spatialement proches d'un cliché à l'autre a ainsi permis de reconstituer un réseau de taille plus importante (Figure 32). La répétition de cette procédure sur l'ensemble de la trajectoire aboutit à la reconstruction d'un réseau dynamique pouvant s'apparenter à un canal d'eau interne.

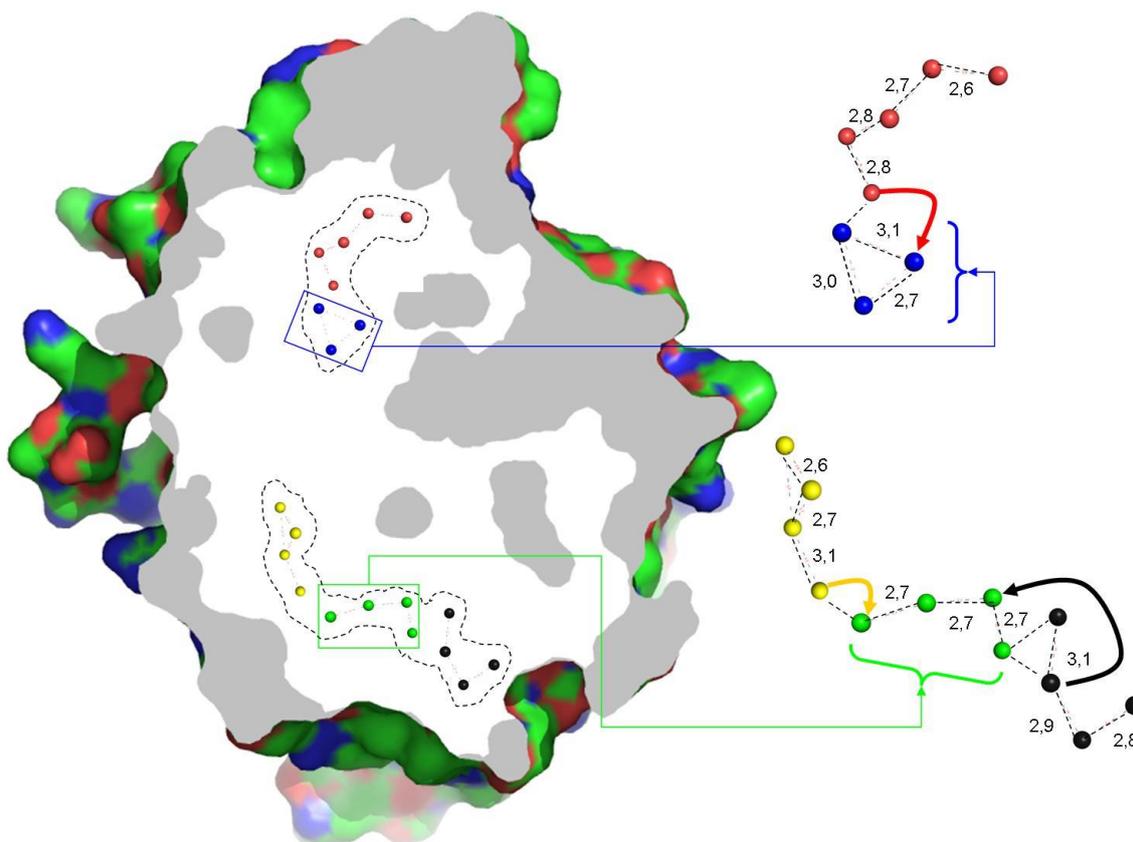


Figure 32. Reconstitution des réseaux d'eau internes

Les sous-réseaux (ou clusters) identifiés sur le cliché précédent (t) sont colorés en jaune, noir, et rouge. Au cliché de trajectoire suivant ($t+100$ ps), deux nouveaux clusters d'eau (bleu et vert) spatialement proches des sous-réseaux isolés à l'instant t sont identifiés. Les flèches rouge, noir, et jaune indiquent respectivement le transfert d'une molécule d'eau de chaque cluster identifié précédemment vers un nouveau cluster formé à $t+100$ ps. Le transfert de molécules d'eau entre des clusters spatialement proches d'un cliché à l'autre permet ainsi de reconstituer un canal d'eau potentiel. Dans cet exemple, deux canaux d'eau potentiels sont formés.

5.2.9.6 Dynamique de l'eau au sein d'un canal

Des scripts ont été développés pour calculer les paramètres décrits ci-dessous.

5.2.9.6.1 Calcul des temps de purge d'un canal

On considère un canal d'eau potentiel défini sur une portion de trajectoire donnée et occupé par un nombre fini de molécules d'eau. Son volume est défini par l'ensemble des

positions occupées par les atomes d'oxygène des molécules d'eau qui le constituent sur la totalité de la simulation. En supposant qu'un ensemble de molécule d'eau se trouve dans ce canal à un instant t de la trajectoire : le temps de purge de ce groupe de molécule d'eau est le temps à l'issue duquel la totalité des molécules d'eau constituant ce groupe ont été évacuées hors du volume défini par ce canal. L'arrivée ultérieure d'un nouveau groupe de molécules d'eau entraîne le calcul d'un nouveau temps de purge. Le calcul du temps de purge d'un canal donné permet ainsi de quantifier la dynamique de son eau interne sur l'ensemble de sa trajectoire.

5.2.9.6.2 Détermination de la probabilité de survie de l'eau d'un canal

La probabilité de survie d'un groupe molécules d'eau au sein d'un canal donné est la probabilité que ce groupe reste confiné dans ce canal. La survie de l'eau interne à un canal donné représente son temps de résidence moyen au sein du canal. Connaissant la fraction de molécules d'eau restantes au sein du canal à un temps donné t_{ref} de la simulation, il est possible de calculer leur probabilité de survie en déterminant la fraction de ces molécules d'eau restante dans le canal au temps $t + \tau$ (τ étant le pas de temps séparant deux clichés de trajectoire). En moyennant les probabilités de survie calculées à chaque temps t_{ref} de la trajectoire, une courbe de survie de l'eau peut être tracée (*Figure 33*) et utilisée pour estimer le temps de survie de l'eau au sein du canal d'intérêt. L'allure de la courbe est décrite par l'équation :

$$P(\tau) = \frac{1}{T} \sum_{t=0}^T \frac{N(t, t + \tau)}{N(t)}$$

Où T est le temps maximal de la simulation et τ le pas de temps entre deux clichés de trajectoire. Le paramètre $N(t)$ est le nombre total de molécules d'eau (tous identifiants confondus) comptées à un temps t de la trajectoire. $N(t, t + \tau)$ est le nombre de molécules d'eau restantes au sein du canal au temps suivant $t + \tau$.

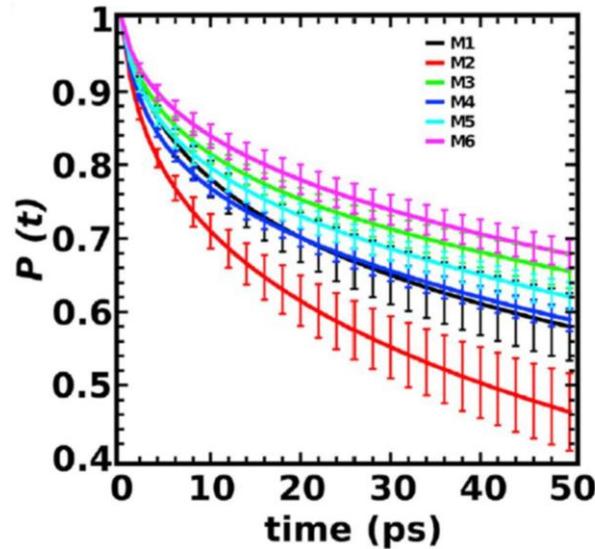


Figure 33. Exemple de tracé de fonctions de survie de l'eau interne à un canal enfoui au sein de la connexine humaine Cx26.

Les courbes de M1 à M6 décrivent l'évolution au cours de la trajectoire de la probabilité de survie de l'eau interne au canal enfoui au sein de chacune des six chaînes monomériques de la connexine Cx26. Reproduit d'après : Araya-Secchi, R.; Perez-Acle, T.; Kang, S.; Huynh, T.; Bernardin, A.; Escalona, Y.; Garate, J.-A.; Martínez, A. D.; García, I. E.; Sáez, J. C.; Zhou, R. *Biophysical Journal* 2014, 107 (3), 599–612.

5.2.9.6.3 Détermination du coefficient de diffusion de l'eau interne

La diffusion est un phénomène physique qui se réfère au transport d'atomes ou de molécules en réponse (en grande partie) à l'agitation thermique du milieu. Le coefficient de diffusion d'une molécule se réfère à sa vitesse de déplacement moyenne en solution. Dans le cadre de ce projet, la détermination du coefficient de diffusion des molécules d'eau internes à une protéine donnée permet d'estimer leur vitesse de déplacement moyenne au sein de sa structure et ainsi de se faire une idée de la contrainte exercée par les résidus enfouis sur le déplacement de ces molécules d'eau. La valeur du coefficient de diffusion de l'eau interne a été déterminée par la relation de Stokes-Einstein ²¹⁸ :

$$D \approx \lim_{\tau \rightarrow \infty} \frac{\langle |r(t+\tau) - r(t)|^2 \rangle}{6\tau}$$

Où D est le coefficient de diffusion et τ l'intervalle de temps entre deux clichés de la trajectoire. $r(t)$ est le vecteur position du centre de masse de l'atome d'oxygène d'une molécule d'eau donnée à un instant t de référence de la trajectoire et $r(t+\tau)$ est le vecteur

position du centre de masse de ce même atome à l'instant $t+\tau$. $|r(t+\tau) - r(t)|^2$ représente le déplacement quadratique de l'atome d'oxygène d'une molécule d'eau donnée dans l'intervalle de temps $[t, t+\tau]$. Les crochets $\langle \dots \rangle$ indiquent le calcul de la moyenne des déplacements quadratiques de toutes les molécules d'eau internes à la protéine à tout instant t de la trajectoire. En pratique, le coefficient de diffusion de l'eau est déterminé graphiquement à partir du tracé des valeurs de déplacements quadratiques moyens des atomes d'oxygène de l'ensemble des molécules d'eau du système étudié calculées à chaque instant de la trajectoire (Figure 34).

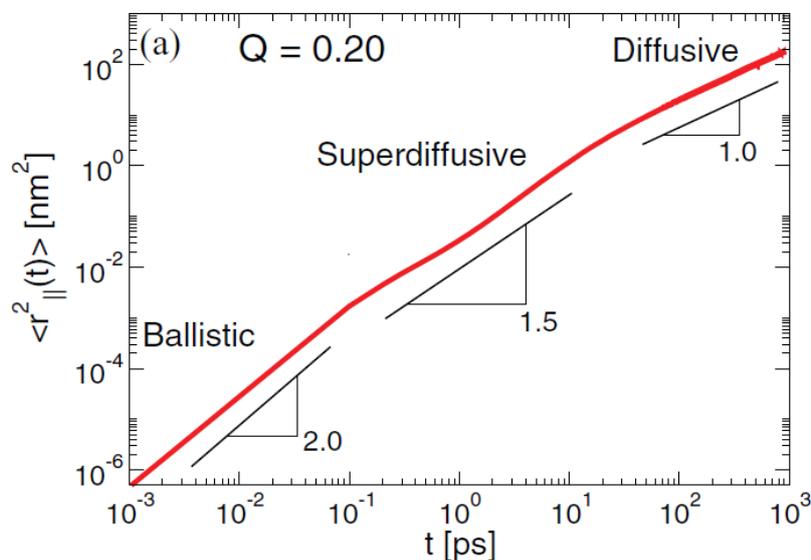


Figure 34. Exemple de détermination de coefficient de diffusion latéral de molécules d'eau entre deux surfaces chargées (coefficient de charge $Q > 0$).

La courbe rouge indique le déplacement quadratique des molécules d'eau entre ces deux surfaces. Les trois segments linéaires indiquent que les molécules d'eau adoptent trois régimes de diffusion distincts à trois intervalles de temps discrets dans la trajectoire. Le régime ballistique se réfère au déplacement sans collision des molécules d'eau les unes par rapport aux autres. Les régimes superdiffusifs et diffusifs se réfèrent au déplacement aléatoire (mouvement Brownien) des molécules d'eau sous l'influence de collisions. Le calcul de la pente de chaque segment linéaire permet d'estimer un coefficient de diffusion. Q est exprimé en unités de charge élémentaire ($e = 1.602 \times 10^{19} C$). Reproduit d'après: Kumar, P.; Han, S. *The Journal of Chemical Physics* 2012, 137 (11), 114510.

- Chapitre III -

Ingénierie de Tt β gly
en transglycosidase

6 Ingénierie de la Tt β gly en transglycosidase

6.1 Présentation du modèle d'étude

La Tt β gly est une β -glycosidase de la famille 1 (GH1) extraite de la souche thermophile de *Thermus thermophilus*. Comme toutes les GH1, cette enzyme se caractérise ainsi par une structure en tonneau (ou TIM-barrel) constitué de 8 brins β et de 8 hélices α (caractéristique du clan A) et présente également un mécanisme catalytique à rétention de configuration. Son site catalytique est constitué de trois sous-sites principaux : le sous-site -1, permettant d'accueillir le sucre donneur, ainsi que les sous-sites +1 et +2, assurant la liaison à l'aglycone (groupe partant) ainsi qu'au sucre accepteur. Les résidus catalytiques de type acide-base (E164) et nucléophile (E338) sont localisés au sein du sous-site -1 (Figure 35).

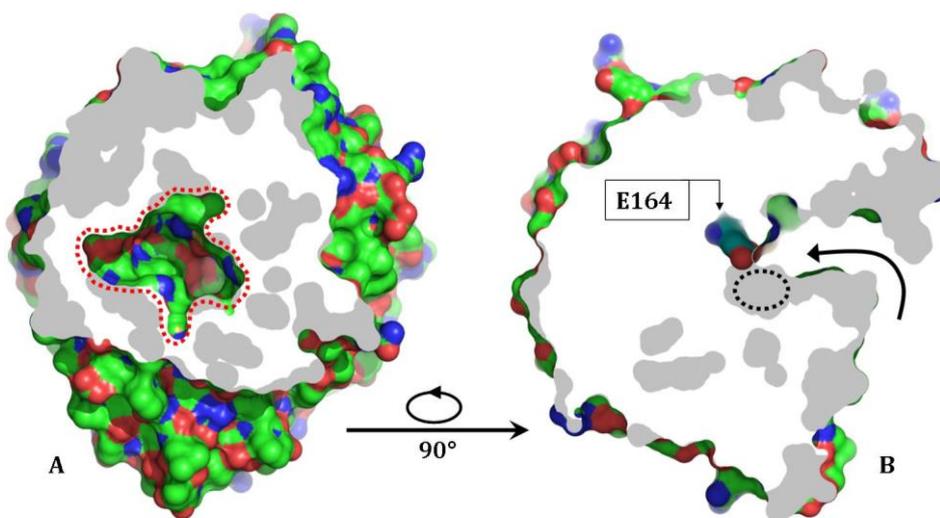


Figure 35. Topologie de surface de Tt β gly et localisation du site catalytique.

Le sous-site -1 est entouré en pointillés. La chaîne latérale du résidu acide-base catalytique (E164) est illustrée en surface. La flèche en B symbolise l'entrée du substrat dans la poche de liaison.

Tt β gly est une exoglycosidase capable d'hydrolyser des β -D-glycosides dérivés du galactose, du glucose, du fucose, et dans une moindre mesure, du xylose²¹⁹. Principalement active sur les liaisons de type β 1-3, cette enzyme présente une spécificité plus contrastée sur les régioisomères de types β 1-2, β 1-4, et β 1-6 avec des taux de conversion respectifs de l'ordre de 71%, 41%, et 10%. Son activité, optimale à

88°C et pour des valeurs de pH comprises entre 6,0 et 7,5, varie également en fonction de la taille du substrat : un rendement de conversion de 100% a ainsi été observé en présence de cellopentaose contre 87,5% et 60% pour le cellotétraose et le cellotriose respectivement²¹⁹. Bien que possédant une activité hydrolytique majoritaire, cette enzyme catalyse également une transglycosylation non-négligeable sur certains substrats. En présence de fucose, son rendement maximal de transfucosylation (transfert de fucose sur le 2-phenyléthanol utilisé comme accepteur) à 88°C a pu être estimé à 63% dans les conditions expérimentales utilisées dans les travaux de Dion et coll.²¹⁹ (substrat accepteur à une concentration 15 fois supérieure à celle du substrat donneur). Son rendement maximal de transgalactosidation en présence d'ONP-β-D-Galactoside avoisine également 50% dans les mêmes conditions.

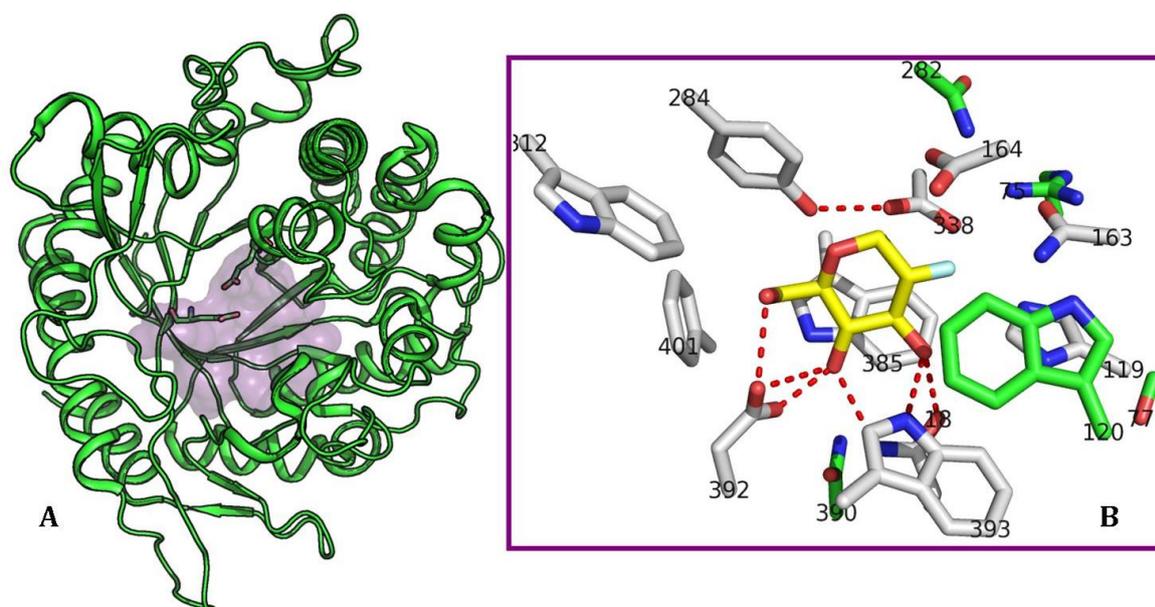


Figure 36. Structure de Ttβgly et résidus conservés du sous-site -1.

A : Structure de la forme sauvage de la Ttβgly. Les acides aminés catalytiques de type acide-base et nucléophile sont représentés en bâtonnets. Le site actif (site -1) est illustré en surface. B : Aperçu des chaînes latérales des acides aminés constitutifs du sous-site -1 et de leurs interactions au substrat donneur. Les chaînes latérales vertes mettent en valeur les résidus de seconde couche.

Un total de 15 acides aminés conservés à plus de 98 %, sur un ensemble des GH1 présentant entre 20 et 90 % d'identité de séquence à la Ttβgly, ont pu être caractérisés. Parmi ces résidus très conservés, la plupart sont localisés au sein du sous-site -1 et sont indispensables à la liaison du sucre donneur. Des expériences de mutagénèse et

d'évolution dirigée ont également pu montrer que les résidus F401, Y284, N390 et N282 étaient fortement impliqués dans le contrôle de la balance entre activités d'hydrolyse et de transglycosylation ^{171,104}. Interagissant respectivement avec le carbone 6 et l'oxygène intracyclique du sucre donneur (*Figure 36*), les résidus F401 et Y284 seraient impliqués dans le maintien d'interactions stabilisant l'état de transition ¹⁷¹. Dans le but de vérifier si une augmentation de l'activité de transglycosylation peut être attribuée à une perturbation de la dynamique de l'eau interne à l'enzyme, la mutagenèse de résidus en interactions avec l'eau interne a été entreprise.

6.2 Recherche de résidus modulant l'activité de transglycosylation

Dans la suite des travaux de Teze et coll. ¹⁷¹, ce travail de thèse s'est consacré à la caractérisation expérimentale de nouveaux mutants susceptibles de favoriser l'activité de transglycosylation de Ttβgly au détriment de l'hydrolyse. Afin de déterminer les positions à muter sur la séquence de Ttβgly, un alignement de sa séquence protéique à deux jeux de séquences de glycoside hydrolases bactériennes et de transglycosylases putatives de plantes issues de la même famille a été préalablement réalisé.

Sur l'intégralité des séquences de GH1 de plantes identifiées par blast comme homologues à celle de la transglycosylase BGlU31 du riz, seules 18 ont été sélectionnées et alignées aux séquences bactériennes. La construction d'un arbre phylogénétique (*Figure 37*) sur l'ensemble du jeu de séquences initial a en effet montré que 18 séquences étaient regroupées sous un même clade en présence de trois des quatre séquences de transglycosylases GH1 de plantes caractérisées dans la littérature (aa5gt ⁶⁷, aa7gt ⁶⁷, et Os9bglu31 ⁶⁸). Apparentées à ces transglycosylases sur le plan évolutif, il a donc été supposé que les autres séquences de plantes appartenant à ce même clade pouvaient être celles de transglycosylases putatives. Le choix des mutations à réaliser s'est finalement basé sur l'identification des positions variables entre l'ensemble des séquences de GH1 bactériennes et eucaryotes après alignement multiple.

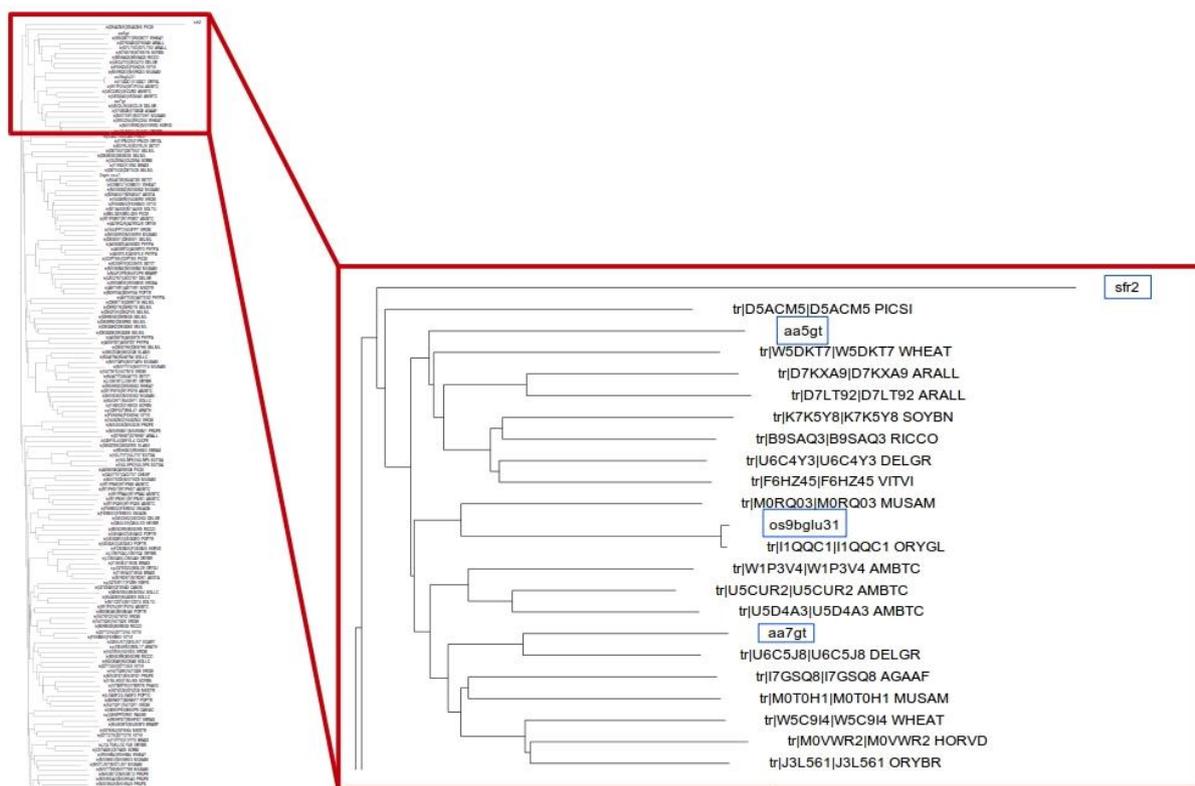


Figure 37. Arbre phylogénétique construit à partir de l'alignement des séquences de GH1.

Les séquences de transglycosidases de plantes sont encadrées en rouge. Parmi celles-ci, celles qui ont été caractérisées expérimentalement sont encadrées en bleu. L'arbre a été visualisé avec le logiciel iTOL v3.4.3 ²²⁰.

En vue de perturber l'organisation de l'eau au sein du canal (ou canal principal) caractérisé par Teze et coll., le choix des mutations candidates s'est volontairement limité aux résidus bordant ce canal. Ce canal, mis en évidence au cours d'une simulation de dynamique moléculaire, canalise l'eau du milieu jusqu'au voisinage du résidu acide-base catalytique et pourrait ainsi jouer un rôle majeur dans l'approvisionnement des molécules d'eau impliquées dans la réaction d'hydrolyse. A l'interface avec le site catalytique, les chaînes latérales des résidus H178 et W120 semblent contrôler l'acheminement de l'eau du canal au voisinage de l'acide-base catalytique. Conjointement aux précédentes expériences de mutagenèse effectuées sur le résidu conservé W120 ¹⁷¹, d'autres travaux réalisés sur un homologue distant de Ttβgly (Tmβgly) ont révélé une contribution importante du résidu H181 (respectivement H178 chez Ttβgly) dans la régulation de la balance entre activité d'hydrolyse et de

transglycosylation (équilibre H/T) chez cette enzyme ¹⁶⁹. Ces résultats suggèrent donc un rôle analogue du résidu H178 chez Ttβglg.

	A188	A187	L182	A179	H178	H173
Conservation de la position chez les GH1 de plantes isolées (transglycosidases putatives)	P : 79.2	F : 62.5	R : 75.0	P : 70.8 A : 29.2	F : 25.0 L : 25.0	Y : 83.3
Conservation de la position chez les GH1 bactériennes (hydrolases)	A : 68.8 G : 14.2	A : 37.8 D : 15.8 E : 10.8	R : 33.0 L : 19.2 H : 10.2	A : 98.9	H : 83.7	Y : 69.4 H : 26.0
Substitutions observées chez les GH1 des plantes caractérisées comme transglycosidases	aa5gt : L os9bglu31 : P aa7gt : P	aa5gt : - os9bglu31 : F aa7gt : -	aa5gt : - os9bglu31 : R aa7gt : R	aa5gt : A os9bglu31 : P aa7gt : P	aa5gt : S os9bglu31 : L aa7gt : I	aa5gt : E os9bglu31 : Y aa7gt : Y
Distance à l'acide base (Å)	16.40	18.87	18.22	13.75	8.73	15.15
Distance au nucléophile (Å)	14.64	17.01	16.05	11.79	6.78	13.02
Mutations envisagées	A188P	A187F	L182R	A179P	H178F H178L H178I	H173Y

	L171	S168	C167	R133	W120	W33
Conservation de la position chez les GH1 de plantes isolées (transglycosidases putatives)	G : 66.7 A : 16.7	F : 29.2 L : 20.8 I : 12.5	I : 33.3 V : 29.2	L : 66.7 I : 12.5	L : 33.3 Y : 25.0 F : 16.7 W : 12.5	W : 66.7 - : 29.2
Conservation de la position chez les GH1 bactériennes (hydrolases)	L : 71.5	S : 48.4 V : 16.7 A : 12.2	C : 72.4 V : 14.5	L : 28.1 P : 23.8 T : 13.3 A : 10.9	W : 99.5	W : 99.6
Substitutions observées chez les GH1 des plantes caractérisées comme transglycosidases	aa5gt : G os9bglu31 : G aa7gt : G	aa5gt : L os9bglu31 : E aa7gt : I	aa5gt : F os9bglu31 : I aa7gt : I	aa5gt : I os9bglu31 : L aa7gt : L	aa5gt : S os9bglu31 : F aa7gt : I	aa5gt : F os9bglu31 : W aa7gt : W
Distance à l'acide base (Å)	8.98	8.85	5.50	19.61	5.32	12.03
Distance au nucléophile (Å)	6.93	6.91	3.39	17.71	3.63	10.32
Mutations envisagées	L171A L171G	S168V S168I	C167I C167V C167M	R133L	W120F W120I	W33F

Tableau 3. Positions ciblées en mutagenèse le long du canal d'eau principal.

Ainsi, sur l'ensemble des résidus conservés au sein des GH1 bactériennes (glycosides hydrolases exclusives) et systématiquement substitués au sein des GH1 de plantes (transglycosylases putatives), ceux situés à moins de 3,5 Å d'au moins une molécule d'eau interne au canal ont préférentiellement été sélectionnés comme candidats pour

une mutagenèse ultérieure. Au total, 13 positions bordant le canal d'intérêt ont finalement été retenues (*Tableau 3*) et testées. Interagissant avec l'eau interne, ces résidus non-catalytiques pourraient contribuer à la canaliser vers le site catalytique et ainsi influencer l'équilibre H/T. La substitution de ces résidus par ceux trouvés au sein des transglycosidases de plantes de la famille 1 (aa5gt, aa7gt, et Os9bglu31) a donc été envisagée dans l'objectif de confirmer leur implication potentielle dans la modulation de l'équilibre H/T chez les GH1 et, le cas échéant, d'appuyer l'hypothèse d'un rôle fonctionnel du canal principal.

6.3 Caractérisation cinétique de la Tt β gly sauvage et de ses mutants

6.3.1 Evaluation de la cinétique de transgalactosylation de l'ONPG

La caractérisation biochimique des 19 mutants sélectionnés a permis de mettre en évidence trois positions influençant significativement l'équilibre H/T : les résidus W120, H178 et C167. En ce qui concerne les deux premiers, ces résultats permettent donc de confirmer les travaux de mutagenèse précédemment réalisés sur Tt β gly¹⁷¹ et son homologue Tm β gly¹⁶⁹. N'ayant aucune participation directe dans l'acte catalytique, au moins deux résidus (W120 et H178) interagissent avec l'eau en sortie du canal et pourraient donc contrôler son passage vers le site actif¹¹⁷. Parmi l'ensemble des mutants présentant une activité de type transglycosidases (réaction de transfert de glycosyle majoritaire sur la réaction d'hydrolyse), les mutants W120I, W120F, H178L, W120FH178L et C167V présentent les plus fortes augmentations d'activité de transglycosylation.

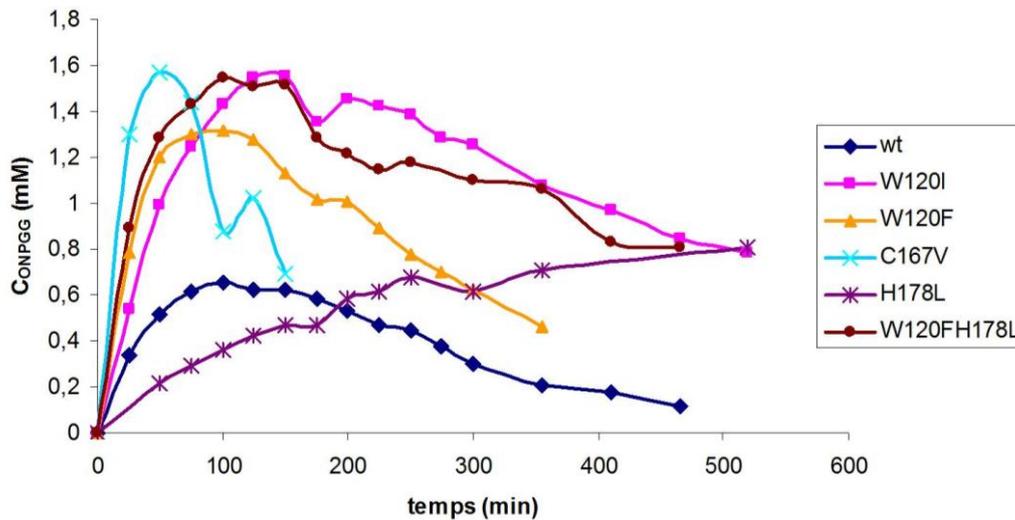


Figure 38. Cinétique de transglycosylation (formation d'ONPG-Gal) de la Ttβgly sauvage et d'une sélection de mutants transglycosidases.

Chaque point mesuré représente un unique essai d'activité cinétique. Les réplicats de ses mesures n'ont pas été représentés sur cette figure par souci de clarté. Le suivi cinétique a été réalisé par électrophorèse capillaire sur ~ 200 μ L de milieu réactionnel (ONPG, 5 mM final) contenant 0.66 μ g d'enzyme et incubé en tampon PBS pH 6,4 à 40°C.

Le suivi de la cinétique de transglycosylation montre cependant que le rendement maximal de synthèse est limité par une hydrolyse secondaire non négligeable des produits de transglycosylation (Figure 38). Cette activité hydrolytique persistante compromet en conséquence l'utilisation de certains de ces mutants à des fins de synthèse. La détermination de la concentration maximale en ONPG-Gal formé au sein du milieu réactionnel associée à la quantité restante de substrat de départ (ONPG) au même temps nous permet d'en déduire un rendement maximal de transglycosylation.

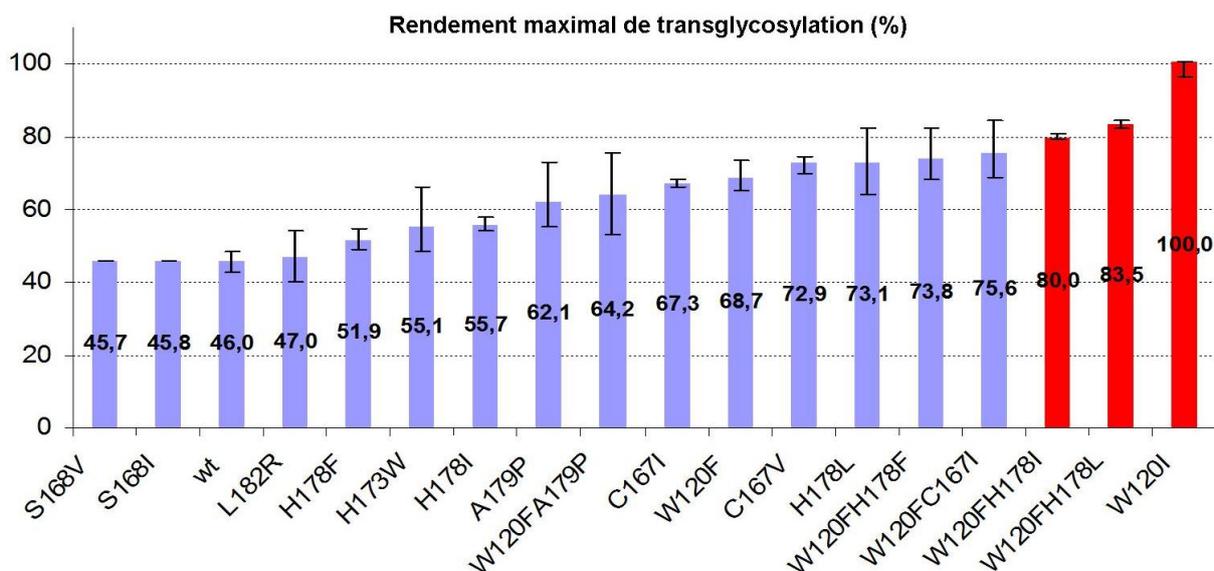


Figure 39. Rendements maximaux de transglycosylation d'une sélection de mutants.

Les mesures ont été effectuées en triplicats. Les valeurs moyennes sont indiquées au centre des barres et les barres inférieures et supérieures précisent les valeurs minimales et maximales respectivement. Les barres colorées en rouge soulignent les mutants présentant un rendement moyen de transglycosylation de plus de 80%. La formule exprimant le rendement maximal de transglycosylation est décrite dans la section 2.2.12 du chapitre 1.

Dans le but d'ajouter les effets des mutations les plus prometteuses et d'augmenter plus fortement le rendement maximal de transglycosylation, des double-mutants ont été produits et testés. La caractérisation des double-mutants W120FH178L, W120FH78I, et W120FC167I montre que l'association de deux mutations provoquant chacune une augmentation de l'activité de transglycosylation induit un effet de synergie sur l'augmentation du rendement maximal de transglycosylation (Figure 39).

6.3.2 Détermination des performances catalytiques des mutants

A l'exception de l'enzyme sauvage, les profils cinétiques de l'ensemble des mutants montrent une absence de saturation malgré l'utilisation de concentrations croissantes en ONPG (Figure 40). L'ONPG jouant à la fois le rôle de donneur et d'accepteur de groupement galactosyle, la cinétique ne suit plus la loi de Michaëlis-Menten lorsque la réaction de transglycosylation devient significative, puisqu'elle est le résultat de la somme de deux réactions simultanées : l'hydrolyse et la transglycosylation. A forte concentration en substrat, la vitesse de transglycosylation augmente plus vite que la réaction d'hydrolyse.

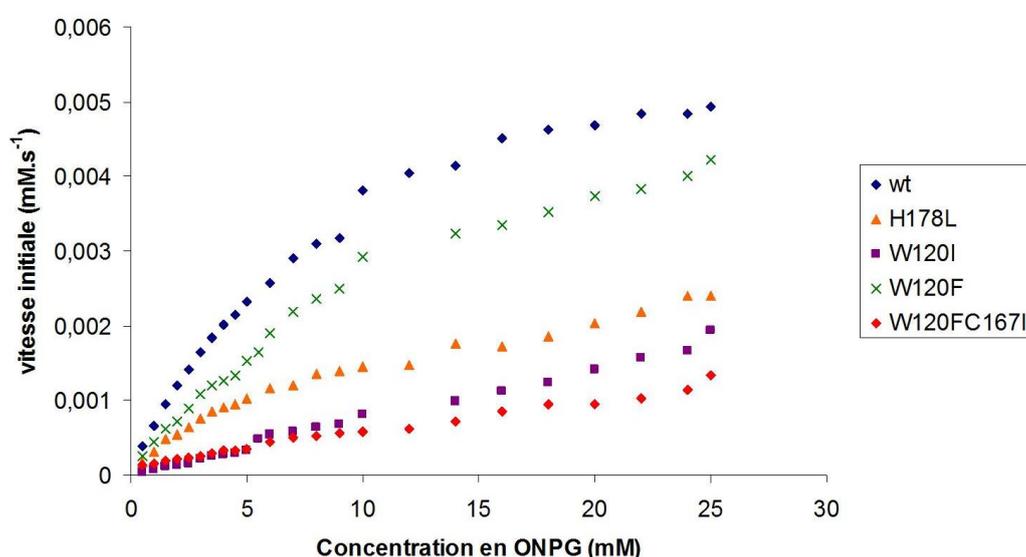


Figure 40. Profil cinétique de la *Ttβgly* sauvage et d'une sélection de mutants transglycosidases en présence de diverses concentrations en substrat.

Le suivi cinétique a été réalisé à 40°C sur une gamme de concentrations d'ONPG de 0,5 à 28 mM, en tampon PBS à pH 6,4.

L'absence de saturation du profil cinétique des mutants aux fortes concentrations en substrat ~~traduit un comportement cinétique non-michaëlien et~~ ne permet donc pas de déterminer les valeurs de leurs constantes cinétiques k_{cat} et K_m . Comme décrit dans le chapitre précédent, l'efficacité catalytique (k_{cat}/K_m) des mutants peut cependant être déterminée en mesurant la pente à l'origine sur les courbes de la figure 40. Le paramètre

k_{cat}/K_m constituant une mesure de la vitesse de glycosylation de l'enzyme, sa mesure permet ainsi déterminer l'effet des mutations sur la première étape du mécanisme.

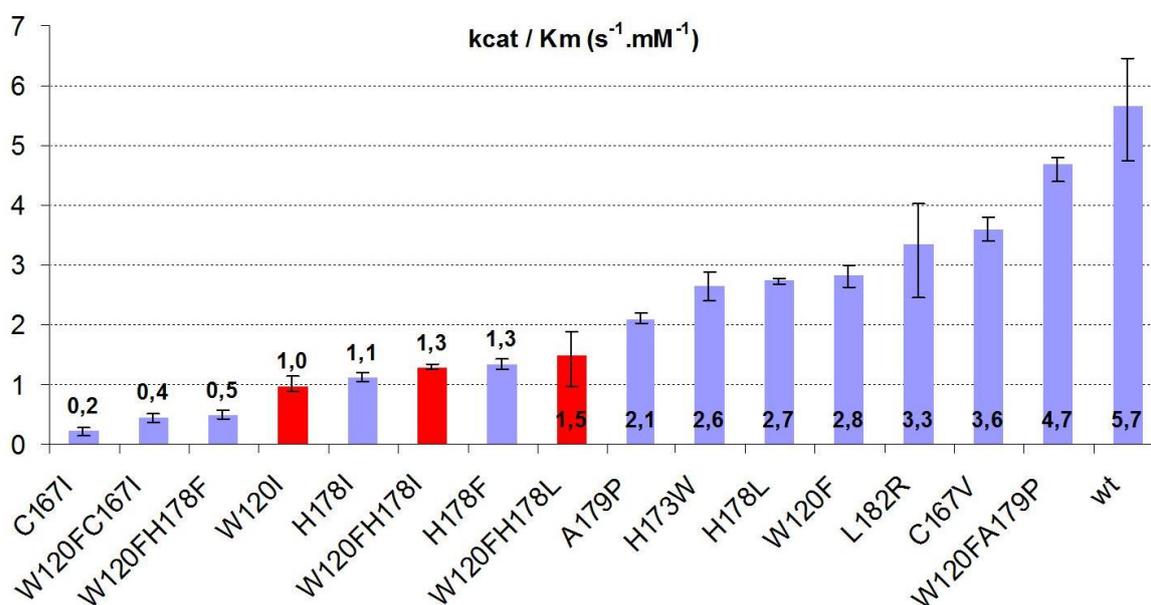


Figure 41. Efficacités catalytiques d'une sélection de mutants transglycosidases.

Les mesures ont été effectuées en triplicats. Les valeurs moyennes sont indiquées au centre des barres et les barres inférieures et supérieures précisent les valeurs minimales et maximales respectivement. Les barres colorées en rouge soulignent les mutants présentant un rendement moyen de transglycosylation de plus de 80%.

Présentant les rendements de transglycosylation les plus élevés, les mutants W120I, W120F, H178L, W120F/H178L et C167V se sont montrés d'autant plus intéressants puisqu'ils conservent une efficacité catalytique (k_{cat}/K_m) modérément réduite (Figure 41) en comparaison à celle de la Tt β gly sauvage (jusqu'à un facteur 5). Les précédents travaux réalisés par Feng et coll.¹⁰² avaient en effet montré que l'efficacité catalytique des variants favorisant le plus la transglycosylation (F401S, N282T) sur l'ONPG chutait d'un facteur 100 environ par rapport à celle de l'enzyme sauvage. Nos mutants sont donc plus intéressants en termes de performances catalytiques.

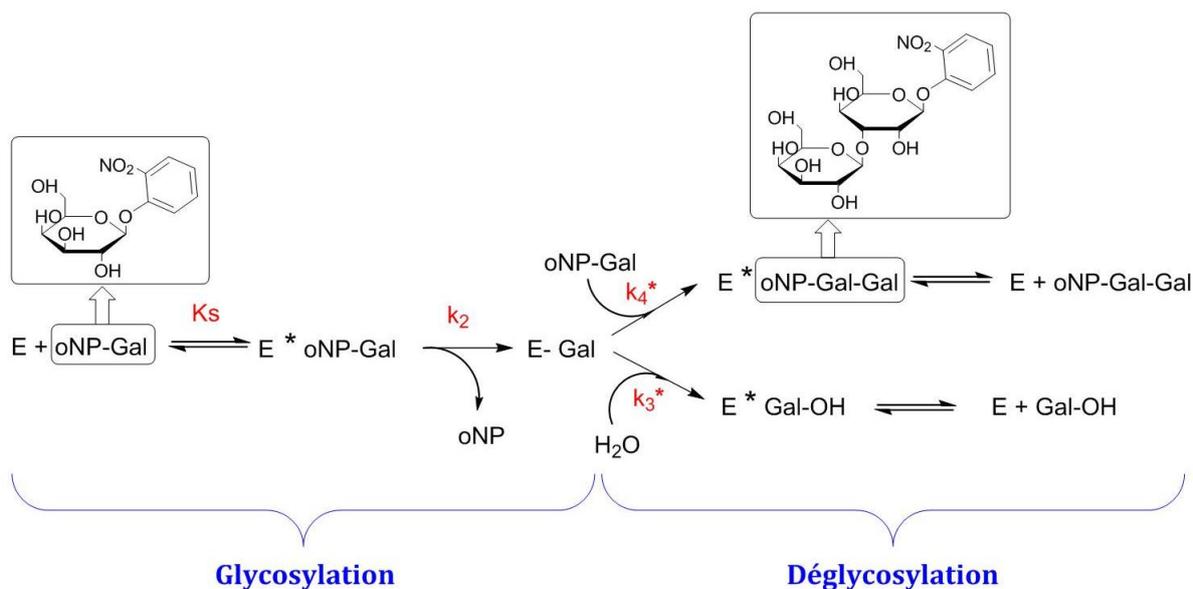


Figure 42. Modèle cinétique des réactions catalysées par la Ttβgly.

Les constantes de vitesses relatives aux différentes étapes réactionnelles sont indiquées en rouge.

Dans le modèle cinétique des réactions catalysées par la Ttβgly (Figure 42), la réaction de glycosylation est constituée de deux étapes distinctes : l'interaction enzyme-substrat régie par la constante d'équilibre K_s , et la formation de l'intermédiaire covalent glycosyl-enzyme régi par la constante de vitesse k_2 . Dans les conditions décrites précédemment, le rapport k_{cat}/K_m équivaut donc au rapport k_2/K_s . La conservation d'une efficacité catalytique relativement proche de celle de l'enzyme sauvage chez ces nouveaux mutants montre ainsi que les mutations introduites aux positions 120, 167, et 178 affectent peu l'état de transition de la réaction de glycosylation de l'enzyme.

6.3.3 Effet des mutations sur la balance H/T: hypothèses

La position 120, située en seconde couche par rapport au sous-site -1, apparaît clairement comme essentielle dans la régulation de la balance hydrolyse/transglycosylation. Bien que l'influence de cette position ait déjà été précédemment répertoriée¹⁷¹, une aussi forte augmentation du rendement de transglycosylation sur l'ONPG n'avait pas encore été décrite. Au vu des rendements maximaux de transglycosylation obtenus pour les mutants W120F et W120I, l'efficacité

de la réaction d'hydrolyse est en grande partie dépendante de la présence d'une chaîne latérale aromatique encombrante en position 120. La substitution du noyau indole par un noyau phényle ou une chaîne aliphatique diminue l'encombrement stérique de la chaîne latérale du résidu 120 et semble en effet être corrélée à une réduction de l'activité d'hydrolyse associée à une augmentation de l'activité de transglycosylation.

La simulation effectuée par Teze et coll.¹¹⁷ avait montré que le noyau indole du résidu 120 jouait le rôle de plateforme réceptionnant les molécules d'eau en sortie du canal à l'interface avec le sous-site -1. Les interactions faibles établies entre l'oxygène des molécules d'eau et le noyau indole du tryptophane pourraient contribuer à les stabiliser et ainsi guider leur repositionnement au voisinage du résidu acide-base catalytique. Un positionnement précis des molécules d'eau catalytiques au sein du sous-site -1 est en effet nécessaire pour stabiliser la géométrie de l'état de transition réactionnel et permettre une catalyse efficace de la réaction d'hydrolyse. Ce rôle hypothétique du résidu 120 permettrait d'expliquer pourquoi la suppression de l'aromaticité de sa chaîne latérale (via la mutation W120I) contribue à augmenter significativement l'activité de transglycosylation. Bien que conservant une relative aromaticité, l'introduction d'un noyau phényle en position 120 réduit considérablement la taille de la plateforme initialement formée par le noyau indole du tryptophane. L'encombrement stérique réduit du noyau phényle lui permettrait également d'explorer un nombre de conformations plus important. En conséquence, les molécules d'eau quittant le canal principal établiraient alors un nombre plus faible d'interactions avec la chaîne latérale du résidu 120 et pourraient donc être moins stabilisées. Ces effets expliqueraient potentiellement l'impact de la substitution W120F sur l'activité d'hydrolyse. Etablissant un contact apolaire avec la chaîne latérale du résidu acide-base catalytique, on ne peut cependant pas exclure la possible implication du résidu 120 dans la stabilisation de la machinerie catalytique.

Associé au résidu 120 en sortie du canal principal, l'histidine 178 est tout comme celui-ci extrêmement conservée au sein des séquences de GH1 bactériennes (*Tableau 3*). Diminuant fortement l'activité hydrolytique, la substitution de sa chaîne latérale par un résidu hydrophobe supprime toute possibilité d'interaction polaire avec l'eau du canal. Chez la Ttβgly sauvage, ce résidu est capable de former une liaison hydrogène avec les

molécules d'eau situées sur la plateforme formée par le tryptophane 120¹¹⁷. Il pourrait alors contribuer à les stabiliser à leur entrée dans le sous-site -1 et ainsi guider leur trajectoire jusqu'au point d'attaque nucléophile. Les pertes d'efficacités catalytiques induites par l'introduction d'une chaîne latérale plus encombrante que celle de la leucine (via les mutations H178I et H178F) à cette position s'expliqueraient par une gêne stérique probable avec le noyau indole du résidu W120.

Le résidu A179, également extrêmement conservé chez les GH1 bactériennes (Tableau 3), est dans une moindre mesure également impliqué dans la modulation de l'équilibre H/T. L'augmentation modérée du rendement maximal de transglycosylation provoquée par la mutation A179P pourrait s'expliquer par le déplacement de la chaîne latérale du résidu H178 induit par la proline en position 179. Le repositionnement du noyau imidazole de l'histidine 178 pourrait affecter son interaction avec l'eau en sortie du canal principal et compromettre dans une moindre mesure l'acheminement des molécules d'eau jusqu'au voisinage de l'acide-base catalytique. Cette diminution très modérée de l'activité de transglycosylation par rapport au mutant W120F a cependant pour mérite d'augmenter considérablement les performances catalytiques de ce mutant (effet additif des rapports k_{cat}/K_m).

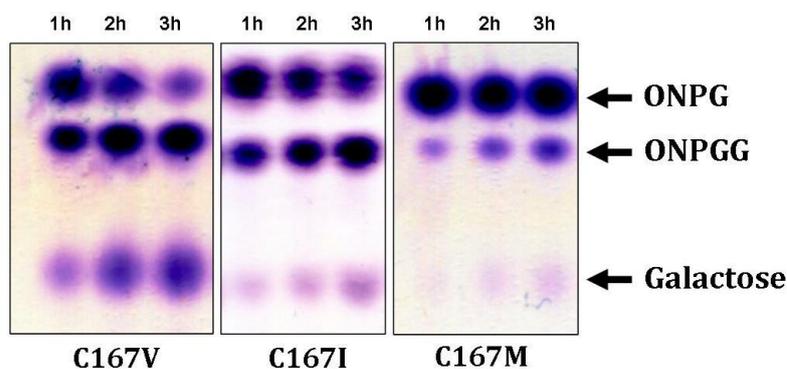


Figure 43. Test de mutants de la position 167 par chromatographie sur couche mince.

La réaction a été réalisée à 40°C pendant 3h sur ~ 200 μ L de milieu réactionnel (ONPG, 5 mM final) contenant 0.66 μ g d'enzyme incubé en tampon PBS pH 6,4. ONPGG indique le produit de transglycosylation (ONPG-Gal).

La substitution de la cystéine en position 167 par une chaîne aliphatique de longueur croissante (valine, isoleucine, méthionine) diminue considérablement les performances

catalytiques de l'enzyme (*Figure 43*). De par sa proximité avec l'acide-base catalytique, le résidu C167 doit donc être impliqué dans la stabilisation de sa chaîne latérale. L'introduction d'une chaîne latérale plus encombrante à cette position pourrait créer une gêne stérique sur celle du résidu E164 et ainsi modifier sa dynamique conformationnelle. La forte diminution des performances catalytiques de l'enzyme en réponse à ces mutations laisse penser que la cinétique de la réaction de glycosylation (première étape du mécanisme catalytique) est fortement ralentie. L'état de transition de la réaction de déglycosylation étant cependant très similaire à celui de la réaction de glycosylation, les effets pourraient également porter sur les deux étapes de la réaction. Au vu de l'augmentation de l'activité de transglycosylation, on pourrait en effet supposer que ces mutations perturberaient sans doute le positionnement de la molécule d'eau catalytique et en conséquence la géométrie de l'état de transition correspondant. Le choix de la taille de la chaîne latérale du résidu en position 167 s'avère donc cruciale afin d'augmenter l'activité de transglycosylation tout en conservant une performance catalytique proche de celle de l'enzyme sauvage.

Provoquant une possible réorientation de la chaîne latérale du résidu H178, l'introduction de la mutation A179P au sein du mutant W120F pénalise légèrement l'effet de la mutation W120F sur l'augmentation du rendement maximal de transglycosylation.

6.4 Analyse de la dynamique de la protéine

La dynamique de la structure protéique des enzymes joue un rôle central dans la catalyse. De nombreux travaux ont en effet montré que l'introduction de mutations dans la séquence de diverses enzymes pouvait modifier significativement la dynamique de leur structure 3D et que cet effet permettait dans certains cas d'expliquer les modifications d'activités associées^{221,222,223}. Une modification de la dynamique des résidus de la machinerie catalytique pourrait, par ailleurs, avoir un impact significatif sur la catalyse. Afin de vérifier si les résultats expérimentaux précédents pouvaient s'expliquer par une modification de la dynamique de la structure protéique de Tt β gly liées à l'introduction des mutations en positions 120, 178, et 167, la réalisation et l'analyse d'une simulation de dynamique moléculaire de l'intermédiaire covalent

glucosyl-enzyme de la Tt β gly sauvage et de ses mutants a été entreprise. L'étude de l'intermédiaire covalent est importante puisqu'il s'agit de l'étape précédant la réaction de déglycosylation au cours de laquelle ont lieu les réactions d'hydrolyse et de transglycosylation.

6.4.1 Stabilité globale et locale du squelette peptidique

Des calculs de RMSD ont été réalisés à partir de la trajectoire de dynamique de chaque mutant afin d'estimer sa stabilité conformationnelle relative par rapport à l'enzyme sauvage. Bien que des différences puissent être observées entre chaque mutant, il ne semble pas y avoir de corrélation évidente entre l'effet de la mutation sur la stabilité globale de la structure et son impact sur l'activité de transglycosylation (Figure 44).

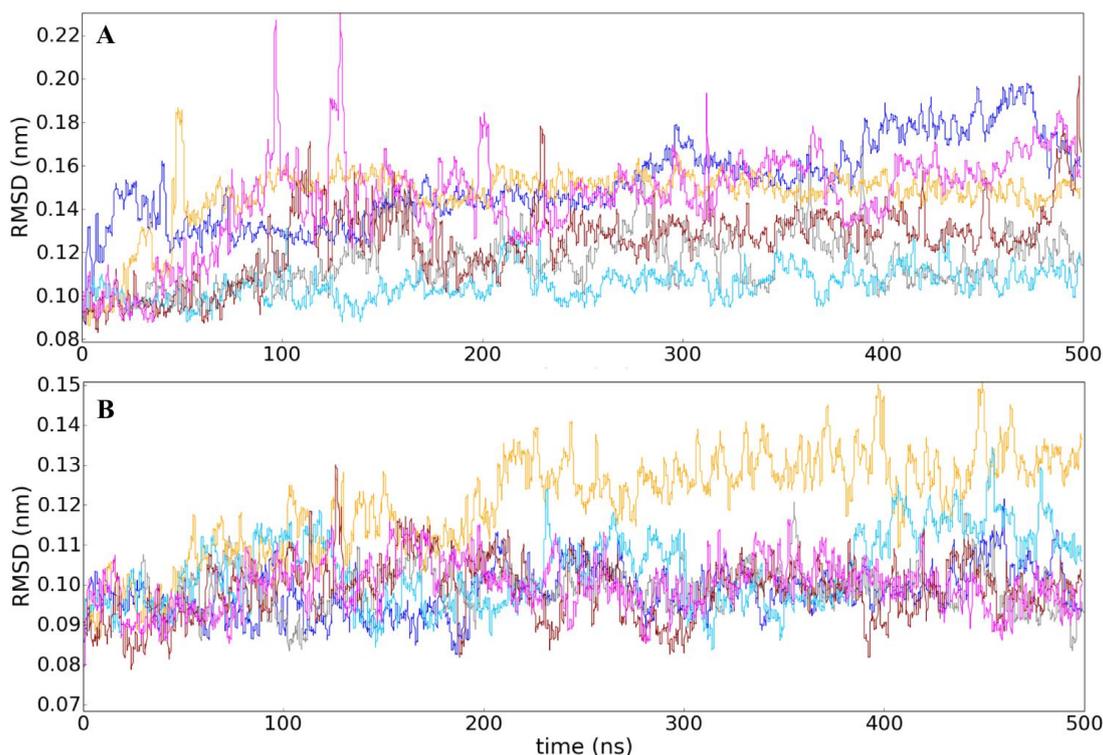


Figure 44. RMSD calculé sur le squelette peptidique.

A – sur l'ensemble de la séquence.

B – sur l'ensemble des régions situées à 10 Å de l'acide-base catalytique.

En rouge : Tt β gly sauvage, en cyan : le mutant W120FH178L, en jaune : le mutant W120I, en bleu : le mutant H178L, en gris : le mutant W120F et en magenta : le mutant C167V.

Il a par ailleurs été montré que la dynamique de la structure protéique est très souvent influencée par celle du solvant interne à celle-ci. Il est donc probable qu'une réorganisation de la trajectoire de molécules d'eau internes sous l'effet d'une mutation puisse en conséquence altérer la dynamique et la stabilité des régions protéiques en contact avec ces mêmes molécules d'eau. Ainsi, en vue de vérifier si la mutagenèse des positions 120, 178, et 167 pourrait affecter la stabilité de la protéine à l'échelle locale, l'ensemble des régions protéiques de chaque mutant présentant une différence de solvation avec l'enzyme sauvage ont été isolées. Un calcul de RMSD a par la suite été réalisé sur chacune de ces régions.

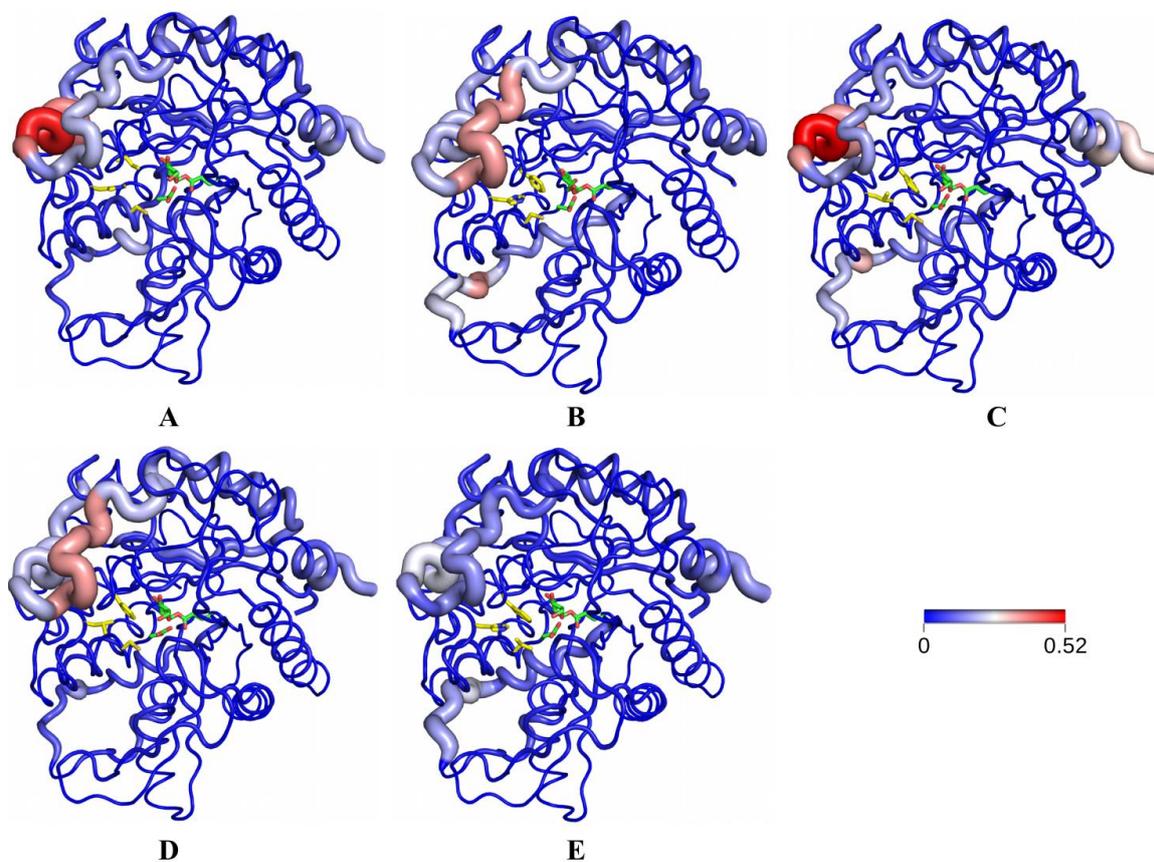


Figure 45. Projection structurale du RMSD (en nm) du squelette peptidique des régions de la protéine montrant des différences de contact avec l'eau enfouie entre la Tt β gly sauvage et chaque mutant.

En A : le mutant W120I, en B : le mutant W120F, en C : le mutant H178L, en D : le mutant H178L, et en E : le mutant C167V.

Les régions montrant les écarts de RMSD les plus élevés ne concernent cependant que des boucles périphériques localisées en surface (Figure 45). La comparaison des

écarts de RMSD observés entre les mutants et l'enzyme sauvage ne révèle également aucune corrélation avec l'effet des mutations sur l'activité de transglycosylation. Pour ces deux raisons, il est donc peu probable que les modifications de la dynamique de ces régions puissent avoir un impact sur l'équilibre de la balance H/T. Le changement conformationnel significatif d'une boucle incluant le résidu acide-base catalytique a pu être en outre observé chez le mutant W120I (*Figure 46*). Ce résultat suggère que la mutation W120I déstabilise le squelette peptidique de ces régions.

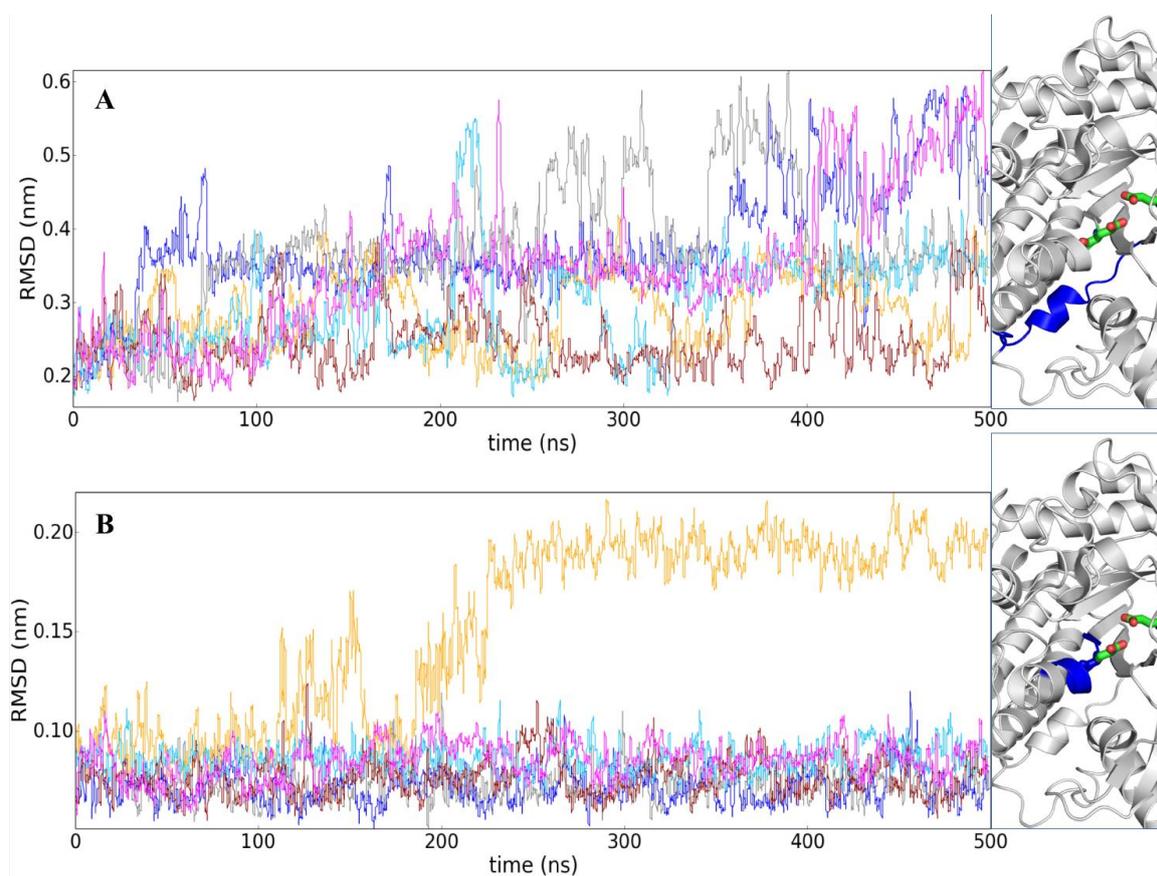


Figure 46. RMSD du squelette peptidique des régions enfouies subissant les plus forts changements conformationnels.

A – région allant du résidu 262 à 277

B – région allant du résidu 162 à 168

A droite de chaque graphe : localisation de la région concernée (en bleu) sur la structure cristallographique de Ttβgly (PDB : 1UG6). En rouge : Ttβgly sauvage, en cyan : le mutant W120FH178L, en jaune : le mutant W120I, en bleu : le mutant H178L, en gris : le mutant W120F et en magenta : le mutant C167V.

6.4.2 Flexibilité du squelette peptidique

Afin d'évaluer l'influence des mutations introduites sur la flexibilité moyenne de la structure protéique de Tt β gly, les fluctuations atomiques moyennées sur les atomes du squelette peptidique de chaque résidu ont été calculés sur l'intégralité des trajectoires. A l'exception de deux boucles externes particulièrement flexibles, les valeurs de RMSF calculées restent extrêmement semblables quelque soit la mutation introduite et ne révèlent ainsi pas de changement conformationnel majeur provoqué par l'introduction des mutations en positions 120, 178, et 167 (*Figure 47 et 48*). Dans le cas des boucles externes présentant les plus fortes variations de RMSF, aucune corrélation évidente ne peut être déduite entre l'impact des mutations sur leur flexibilité moyenne et leur effet sur la balance H/T.

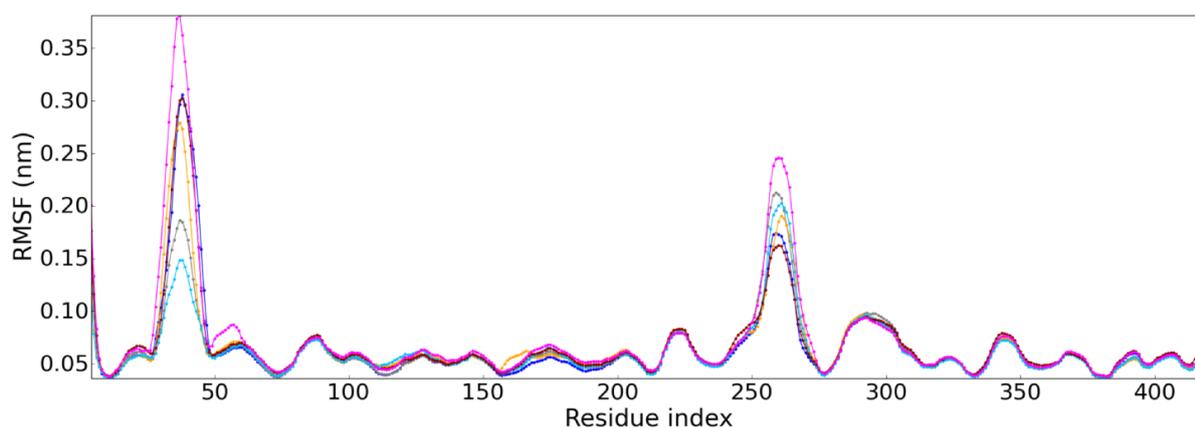


Figure 47. RMSF moyenné sur les atomes du squelette peptidique.

En rouge : Tt β gly sauvage, en cyan : le mutant W120FH178L, en jaune : le mutant W120I, en bleu : le mutant H178L, en gris : le mutant W120F et en magenta : le mutant C167V.

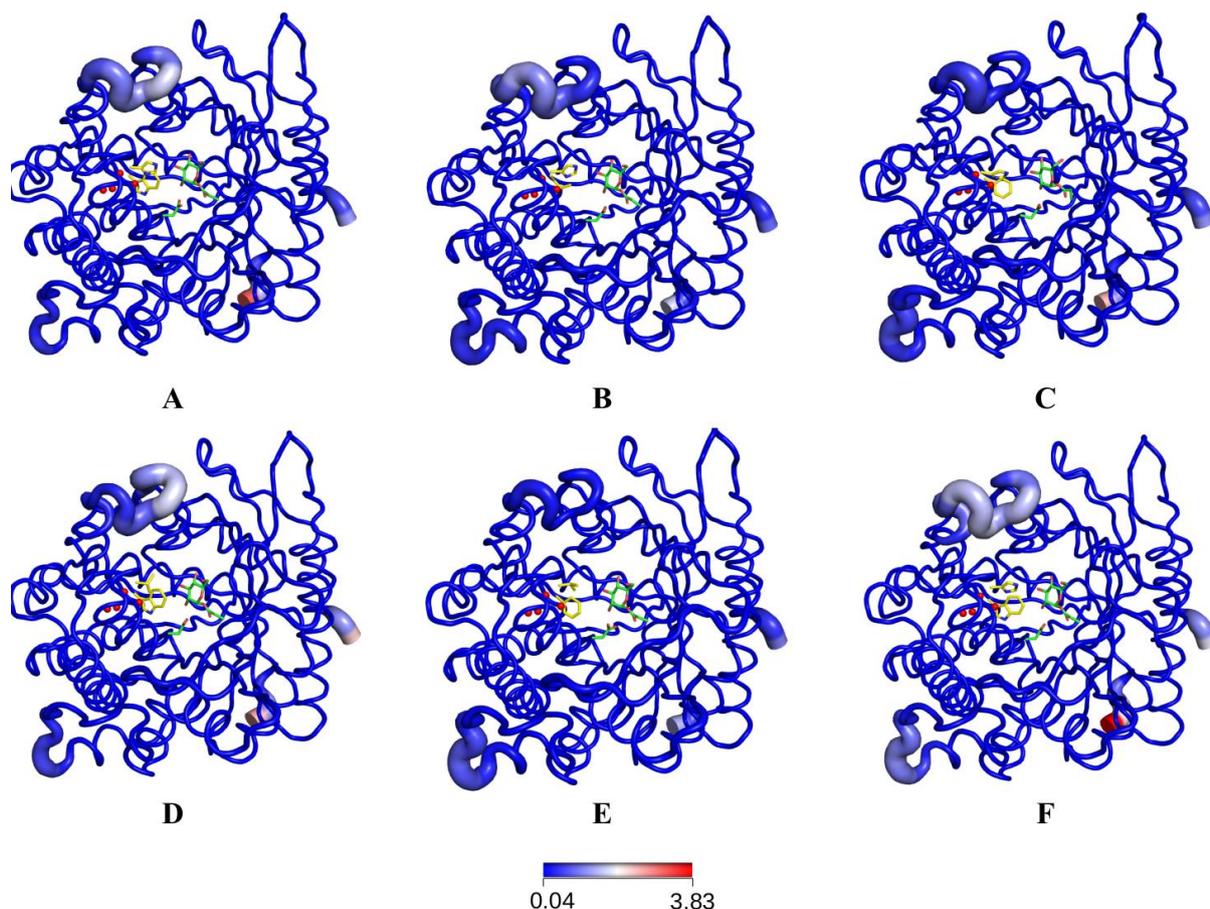


Figure 48. Projection structurale des facteurs B moyens (nm^2) de chaque résidu sur le squelette peptidique des structures dynamiques moyennes (sur 500 ns de trajectoire) de la Ttβgly sauvage (A), et des mutants W120I (B), W120F (C), H178L (D), W120FH178L (E), et C167V (F).

Les valeurs de facteur B ont été calculées à partir des valeurs de RMSF associées à chaque résidu. L'épaisseur du squelette peptidique est d'autant plus importante et sa couleur d'autant plus chaude que la valeur du facteur B est grande.

6.4.3 Flexibilité des chaînes latérales de résidus en contact avec le substrat

Si l'évaluation du RMSD et du RMSF est utile pour caractériser les changements conformationnels moyens du squelette peptidique ou de l'ensemble de la protéine, il est également nécessaire de vérifier si l'introduction des mutations n'affecte pas la dynamique locale des chaînes latérales des résidus intervenant dans la catalyse ou dans la stabilisation des interactions au substrat au sein du site catalytique. La disposition des chaînes latérales de tels résidus est cruciale pour stabiliser l'état de transition associé à l'acte catalytique.

La molécule d'eau catalytique doit en effet être stabilisée dans une géométrie optimale pour pouvoir assurer l'attaque nucléophile initiant la réaction d'hydrolyse. Il paraît ainsi probable qu'une modification de la dynamique conformationnelle des chaînes latérales environnantes pouvant être provoquée par l'introduction de mutations puisse impacter de manière significative la stabilisation de la molécule d'eau catalytique dans sa géométrie d'attaque et ainsi diminuer l'efficacité de la réaction d'hydrolyse au profit de la transglycosylation.

La dynamique des chaînes latérales des résidus catalytiques (E164 et E338) peut être quantifiée par l'évaluation des angles de torsion χ_1 , χ_2 , et χ_3 associés à chaque liaison C-C de leur chaîne latérale. Afin de vérifier si l'introduction des mutations en positions 120, 178, et 167, pourraient affecter l'espace conformationnel exploré par les chaînes latérales de ces résidus, les angles de torsion χ_1 , χ_2 , et χ_3 ont été échantillonnés sur l'intégralité de la trajectoire de la Tt β gly sauvage et de ses mutants. Seule la distribution des angles de torsion χ_1 et χ_3 de ces résidus est cependant présentée ici puisqu'elle permet de mieux différencier l'enzyme sauvage de ses mutants. L'échantillonnage des angles des torsions des chaînes latérales des résidus catalytiques (acide-base et nucléophile) chez les mutants montre que ces résidus adoptent des conformations similaires à celles retrouvées chez l'enzyme sauvage (*Figure 49 et 50*). Si de nouveaux conformères peuvent néanmoins être remarqués chez certains mutants (W120I et W120FH178L), ils ne représentent qu'une fraction très minoritaire (< 5%) de l'ensemble des populations de conformères identifiées sur la trajectoire. Cette observation montre ainsi que le résidu W120 ne semble pas significativement impliqué dans la stabilisation de la chaîne latérale de l'acide-base catalytique ou du glycosyl-enzyme. Il est également intéressant de constater que la mutation C167V ne semble pas perturber la dynamique conformationnelle de la chaîne latérale du résidu acide-base (malgré sa proximité à celle du résidu C167). Bien que plus encombrante, il est possible que la chaîne latérale de la valine soit encore trop courte pour pouvoir créer une gêne stérique avec la chaîne latérale du résidu acide-base catalytique et ainsi influencer sa trajectoire.

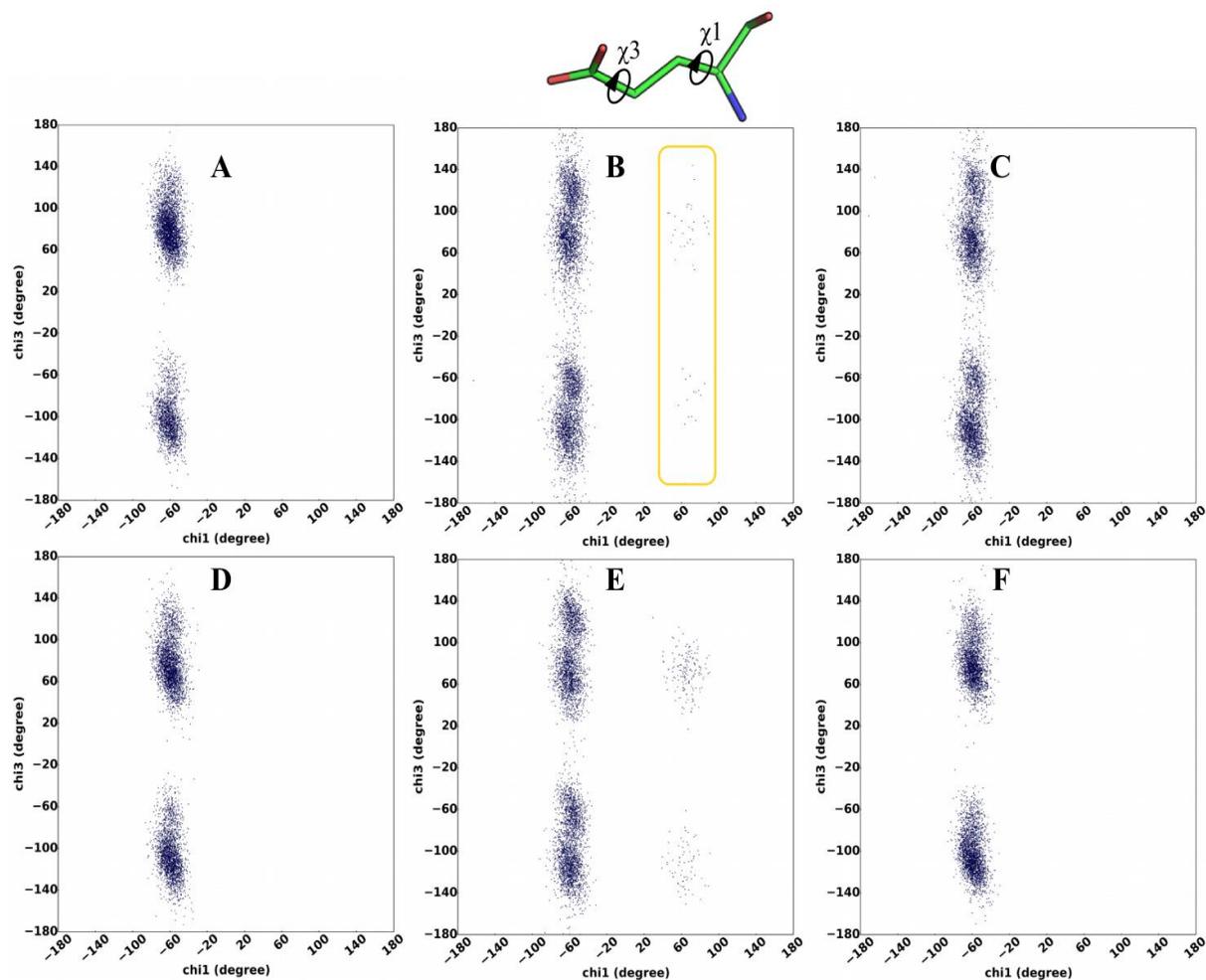


Figure 49. Angles de torsion (χ_1, χ_3) de la chaîne latérale du résidu acide-base catalytique (E164). Les valeurs d'angles remarquables sont encadrées de jaune. En A : la Tt β gly sauvage, en B : le mutant W120I, en C : le mutant W120F, en D : le mutant H178L, en E : W120FH178L et en F : C167V.

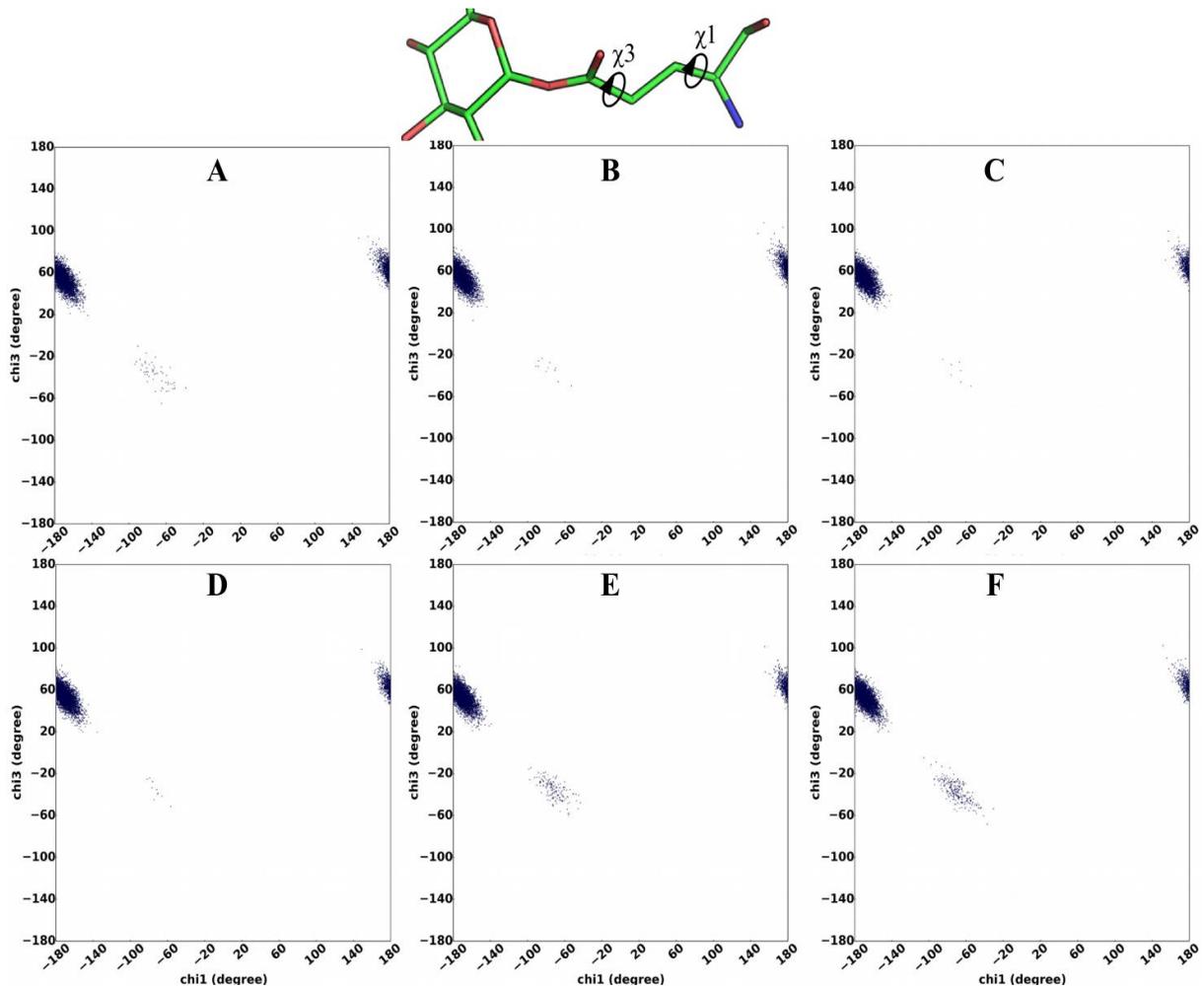


Figure 50. Angles de torsion (χ_1 , χ_3) de la chaîne latérale du résidu nucléophile catalytique (E338). En A : la Tt β gly sauvage, en B : le mutant W120I, en C : le mutant W120F, en D : le mutant H178L, en E : W120FH178L et en F : C167V.

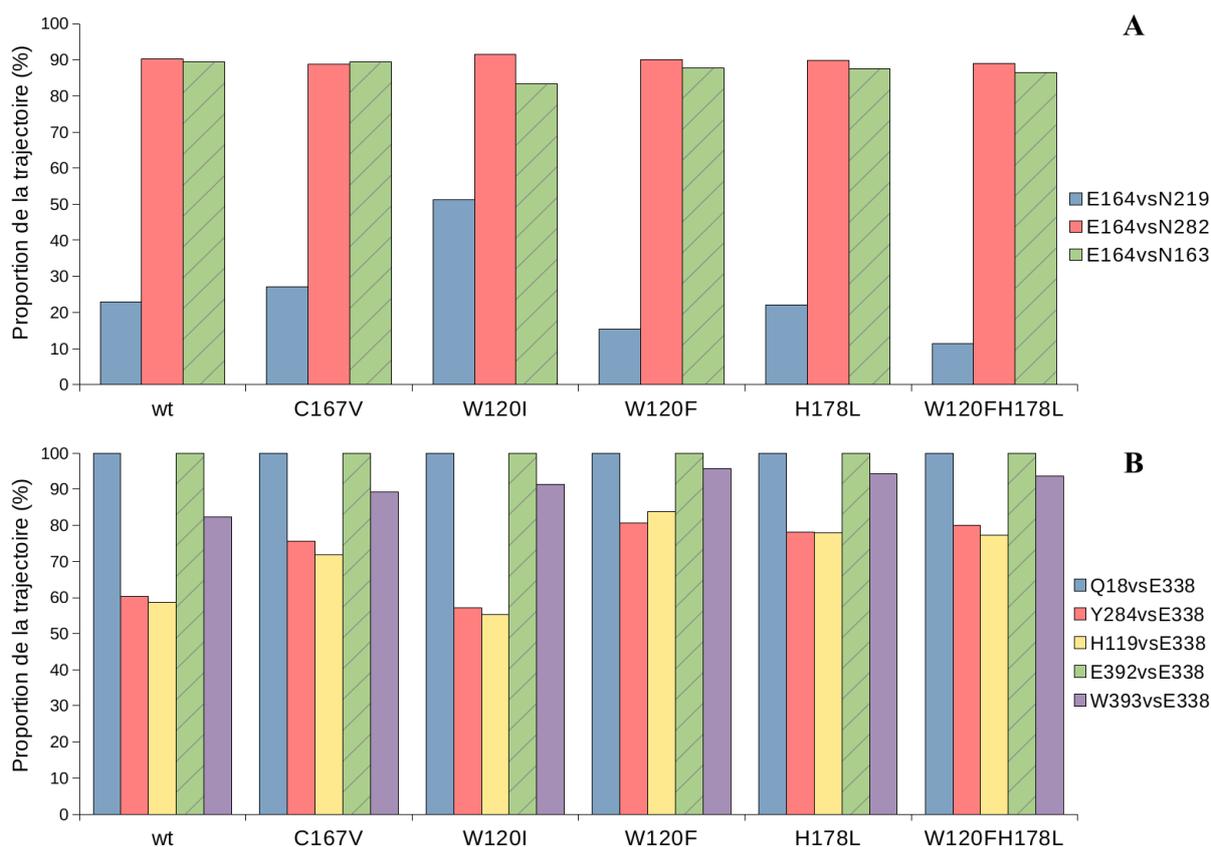


Figure 51. Interactions de résidus conservés du sous-site -1 avec l'acide-base catalytique.

En A : Proportion de clichés de la trajectoire pour lesquels au moins une liaison hydrogène est établie entre la chaîne latérale du résidu acide-base catalytique (E164) et celles des résidus du sous-site -1 impliqués dans sa stabilisation. En B : même graphique pour le résidu nucléophile catalytique (liaisons hydrogènes établies avec la partie glucosyle).

Le réseau d'interaction par liaisons hydrogènes établies entre les résidus catalytiques et les résidus impliqués dans leur stabilisation¹⁷¹ semble peu altéré par l'introduction des mutations en positions 120, 178 et 167 (Figure 51), à l'exception des réseaux d'interactions impliquant le résidu N219 et l'acide-base catalytique (E164) ainsi que les résidus Y284 et H119 avec le nucléophile catalytique. Pour la majorité des résidus étudiés, le nombre d'interactions qu'ils établissent avec les résidus catalytiques reste identique à celui déterminé pour l'enzyme sauvage. Ce résultat est donc en accord avec le fait que ces mutations n'affectent pas de manière significative l'orientation des chaînes latérales des résidus catalytiques (Figure 49 et 50).

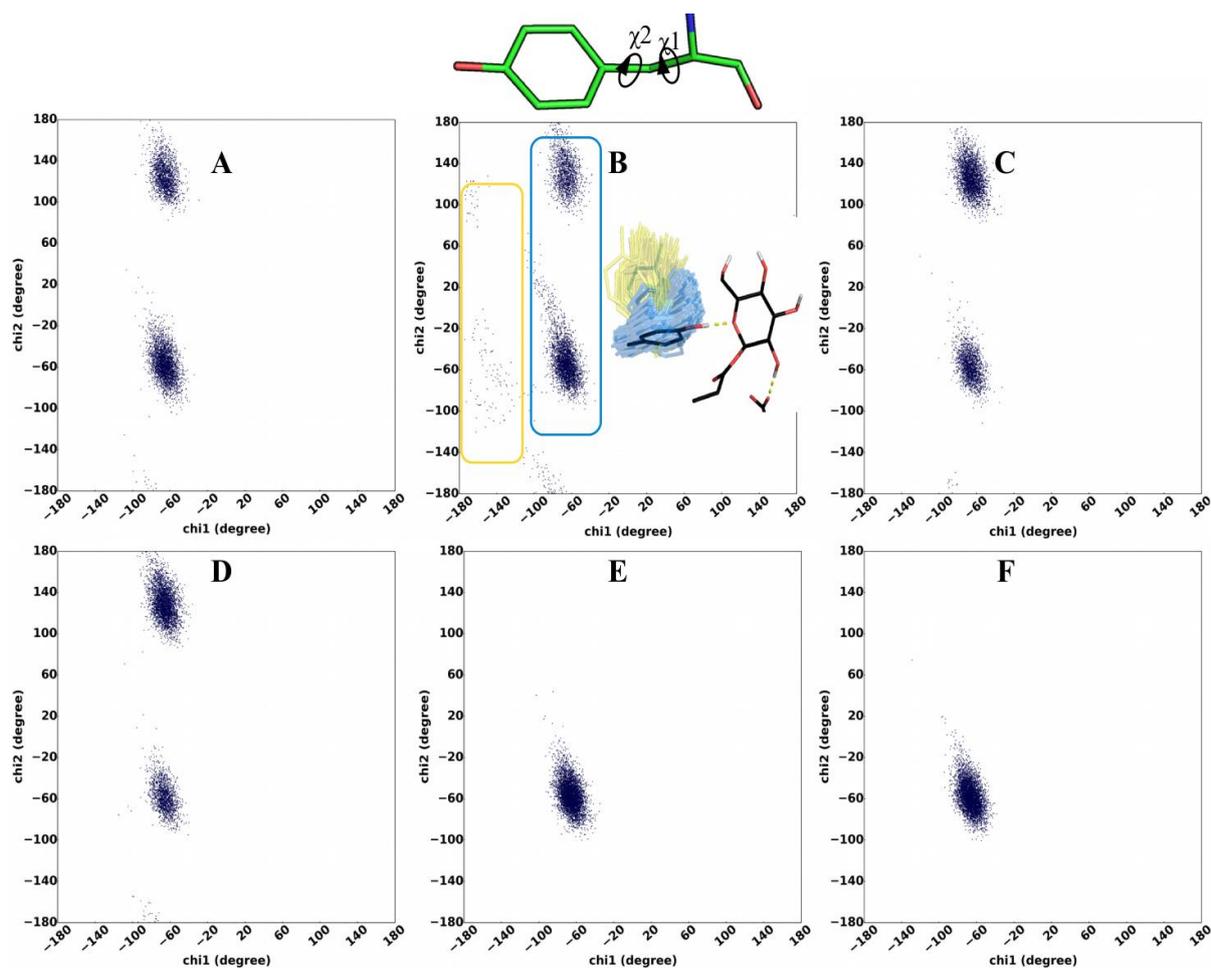


Figure 52. Angles de torsion (χ_1, χ_2) de la chaîne latérale du résidu Y284.

Les couples de valeurs d'angles représentatives des différentes populations de conformères sont encadrées. En A : la Tt β gly sauvage, en B : le mutant W120I, en C : le mutant W120F, en D : le mutant H178L, en E : W120FH178L et en F : C167V.

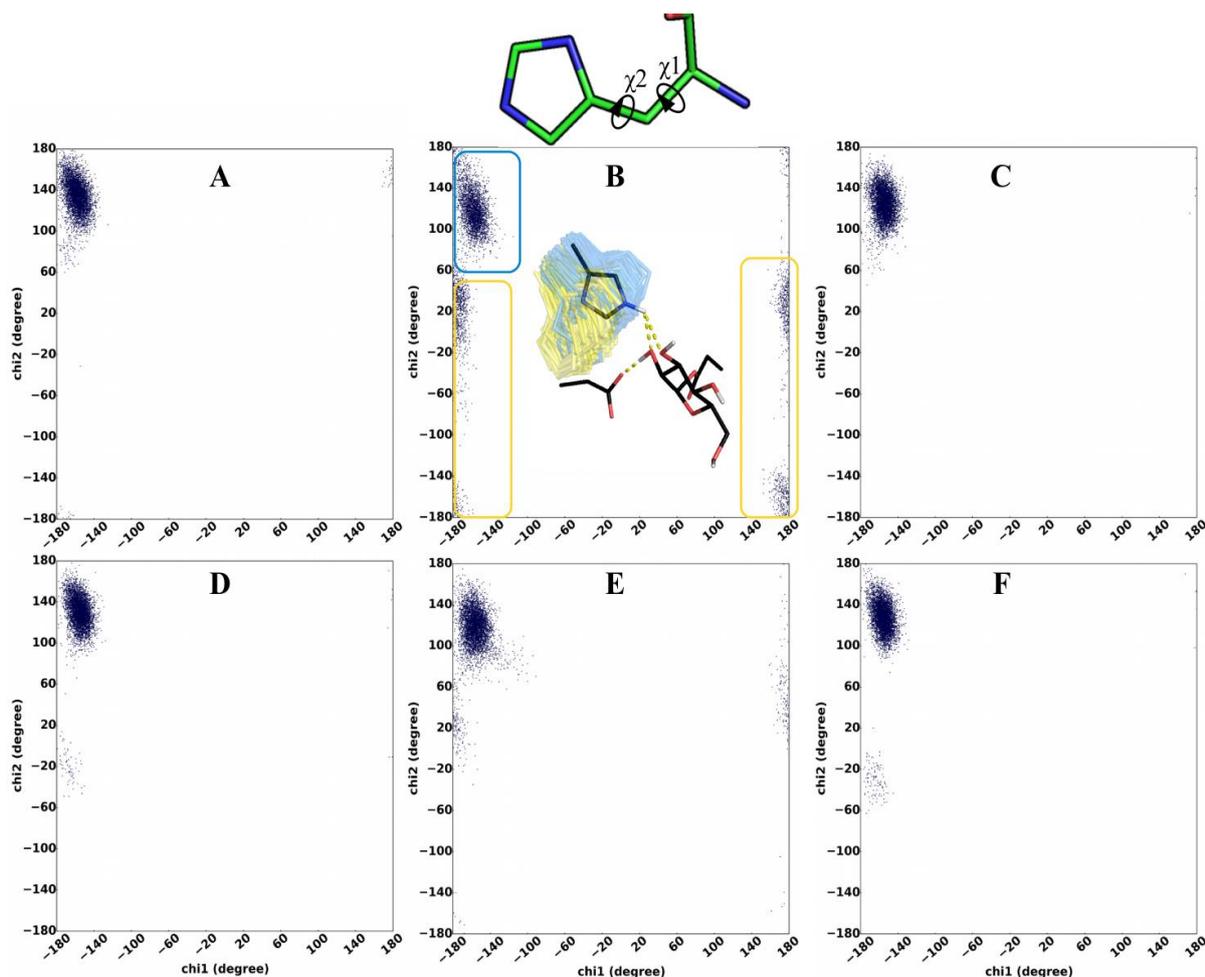


Figure 53. Angles de torsion (χ_1, χ_2) de la chaîne latérale du résidu H119.

Les couples de valeurs d'angles représentatives des différentes populations de conformères sont encadrées. En A : la Ttβgly sauvage, en B : le mutant W120I, en C : le mutant W120F, en D : le mutant H178L, en E : W120FH178L et en F : C167V.

Parmi les résidus de première et seconde couche intervenant dans la stabilisation de l'intermédiaire covalent, seuls les résidus Y284 (Figure 52), H119 (Figure 53), et dans une moindre mesure N219, présentent une dynamique conformationnelle altérée chez certains mutants. Ces modifications n'altèrent cependant pas significativement la dynamique de la chaîne latérale du nucléophile catalytique dont l'orientation reste de manière générale sensiblement identique à celle rencontrée chez la Ttβgly sauvage.

6.4.4 Flexibilité des chaînes latérales des résidus W120 et H178

Formant un goulet d'étranglement en bordure du sous-site -1, ces résidus ont été suspectés d'intervenir dans le contrôle de l'équilibre de la balance H/T chez Ttβgly¹¹⁷ et

son homologue Tm β gly¹⁶⁹. Bien qu'ils ne soient pas directement impliqués dans la catalyse ou dans la stabilisation de la machinerie catalytique, la dynamique de leurs chaînes latérales respectives pourrait cependant jouer un rôle dans le contrôle de l'acheminement des molécules d'eau du milieu au sous-site -1. Leur caractère fonctionnel a par ailleurs bien été confirmé par les résultats cinétiques évoqués précédemment. Si la mutation C167V ne perturbe pas la mobilité de la chaîne latérale de l'acide-base catalytique et du résidu 120, il n'est cependant pas exclu que l'inversion de balance H/T observée chez ce mutant puisse être liée à une perturbation de la dynamique conformationnelle de la chaîne latérale du résidu H178 provoquée par l'introduction de la valine en position 167.

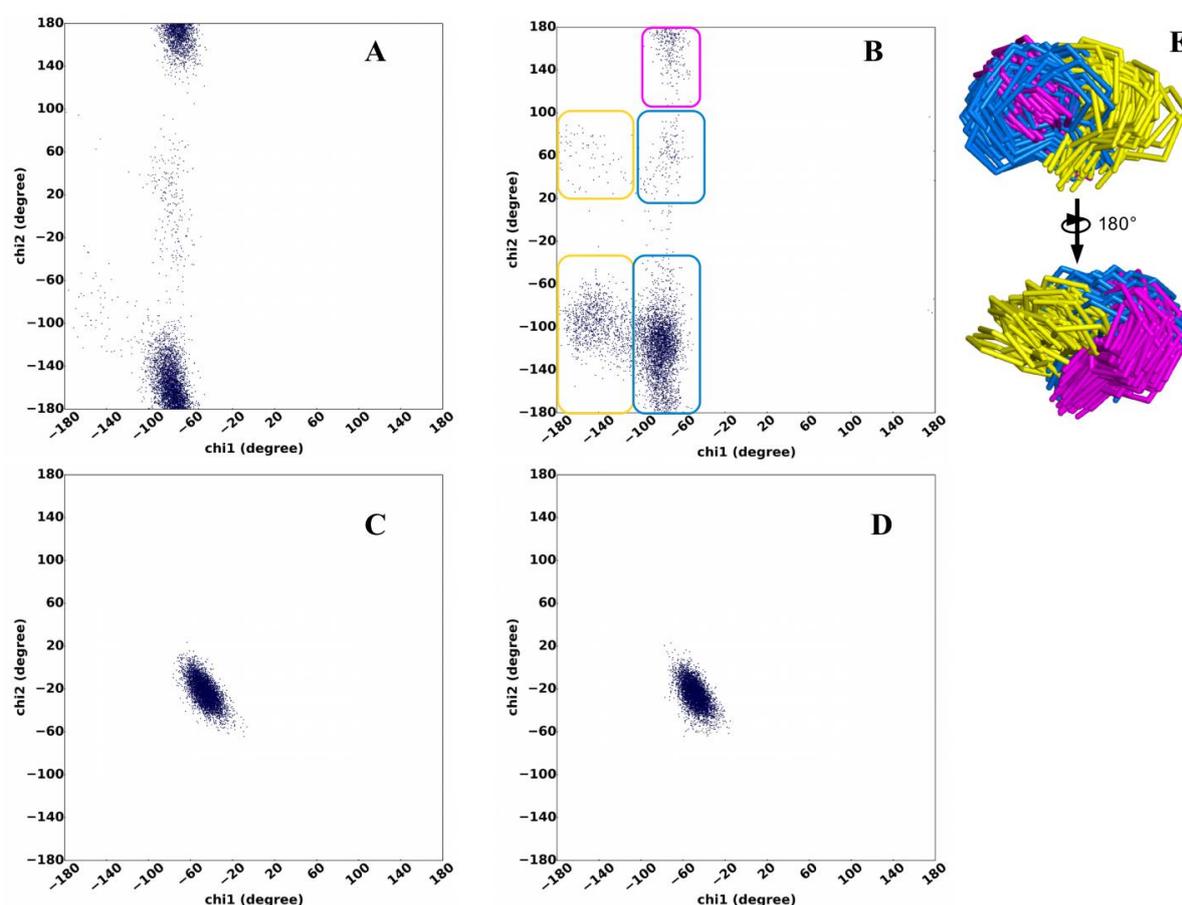


Figure 54. Angles de torsion (χ_1 , χ_2) de la chaîne latérale des résidus H178 et W120.

En A et B : Angles de torsion (χ_1 , χ_2) de la chaîne latérale du résidu H178 chez la Tt β gly sauvage et le mutant C167V. En C et D : Angles de torsion (χ_1 , χ_2) de la chaîne latérale du résidu W120 chez la Tt β gly sauvage et le mutant C167V. Les couples de valeurs d'angles représentatives des différentes populations de conformères du résidu H178 sont encadrés et les conformères correspondants sont illustrés en E.

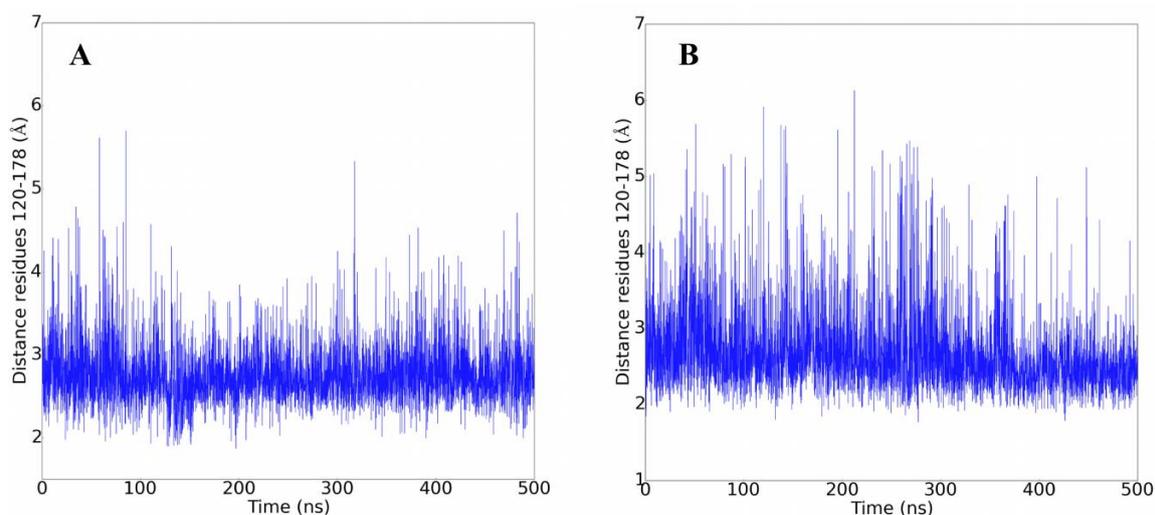


Figure 55. Suivi de la plus petite distance interatomique entre les chaînes latérales des résidus W120 et H178 chez la *Ttβgly* sauvage (A) et le mutant C167V (B).

L'échantillonnage des angles de torsion de la chaîne latérale de l'histidine 178 montre que la mutation C167V modifie bien sa dynamique (Figure 54). L'évaluation de la distance entre les chaînes latérales des résidus H178 et W120 sur l'ensemble des clichés de la trajectoire révèle également que cette mutation tend à les éloigner l'une de l'autre (Figure 55, B). Compte tenu de la distance importante entre les résidus 167 et 178, il est possible que le déplacement de la chaîne latérale du résidu W120 par rapport à celle du résidu H178 puisse expliquer l'effet de la mutation sur sa dynamique. Ces résultats permettent donc de supposer que l'introduction d'une valine en position 167 pourrait donc augmenter l'activité de transglycosylation en modifiant la dynamique conformationnelle du résidu H178.

6.4.5 Conclusions

Les résultats précédents permettent de supposer que les mutagenèses effectuées sur les positions 120, 178, et 167 ne semblent de manière générale pas déstabiliser la machinerie catalytique. Le mutant W120I se distingue cependant des autres mutants, puisqu'une courte région de la protéine incluant l'acide-base catalytique subit un changement conformationnel non-négligeable en comparaison avec l'enzyme sauvage. Bien que l'orientation des chaînes latérales des résidus catalytiques reste préservée, la

plus forte mobilité du squelette peptidique de cette région pourrait avoir un impact important sur la stabilisation de la molécule d'eau catalytique et ainsi diminuer l'efficacité de la réaction d'hydrolyse chez le mutant W120I. Sachant que les autres mutants transglycosidases ne montrent aucune perturbation conformationnelle comparable au voisinage du site catalytique, il est difficile de dire si l'augmentation de la flexibilité de cette région serait exclusivement responsable de la diminution importante de l'activité hydrolytique observée chez ce mutant. Bien que l'introduction des mutations en positions 120, 178, et 167 modifie la flexibilité de certaines régions en surface de la protéine, il paraît cependant peu probable que leur dynamique puisse affecter la catalyse compte tenu de leur distance par rapport au site actif.

De manière générale, les résultats obtenus ne permettent pas d'affirmer avec certitude si les perturbations de la dynamique de la structure protéique observées par l'analyse des trajectoires des mutants contribueraient de manière significative à l'inversion de la balance H/T. Il est donc probable qu'un ensemble d'autres facteurs auxiliaires puissent jouer un rôle dans l'augmentation de l'activité de transglycosylation observée chez les mutants. Par ailleurs, l'efficacité de la réaction d'hydrolyse est conditionnée par l'accès de l'eau au voisinage du résidu acide-base catalytique et du point d'attaque nucléophile. Il n'est donc pas exclu que l'introduction des mutations aux positions 120, 178, et 167 puisse également perturber la dynamique de l'eau interne à l'enzyme et diminuer son accessibilité au site catalytique. Afin de vérifier cette hypothèse, une analyse comparée de la trajectoire des molécules d'eau internes à la Tt β gly sauvage et aux mutants présentant les meilleurs rendements de transglycosylation a donc été entreprise.

6.5 Analyse de la dynamique de l'eau interne à la protéine

6.5.1 Diffusion de l'eau interne au sein de la protéine

Le coefficient de diffusion de l'eau interne à l'ensemble de la protéine a tout d'abord été déterminé afin de vérifier si la plus forte activité de transglycosylation observée chez les mutants pouvait être corrélée à une perturbation de la cinétique de diffusion de l'eau interne à l'enzyme. L'ordre de grandeur des valeurs de coefficient de diffusion calculées sur les 10 premières nanosecondes de la trajectoire est environ 40000 à 80000 fois plus

faible que celui des valeurs expérimentale²²⁴ et théorique²¹⁸ respectivement déterminées sur l'eau liquide et le modèle d'eau TIP3P à 298 K (respectivement égales à $2,39 \cdot 10^{-9} \text{ m}^2 \cdot \text{s}^{-1}$ et $5,6 \cdot 10^{-9} \text{ m}^2 \cdot \text{s}^{-1}$). En dépit de la température de simulation utilisée (350 K), l'obtention de telles valeurs s'explique facilement du fait du confinement de l'eau interne au sein de la structure protéique. Contrairement à l'eau du milieu, les molécules d'eau enfouies dans la protéine établissent de nombreuses interactions avec les chaînes latérales des aminoacides ainsi qu'avec les groupements amide du squelette peptidique et sont donc en conséquence fortement ralenties. Il est donc logique d'observer une diminution importante de la vitesse de diffusion des molécules d'eau enfouies par rapport à celles du milieu.

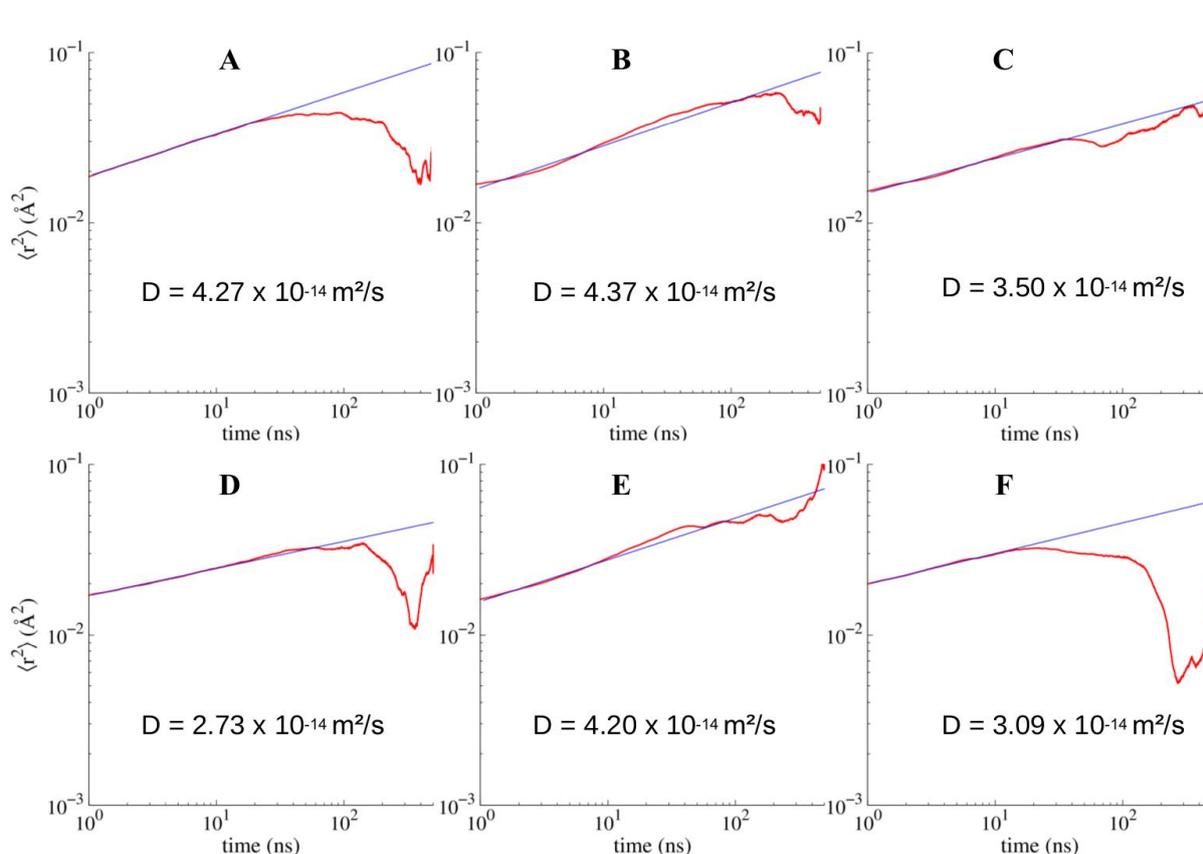


Figure 56. Déplacement quadratique moyen des molécules d'eau internes à la Ttβgly sauvage (A) et les mutants W120I (B), W120F (C), H178L (D), W120FH178L (E) et C167V (F).

La courbe de régression quadratique utilisée pour la détermination du coefficient de diffusion D est représentée en bleu.

En raison de leur confinement au sein de la protéine, le déplacement quadratique des molécules d'eau n'évolue cependant pas de manière linéaire au cours du temps (Figure

56). Le coefficient de diffusion de l'eau interne à la protéine n'est donc pas constant sur l'intégralité de la trajectoire. Au temps longs, une fluctuation importante du déplacement quadratique moyen des molécules d'eau est en effet souvent observée. Ces fluctuations s'accompagnent d'un ralentissement de la cinétique de diffusion de l'eau interne qui adopte alors un régime sous-diffusif. Ce phénomène avait déjà été caractérisé lors d'une étude théorique de la diffusion de l'eau au sein de la GFP ²⁵⁶. Il est ainsi difficile d'évaluer l'impact des mutations sur la diffusion de l'eau interne à l'enzyme en se basant sur cette méthode. Une méthodologie alternative pour estimer la diffusion des molécules d'eau internes à la protéine consiste à calculer le temps moyen mis par ces dernières pour réaliser un déplacement d'au minimum 3 Å (longueur approximative d'une molécule d'eau).

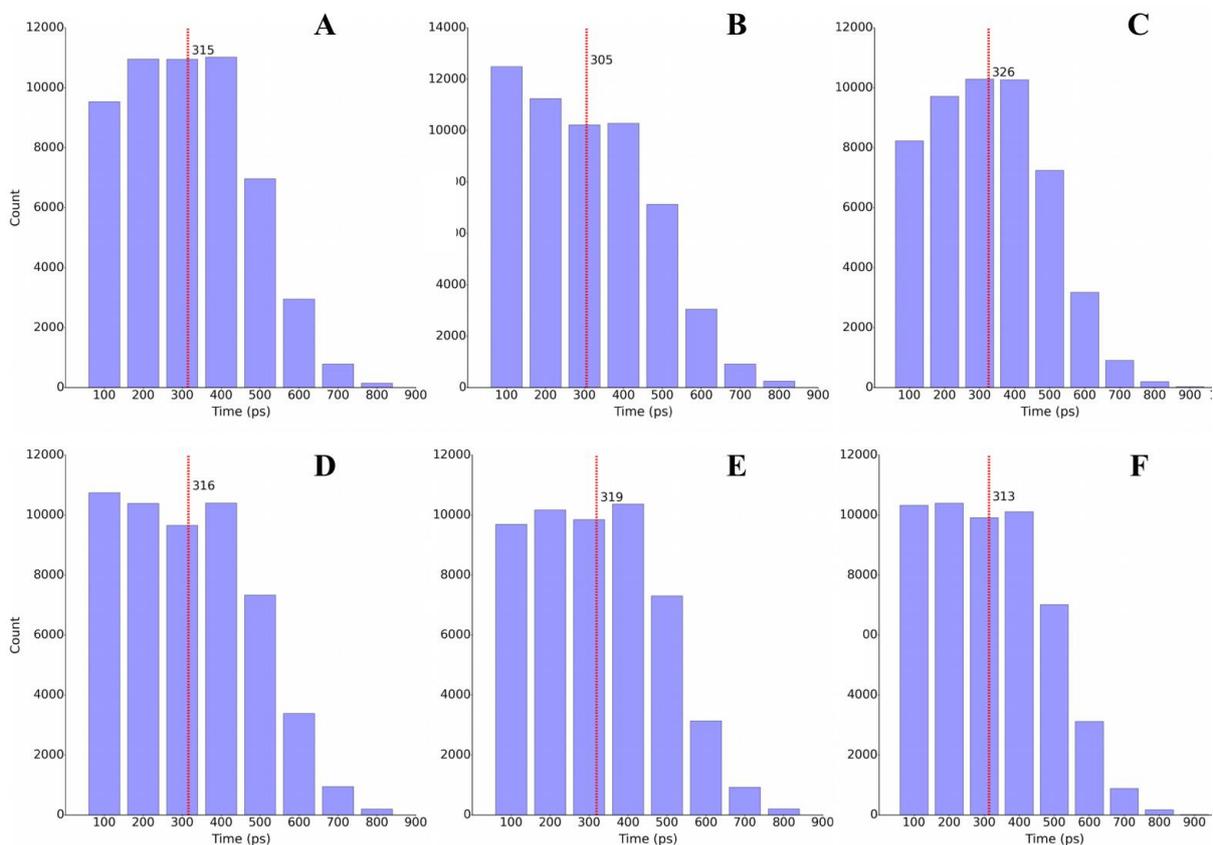


Figure 57. Distribution du temps mis par les molécules d'eau internes à Ttβgly pour parcourir une distance d'au minimum 3 Å au sein de la protéine.

En A : la Ttβgly sauvage, en B : le mutant W120I, C : W120F, D : H178L, E : W120FH178L, et F : C167V. La valeur moyenne est indiquée par une ligne verticale rouge.

Quelque soit le système étudié (enzyme sauvage ou mutée), le temps moyen nécessaire aux molécules d'eau internes pour parcourir une telle distance n'est pas significativement modifié par l'introduction des mutations (*Figure 57*). La vitesse de diffusion de l'eau interne à l'enzyme ne semble donc pas globalement affectée par l'introduction des mutations.

Ce résultat conforte ainsi les analyses précédentes qui ont montré que les mutations n'avaient pas d'effets significatifs sur la dynamique conformationnelle de l'enzyme. Impliquée dans un réseau de liaisons hydrogène avec les résidus enfouis, les molécules d'eau internes jouent un rôle fondamental dans la stabilisation de l'architecture de l'enzyme. Il paraît ainsi logique que toute modification de leur vitesse de diffusion altérerait en conséquence le réseau de liaisons hydrogène dans lequel elles sont impliquées et devrait ainsi s'accompagner d'une modification de la flexibilité de la structure protéique. Il n'est donc pas surprenant de ne pas observer de modification majeure de la cinétique de diffusion de l'eau interne en l'absence de changements conformationnels notables au sein de la protéine.

Si la diffusion de l'eau interne prise sur l'ensemble de la structure ne semble pas être modifiée par les mutations, il n'est cependant pas exclu que la dynamique des molécules d'eau enfouies ne soit affectée que localement. Comme suggéré au chapitre 1, il est par ailleurs possible que l'eau catalytique puisse être acheminée jusqu'au sous-site -1 par le biais des crevasses ou des canaux internes formés au sein de la structure protéique. En vue de vérifier si les mutations introduites pourraient influencer la dynamique de l'eau interne à ces cavités, il convient tout d'abord de les identifier.

6.5.2 Identification de canaux d'eau internes à la protéine

Un total de près de 4000 réseaux de molécules d'eau internes ont pu être isolés sur 500 ns de trajectoire de dynamique moléculaire de la Tt β gly sauvage et de ses mutants. Sur l'ensemble de ces réseaux, entre 10 et 20 sont localisés au voisinage de six clusters de molécules d'eau cristallographiques conservés au sein de plus de 90 % des structures de GH1 connues (*Figure 58*).

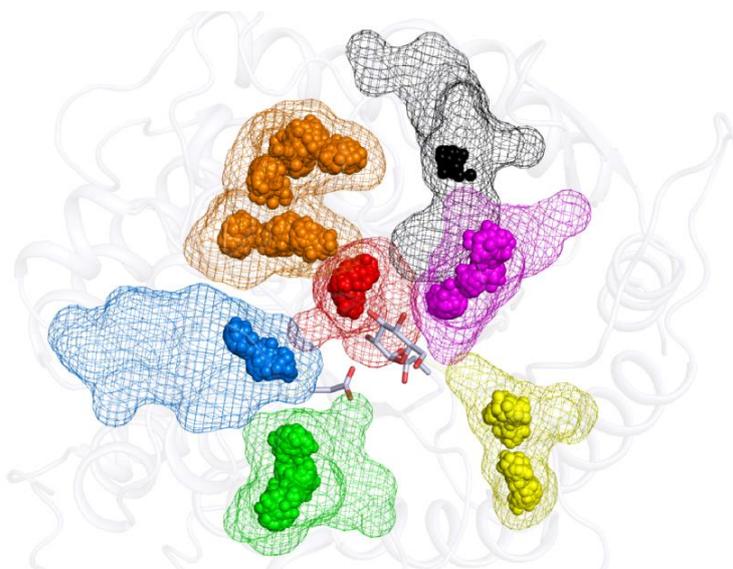


Figure 58. Localisation des clusters de molécules d'eau cristallographique chez la Ttβgly sauvage. Les réseaux de molécules d'eau isolés sur 500 ns de trajectoire de la Ttβgly sauvage et présentant une trajectoire commune sont délimités par un grillage de couleur unique. Les positions de molécules d'eau internes trouvées au plus à 1.6 Å d'un des six clusters de molécules d'eau cristallographiques conservés sont identifiées par une sphère.

Identifiés par Teze et coll. en bordure du sous-site -1, ces clusters de molécules d'eau pourraient de par leur très forte conservation remplir un éventuel rôle fonctionnel (structural ou catalytique). Compte tenu du nombre important de réseaux de molécules d'eau isolés sur les trajectoires, seuls les réseaux situés au voisinage de ces clusters ont été pris en compte pour la caractérisation des canaux d'eau bordant le sous-site -1. L'identification d'un canal d'eau a été réalisée en regroupant les réseaux de molécules d'eau situés au voisinage d'un même cluster d'eau cristallographique.

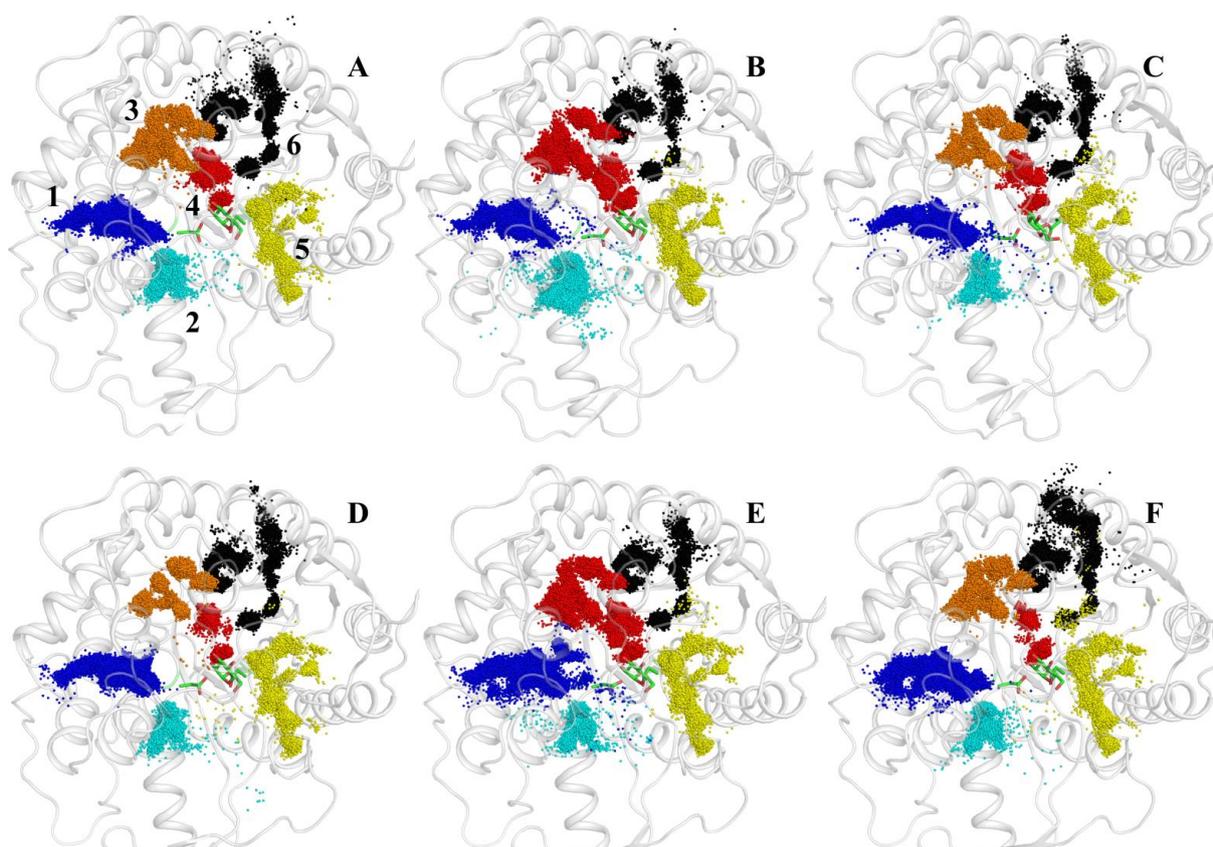


Figure 59. Canaux d'eau isolés sur 500 ns de trajectoire de dynamique moléculaire chez la Ttβgly sauvage (A) et les mutants W120I (B), W120F (C), H178L (D), W120FH178L (E) et C167V (F). Les résidus catalytiques sont représentés en bâtonnets verts.

Dans ce cadre, six canaux d'eau internes ont pu être identifiés (Figure 59) après regroupement des réseaux de molécules d'eau internes précédemment caractérisés. Parmi ces canaux, quatre d'entre eux débouchent sur le sous-site -1. Ces observations suggèrent donc qu'au moins un de ces canaux pourrait éventuellement être impliqué dans l'approvisionnement des molécules d'eau catalytiques au sein du site actif de la Ttβgly. L'isolement des chaînes latérales provenant des résidus bordant chacun de ces canaux montre que ceux-ci sont très majoritairement ceinturés de nombreux résidus apolaires (G, A, V, L, I) et aromatiques (F, Y, W) (Tableau 4).

	P	N, Q	F, Y, W	G, A, V, L, I	K, R, H	D, E	C, M	S, T
Canal 1	0.04	0	0.2	0.48	0.12	0.04	0.04	0.08
Canal 2	0.05	0.11	0.32	0.21	0.16	0.11	0.05	0
Canal 3	0.1	0	0.38	0.29	0.05	0.1	0.05	0.05
Canal 4	0.06	0.03	0.24	0.27	0.09	0.12	0	0.18
Canal 5	0	0.07	0.15	0.26	0.19	0.11	0.11	0.11
Canal 6	0	0.11	0.37	0.11	0.11	0.16	0.05	0.11
Moyenne	0.04	0.05	0.28	0.27	0.12	0.11	0.05	0.09

Tableau 4. Proportion des acides aminés bordant les canaux d'eau.

Il est probable que ces résidus doivent jouer un rôle dans la formation des cavités au sein desquelles se forment ces canaux. L'acheminement des molécules d'eau à proximité du glucosyle pourrait être contrôlé par la présence de chaînes latérales aromatiques impliquant majoritairement un tryptophane. Des couples de résidus histidine/tryptophane (W120/H178, W166/H238, W393/H119) ou dans certains cas de deux résidus aromatiques (W312/F401) sont par également souvent retrouvés à l'embouchure des canaux 1, 2, 5 et 6 avec le sous-site -1.

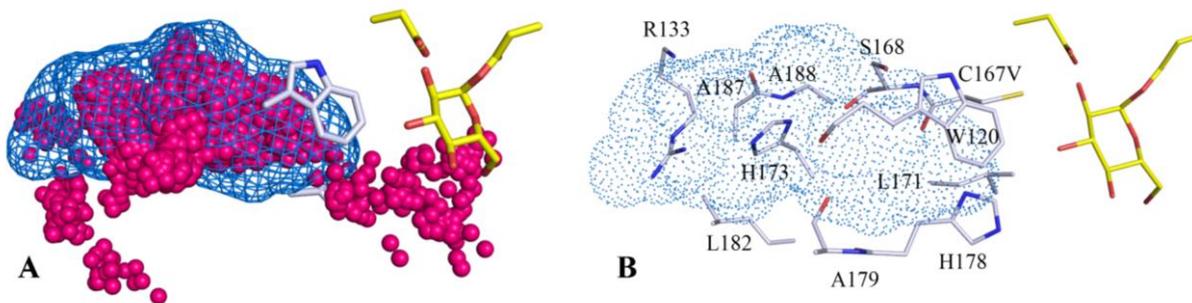


Figure 60. Canal d'eau principal (ou canal 1) isolé sur 500 ns de trajectoire de dynamique moléculaire chez la *Ttβgly* sauvage.

En A : comparaison de la trajectoire des molécules d'eau obtenue dans ces travaux (grillage bleu) à celle obtenue par Teze et coll. (sphères magenta). En B : Aperçu des résidus bordant le canal et mutés dans ces travaux.

L'identification d'un canal (canal 1 ou canal principal) analogue à celui mis en évidence par Teze et coll. permet de confirmer les résultats des précédents travaux. La trajectoire du canal principal identifiée par notre simulation diffère cependant de celle décrite dans ces travaux. Bien que la majeure partie des positions d'oxygènes de molécules d'eau restent conservés entre les deux trajectoires, le canal isolé au cours de notre simulation ne débouche pas au voisinage de l'acide-base catalytique (*Figure 60*). Les résidus W120 et H178 semblent en apparence bloquer la progression de l'eau vers le sous-site -1. En dépit de la dynamique conformationnelle de la chaîne latérale du résidu H178, la rigidité de celle du résidu W120 conforte cette observation. A l'inverse, la trajectoire générée par Teze et coll. révèle un canal d'eau plus étendu, marqué notamment par la présence de positions supplémentaires à l'interface avec le sous-site -1. Il est par ailleurs remarquable de noter que le canal obtenu par Teze et coll. s'identifie

à la trajectoire d'une unique molécule d'eau sur 50 ns. Bien qu'étendue à 500 ns, notre simulation ne permet cependant pas d'identifier à première vue la présence de molécules d'eau provenant du canal dans le sous-site -1 chez l'enzyme sauvage (*Figure 61, A*).

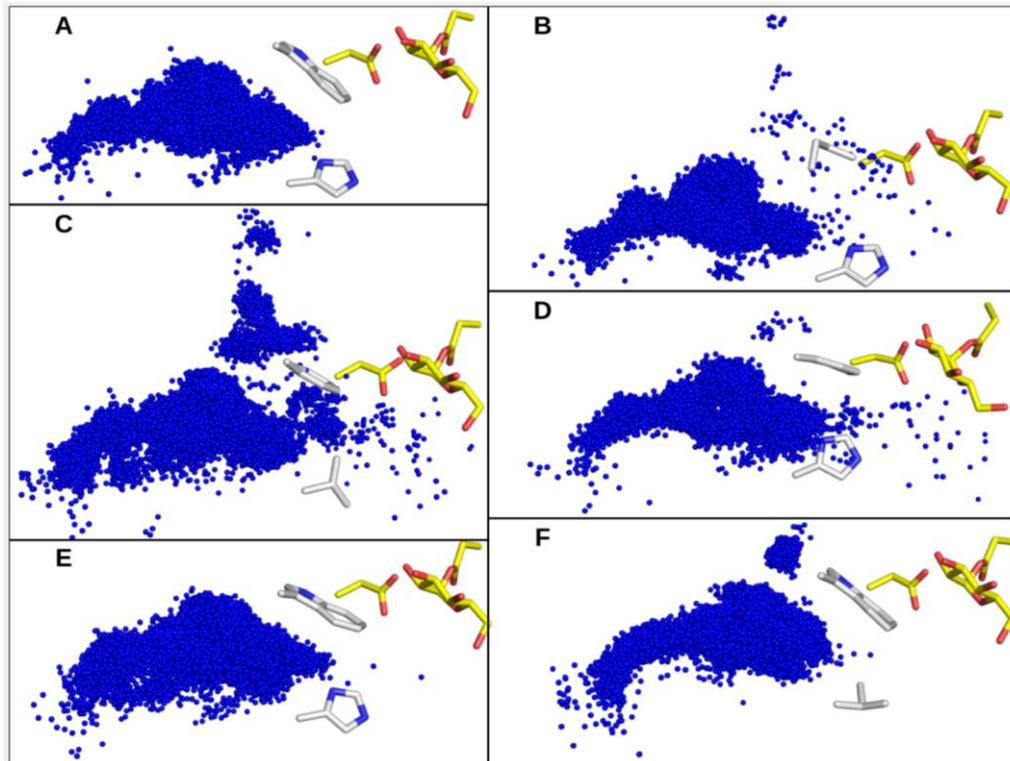


Figure 61. Trajectoire du canal principal isolé sur 500 ns de trajectoire de dynamique moléculaire chez la *Ttβgly* sauvage (A) et les mutants W120I (B), W120FH178L (C), W120F (D), C167V (E) et H178L (F).

Les résidus catalytiques sont représentés en bâtonnets jaunes. Les résidus 120 et 178 sont illustrés en bâtonnets gris.

La suppression du noyau indole en position 120 par une chaîne latérale moins encombrante provoque chez les mutants un relargage de molécules d'eau du canal principal vers le sous-site -1. Cette observation confirme donc que l'encombrement des chaînes latérales des résidus W120 et H178 semble contraindre la progression des molécules d'eau internes au canal 1 vers le sous-site -1.

6.5.3 Comparaison aux précédents résultats publiés

L'utilisation d'un protocole de simulation alternatif à celui choisi par Teze et coll. n'a permis de reproduire qu'une partie des résultats précédemment obtenus. En dépit de la variation d'un grand nombre de paramètres entre les deux méthodologies, le choix d'un champ de force distinct (CHARMM27 vs. AMBER99SB-ildn) pourrait justifier les différences observées entre les deux simulations. Une comparaison antérieure des champs de force AMBER99SB et AMBER99SB-ildn a en effet montré qu'une variation de quelques paramètres (constantes de torsion d'angles dièdres) suffisait à influencer de manière significative la stabilité et la flexibilité de l'architecture protéique. Il pourrait également être possible que le champ de force AMBER99SB-ildn stabilise plus fortement la structure de Tt β gly en comparaison à CHARMM27. Comme démontré précédemment, cet effet pourrait se traduire par une diminution de la flexibilité moyenne du squelette peptidique et des chaînes latérales. Evoluant au sein d'une structure protéique potentiellement plus rigide sous le champ de force AMBER99SB-ildn, les molécules d'eau internes à la protéine pourraient donc être susceptibles d'explorer une trajectoire plus restreinte en comparaison à celle générée sous le champ de force CHARMM27. Ceci expliquerait alors l'absence de molécules d'eau reliant la sortie du canal principal au sous-site -1. Une comparaison détaillée des champs de force AMBER99SB-ildn et CHARMM27 nécessite cependant d'être réalisée afin de pouvoir confirmer cette hypothèse.

Le sous-site -1 étant fortement accessible au solvant, il est également probable que le temps de résidence des molécules d'eau situées en sortie du canal principal à l'interface avec le sous-site -1 soit inférieur à 100 ps à la température de simulation utilisée (350 K). De précédentes études statistiques ont en effet montré que le temps de résidence des molécules d'eau localisées au sein de la première couche d'hydratation des protéines semblait se situer entre la dizaine et la centaine de picosecondes à 300 K¹²⁴. Il est donc fortement possible que les molécules d'eau quittant le canal en direction du sous-site -1 puissent s'échapper dans le milieu en quelques dizaines de picosecondes à 350 K. Par ailleurs, une mesure des distances parcourues par l'ensemble des molécules d'eau internes à l'enzyme sauvage d'un cliché à l'autre sur l'intégralité de sa trajectoire (500 ns) montre que certaines molécules d'eau internes sont capables de

parcourir des distances de l'ordre de 5 à 7 Å en 100 ps (données non montrées). Il est donc possible que l'intervalle d'échantillonnage des coordonnées atomiques au cours de la simulation (100 ps) puisse être trop large pour pouvoir suivre les déplacements rapides de l'eau interne en sortie du canal 1 jusqu'au sein du sous-site -1. La trajectoire générée par Teze et coll. avait d'ailleurs été échantillonnée tous les 10 ps afin de décrire plus finement les déplacements atomiques de l'eau. Bien que la trajectoire présentée ici soit plus longue (500 ns), un échantillonnage plus fréquent des coordonnées atomiques (au détriment du temps d'analyse et de l'espace de stockage sur disque) au cours de la simulation aurait peut être permis de détecter les molécules d'eau « manquantes » à la sortie du canal principal débouchant sur le sous-site -1 et ainsi de confirmer les résultats précédemment publiés. Cette hypothèse reste cependant à être vérifiée. Comme évoqué plus haut, il est également probable que le champ de force utilisé puisse avoir un effet non négligeable sur la dynamique de la protéine et de ses interactions avec les molécules d'eau internes.

Les résultats obtenus ne permettent donc pas de confirmer le caractère fonctionnel du canal principal puisque celui-ci ne semble fournir aucune molécule d'eau au sous-site -1 chez l'enzyme sauvage (*Figure 61, A*) malgré la réalisation d'une simulation dix fois plus longue que celle effectuée par Teze et coll.¹¹⁷. Parmi l'ensemble des canaux d'eau isolés sur la trajectoire, il est probable que tous ne jouent pas un rôle dans l'activité de Ttβgly. Parce qu'ils débouchent sur le sous-site -1, un potentiel rôle fonctionnel des canaux 2, 4, 5, et 6 dans l'approvisionnement de l'eau catalytique au sous-site -1 chez la Ttβgly sauvage a précédemment été supposé. En vue de le confirmer, il est donc nécessaire d'identifier les molécules d'eau potentiellement catalytiques et de caractériser leur trajectoire à l'intérieur de la protéine afin de déterminer si celles-ci auraient pu être acheminées jusqu'au point d'attaque nucléophile par au moins l'un de ces canaux.

6.5.4 Identification de l'origine des molécules d'eau catalytiques

Les molécules d'eau considérées comme potentiellement catalytiques sont celles qui se situent à environ 3,7-3,8 Å du carbone anomérique du sucre donneur et entre 2,5 et 2,8 Å d'un des deux oxygènes du carboxylate de l'acide-base catalytique^{225,226}.

L'identification de ces molécules au point d'attaque nucléophile ainsi que de leur trajectoire chez la Tt β gly sauvage montre qu'aucune d'entre elles n'a pu être identifiée dans le canal 1 correspondant à celui caractérisé par Teze et coll. (*Figure 62*). A l'exception d'une seule molécule d'eau dans le cas mutant W120FH178L, aucune des molécules d'eau libérées au sein du sous-site -1 ne parvient cependant au niveau du point d'attaque nucléophile.

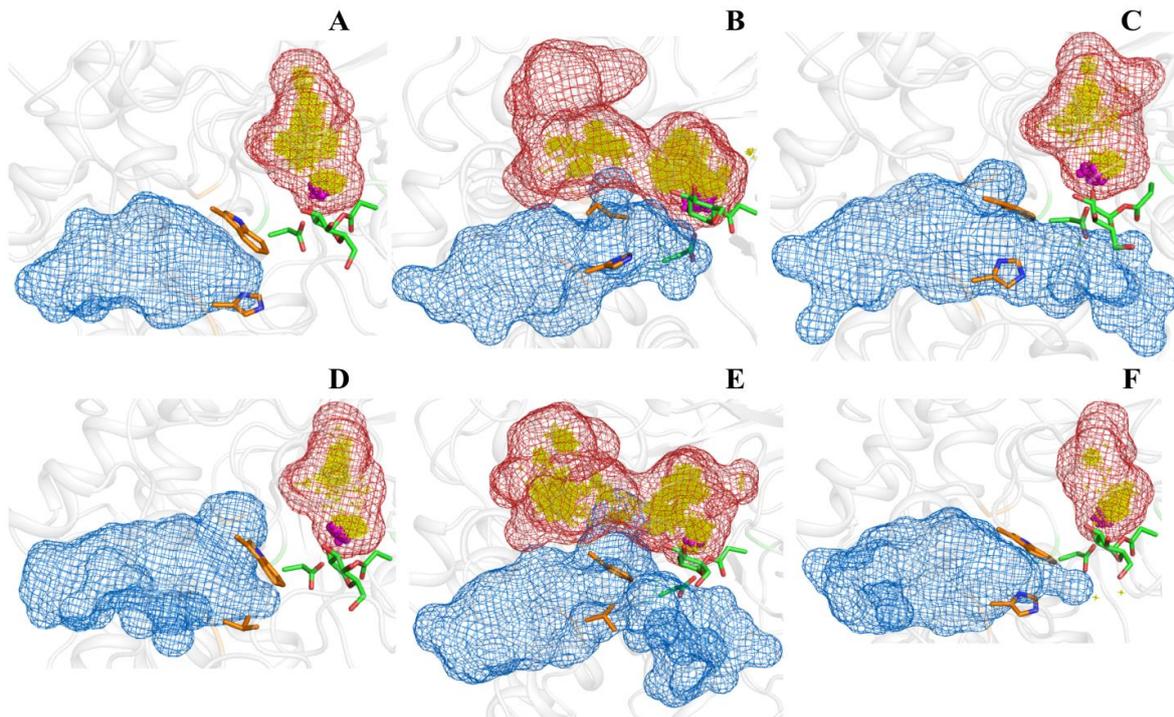


Figure 62. Localisation des molécules d'eau potentiellement catalytiques.

Les atomes d'oxygène des molécules d'eau situées au point d'attaque nucléophile sont représentés sous la forme de sphères violettes. Les positions explorées par ces molécules d'eau au sein du canal 4 sont colorées en jaune. Les trajectoires des canaux 1 et 4 sont respectivement représentées par un grillage bleu et rouge chez la Tt β gly sauvage (A) et les mutants W120I (B), W120F (C), H178L (D), W120FH178L (E) et C167V (F). Les résidus catalytiques sont représentés en bâtonnets verts.

Parmi les autres canaux isolés, un seul d'entre eux (le canal 4) s'avère en revanche impliqué dans leur acheminement au sous-site -1. Plus de 20% de la totalité des molécules d'eau catalytiques accèdent au moins une fois au sous-site -1 en empruntant ce canal chez la Tt β gly sauvage (*Figure 64, A*).

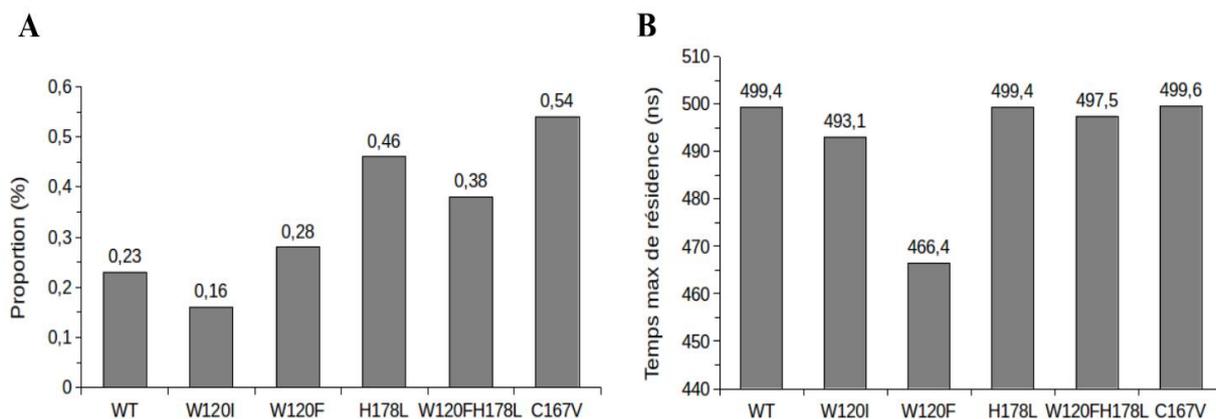


Figure 63. A : Proportion de molécules d'eau issues du canal 4 parmi l'ensemble des molécules d'eau (internes) échantillonnées au point d'attaque nucléophile. B : Temps de résidence maximal de ces molécules au sein du canal 4.

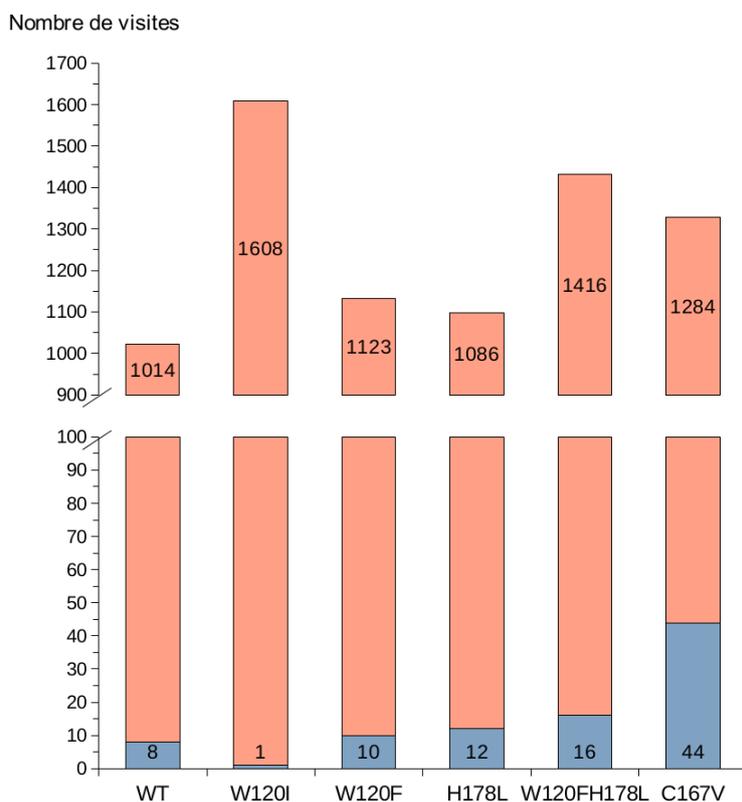


Figure 64. Nombre de visite(s) des molécules d'eau issues du canal 4 (bleu) ou du milieu (rouge) au point d'attaque nucléophile.

Le nombre de visite au point d'attaque nucléophile des molécules d'eau catalytiques du canal 4 reste cependant très faible (moins de 1%) en comparaison à celles provenant du milieu (Figure 64). La plus faible fréquence de visite des molécules d'eau catalytiques

issues de ce canal s'explique aisément de par leur proportion minoritaire au point d'attaque par rapport à celles provenant du milieu. Ce résultat est également favorisé par une résidence importante des molécules d'eau au sein de ce canal avant de rejoindre le point d'attaque nucléophile. L'identification de la trajectoire des molécules d'eau catalytiques issues du canal 4 montre par ailleurs que certaines d'entre elles sont capables de parcourir ce canal sur toute sa longueur en quelques dizaines de nanosecondes avant d'accéder au sous-site -1 (*Figure 65*). Ce canal est par ailleurs bordé de résidus extrêmement conservés dont la mutagenèse a systématiquement contribué à fortement diminuer les performances catalytiques de Ttβgly sans pour autant augmenter significativement le rendement de transglycosylation. Le canal 4 pourrait donc, pour toutes ces raisons, jouer un rôle important dans la catalyse.

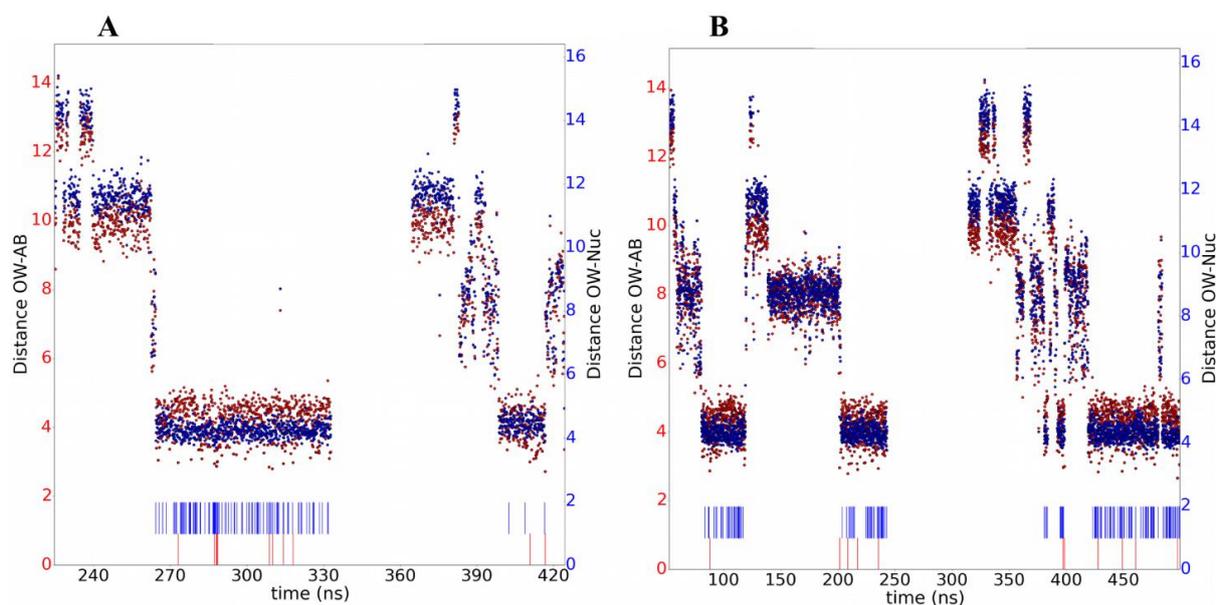


Figure 65. Trajectoire de deux molécules d'eau (A : OW 481 et B : OW 532) empruntant le canal 4 pour accéder au point d'attaque nucléophile chez la Ttβgly sauvage.

La distance entre l'atome d'oxygène de la molécule d'eau et l'un des deux oxygènes du groupement carboxylate du résidu acide-base (OW-AB) ou du résidu nucléophile catalytique (OW-Nuc) est symbolisée par un point rouge ou bleu, respectivement. Une barre verticale rouge/bleue est ajoutée lorsque l'atome d'oxygène de la molécule d'eau se trouve à une distance inférieure ou égale à 3.5/4.0 Å des oxygènes du groupement carboxylate du résidu acide-base/nucléophile catalytique, respectivement.

Dans le cas des mutants augmentant le plus fortement la transglycosylation (W120FH178L, W120I, C167V), une accessibilité plus forte de l'eau (issue du canal et du milieu) au point d'attaque nucléophile peut être observée chez tous les mutants et semble de surcroît augmenter avec l'activité de transglycosylation. Ce résultat peut paraître d'autant plus contradictoire qu'une accessibilité plus importante de l'eau à ce site impliquerait logiquement une plus forte probabilité d'attaque nucléophile conduisant à l'hydrolyse du substrat.

En dépit d'une absence de corrélation évidente entre les résultats théoriques précédents et expérimentaux, la plus forte accessibilité de l'eau des canaux 1 et 4 au sous-site -1 chez les mutants W120I, W120F, W120F/H178L et C167V suggère que la dynamique l'eau interne à ces canaux semble affectée lorsque les positions 120, 167, et 178 sont mutées. Une fusion des canaux 3 et 4 en un canal unique (*Figure 66, E*) est par ailleurs observée chez le mutant W120FH178L. Ce phénomène traduit une réorganisation de la trajectoire de l'eau interne à l'enzyme et suggère ainsi un rôle probable des résidus W120 et H178 dans la modulation de la dynamique de l'eau interne aux GH1. Il est par ailleurs possible que les modifications de la dynamique de l'eau interne au sein des canaux desservant le sous-site -1 puissent influencer l'équilibre de la balance H/T. Plusieurs analyses comparées de la dynamique des molécules d'eau internes aux canaux d'eau connectant le milieu au sous-site -1 ont donc été menées afin de caractériser l'impact des mutations sur leurs propriétés dynamiques.

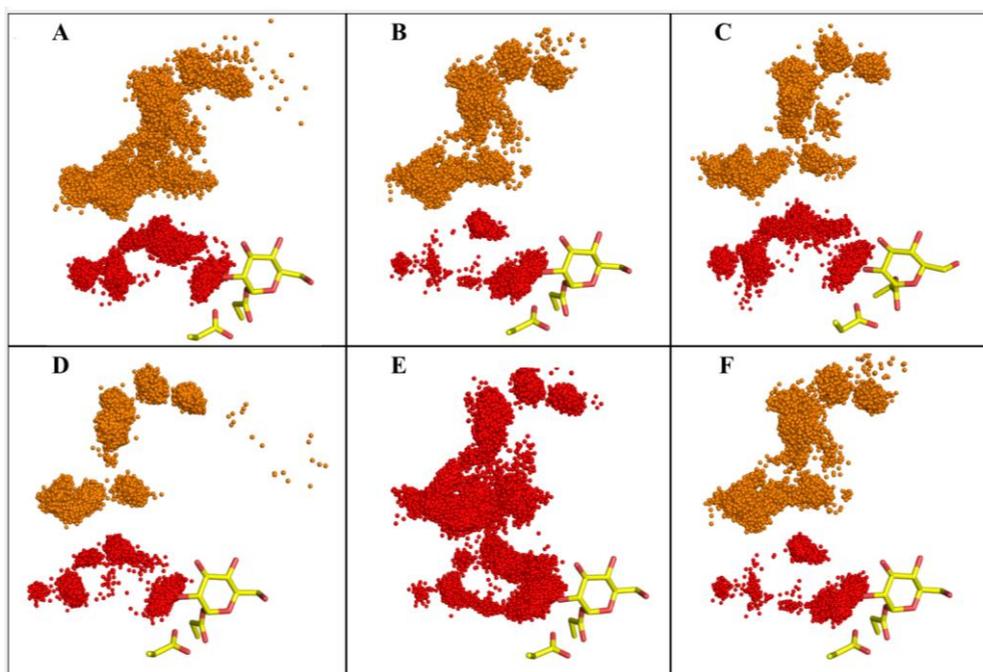


Figure 66. Trajectoire des canaux 3 et 4 isolés sur 500 ns de trajectoire de dynamique moléculaire. L'allure des canaux est présentée chez *Ttβgly* sauvage (A) et les mutants W120I (B), W120F (C), H178L (D), W120FH178L (E) et C167V (F). Les atomes de carbone des résidus catalytiques sont représentés en bâtonnets jaunes. Les atomes d'oxygène des molécules d'eau formant le canal 3 et 4 sont respectivement colorés en orange et rouge, à l'exception du mutant W120FH178L (E) pour lequel une fusion des deux canaux est observée.

6.5.5 Purges des canaux d'eau desservant le sous-site -1

Comme décrit dans le chapitre précédent, le temps de purge se rapporte au temps mis par un groupe de molécules d'eau interne à un canal pour le quitter. Ce paramètre constitue donc une mesure de la dynamique des molécules d'eau échantillonnées sur chaque cliché de simulation dans la trajectoire d'un canal d'eau donné. Une comparaison des valeurs de temps de purge calculés pour les canaux 1, 2, 4, et 5 montre que l'eau circule de manière générale plus rapidement au sein des canaux 1 et 2 par rapport aux canaux 4 et 5 qui présentent des temps de purge significativement plus élevés (Figure 67). Si des variations significatives du temps de purge de l'eau des canaux 2, 4, et 5 sont presque systématiquement observées chez les mutants par rapport à l'enzyme sauvage, aucune corrélation liant ce paramètre aux résultats expérimentaux ne peut cependant être établie pour ces canaux.

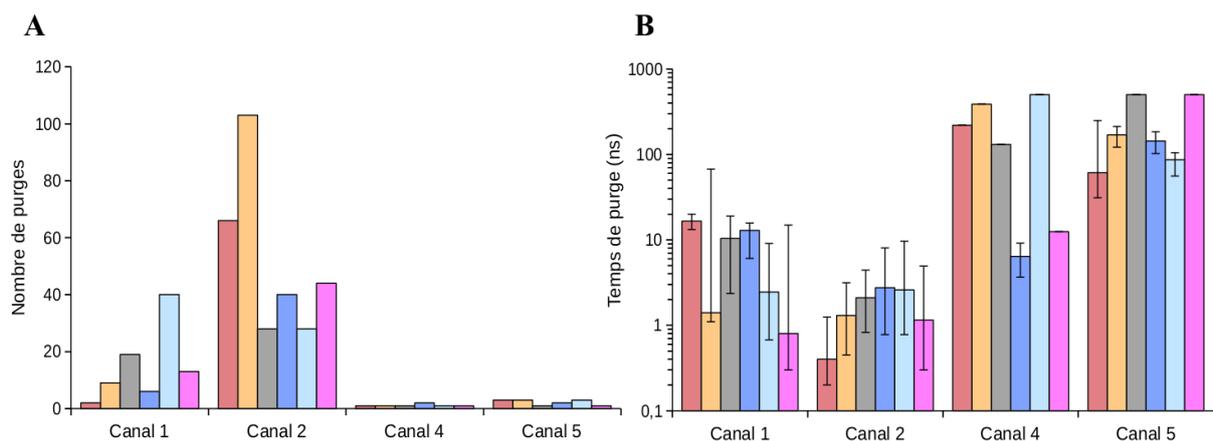


Figure 67. Temps de purge des molécules d'eau au sein des canaux desservant le sous-site -1. En rouge : Ttβgly sauvage, en cyan : le mutant W120FH178L, en jaune : le mutant W120I, en bleu : le mutant H178L, en gris : le mutant W120F et en magenta : le mutant C167V. En B : les bâtons affichent les valeurs médianes et les barres inférieure et supérieure indiquent respectivement les 1^{er} et 3^{ème} quartiles.

Si le temps de purge médian des molécules d'eau du canal 1 semble décroître suivant l'augmentation de l'activité de transglycosylation, la forte dispersion des valeurs de temps de purge chez certains mutants ne permet pas de dire si ce résultat est bien significatif. Une diminution du temps de purge des molécules d'eau du canal 1 semble cependant logique en considérant la nature des substitutions réalisées aux positions 120 et 178. En effet, les résultats précédents ont montré que l'introduction de chaînes latérales plus courtes sur ces positions provoquait une sortie des molécules d'eau vers le sous-site -1 tandis que celles-ci semblaient bloquées par l'encombrement des chaînes latérales des résidus W120 et H178 chez la Ttβgly sauvage. Il paraît donc cohérent qu'un relargage de molécules d'eau vers le sous-site -1 (ou le milieu) généré par l'ouverture de la sortie du canal 1 puisse avoir lieu et provoquer ainsi une accélération de la dynamique de son eau interne. L'augmentation de la fréquence de purge de l'eau du canal 1 également observée chez le mutant C167V pourrait par ailleurs s'expliquer en partie par l'influence de la valine en position 167 sur la dynamique de la chaîne latérale du résidu H178. La sortie de molécules d'eau du canal vers le sous-site -1 ayant par ailleurs pu être confirmée précédemment (*Figure 60*) chez ce mutant, il est donc probable que l'altération de la mobilité du résidu H178 puisse en être responsable.

6.5.6 Résidence de l'eau des canaux desservant le sous-site -1

Une autre méthode permettant de quantifier la dynamique des molécules d'eau internes à ces canaux consiste à déterminer leur temps de résidence moyen. Ce temps est déterminé pour chaque canal à partir du calcul de la fonction de survie de son eau interne. L'observation des courbes de survie de l'eau interne aux canaux 2 et 4 ne montre pas de corrélation entre les temps de résidence moyens calculés et les données expérimentales issues des cinétiques. En revanche, la comparaison des courbes de survie des canaux 1 et 4 révèle deux tendances opposées : l'introduction des mutations semble diminuer significativement le temps de résidence moyen des molécules d'eau internes au canal 1 tandis que l'effet inverse se produit au sein du canal 4 (*Figure 68*). Dans le cas de ce dernier, l'augmentation du temps de résidence moyen de son eau interne pourrait être cohérente en tenant compte du rôle de ce canal dans l'approvisionnement de molécules d'eau au point d'attaque nucléophile. Une plus forte résidence des molécules d'eau au sein du canal 4 compromettrait ainsi sa capacité à fournir des molécules d'eau au point d'attaque nucléophile et du même coup l'efficacité de la réaction d'hydrolyse.

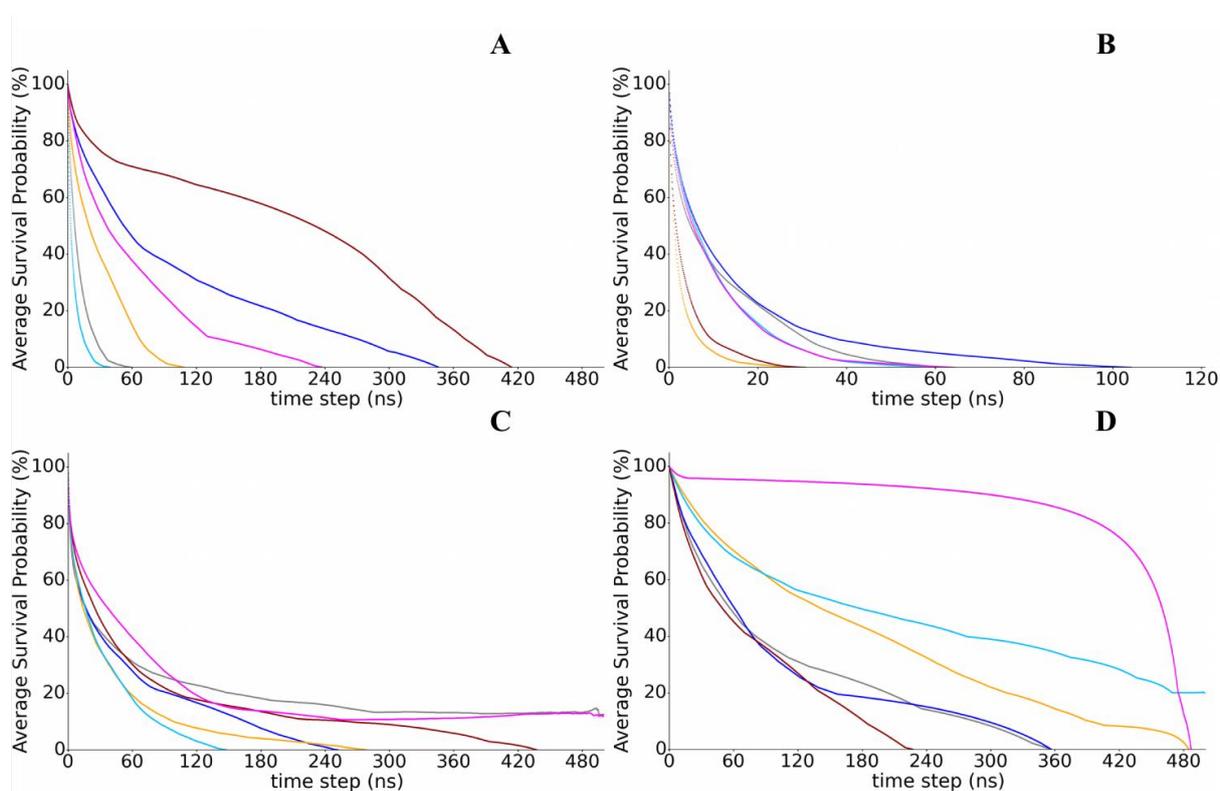


Figure 68. Courbes de survie de l'eau interne aux canaux desservant le sous-site -1.

En A : le canal 1, en B : le canal 2, en C : le canal 3, en D : le canal 4.

En rouge : Ttβgly sauvage, en cyan : le mutant W120FH178L, en jaune : le mutant W120I, en bleu : le mutant H178L, en gris : le mutant W120F et en magenta : le mutant C167V.

Si cette hypothèse s'applique bien au cas du mutant W120I, la fréquence de visites du point d'attaque nucléophile par les molécules d'eau provenant de ce canal reste néanmoins identique voire supérieure (chez les mutants W120FH178L et C167) à celle relevée pour l'enzyme sauvage chez les autres mutants. Cette observation montre qu'une augmentation du temps de résidence moyen de l'eau de ce canal n'empêche cependant pas l'accès de certaines molécules d'eau jusqu'au point d'attaque nucléophile. Il n'est donc pas exclu que la plus forte résidence de l'eau de ce canal chez ces mutants puisse influencer l'équilibre de la balance H/T *via* un autre mécanisme.

A l'inverse, la diminution générale du temps de résidence des molécules d'eau internes au canal 1 traduit une accélération de leur dynamique et confirme ainsi les résultats précédents. Comme supposé précédemment, il est probable que cet effet chez les mutants soit justifié par une accessibilité plus importante de l'eau du canal au sous-site -1 suite à la suppression des chaînes latérales encombrantes des résidus W120 et H178 ou à une modification de leur dynamique (mutant C167V). Il n'est en revanche pas aisé d'expliquer comment l'accélération de la dynamique de l'eau du canal 1 contribuerait à inverser la balance H/T. Chez les mutants, il est possible qu'un relargage important de molécules d'eau du canal principal au voisinage de l'acide-base puisse compromettre le bon positionnement de la molécule d'eau catalytique en perturbant le réseau de liaisons hydrogène qui la stabilise au point d'attaque nucléophile.

6.5.7 Conclusions

L'ensemble des résultats précédents permettent de supposer qu'une perturbation significative de la dynamique de l'eau interne des canaux 1 et 4 au sein de Ttβgly est nécessaire pour provoquer une inversion de la balance H/T. De manière générale, les mutations augmentant le plus la transglycosylation semble favoriser une plus forte accessibilité de l'eau au point d'attaque nucléophile malgré une activité hydrolytique plus faible. Cet effet, qui semble de surcroît augmenter avec l'activité de

transglycosylation, est donc en contradiction avec les résultats cinétiques. Il est donc très probable que l'effet des mutations sur l'équilibre de la balance H/T ne puisse pas s'expliquer sur la base d'une différence d'accessibilité de l'eau au point d'attaque nucléophile. Les résultats précédents ont en revanche confirmé le rôle des résidus W120 et H178 dans le contrôle de passage des molécules d'eau du canal 1 vers le sous-site -1. Chez l'enzyme sauvage, les chaînes latérales de ces résidus pourraient jouer un rôle de barrage servant à contenir les molécules d'eau du milieu empruntant le canal 1 hors du site actif. Bien qu'à première vue contre-intuitif, le relargage de molécules d'eau hors du canal 1 provoqué par l'ouverture de ce barrage moléculaire par mutagenèse semble également être nécessaire pour aboutir à une réduction significative de l'activité hydrolyse au profit de l'activité de transglycosylation.

- Chapitre IV -
Ingénierie de l'AgaD
en trans-agarase

7 Ingénierie d'AgaD en trans-agarase

7.1 Présentation du modèle d'étude

L'agarase D (AgaD) est une β -endoglycosidase de la famille 16 extraite de la bactérie marine *Zobellia galactanivorans*. Comme l'ensemble des membres de cette famille, AgaD possède une structure de type *jelly-roll* qui se caractérise par un sandwich de deux feuillets de 5 à 7 brins β accompagnés de 4 hélices α et 2 hélices 3/10 en surface ¹⁸⁷ (Figure 69). Le gène de cette enzyme est isolé et ne fait pas partie d'un cluster de gènes du type PUL (PUL : Polysaccharide Utilizing Loci). Dans sa forme native, la protéine est sécrétée grâce à un peptide de 20 acides aminés en N-terminal et un système de sécrétion propre aux bacteroidetes ¹⁸⁷. Impliqués dans l'activité d'hydrolyse, les résidus catalytiques de type acide-base (E159) et nucléophile (E154) appartiennent à un motif (EXDXXE) très conservé chez la plupart des enzymes de la famille GH16 ¹⁸⁶. AgaD catalyse l'hydrolyse de liaisons glycosidiques de type β -1,4. L'hydrolyse de polymères d'agarose conduit à la formation de produits de tailles variables finalement dégradés en néoagarohexose et néoagarotétraose.

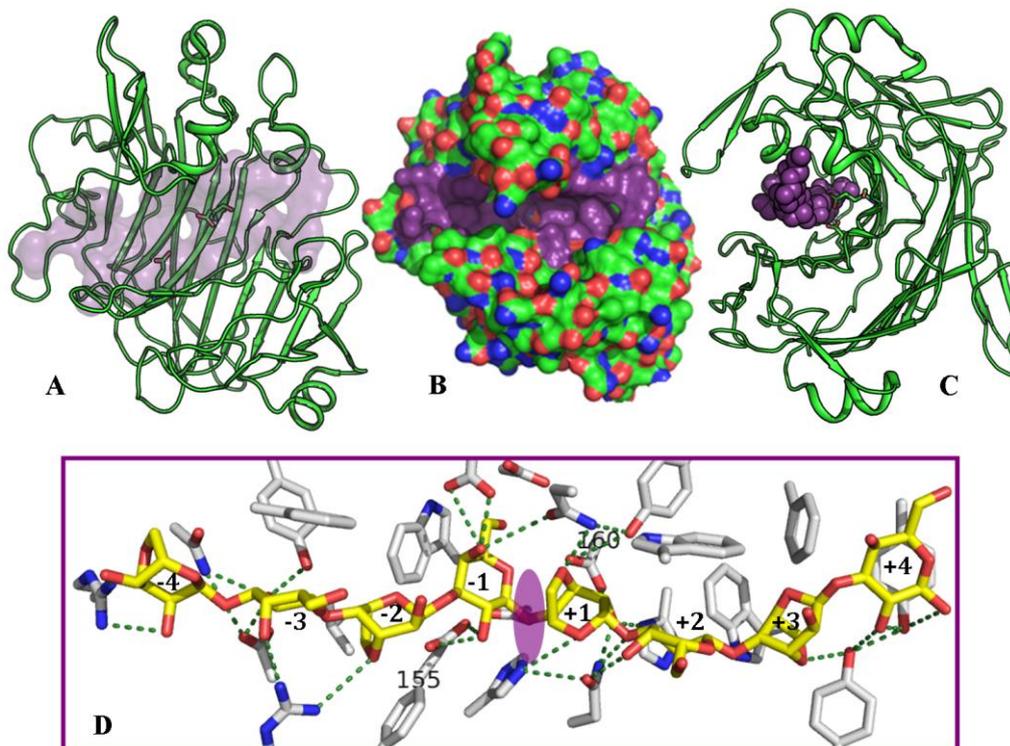


Figure 69. Structure d'AgaD et résidus conservés du site -1.

A : Structure cristallographique du domaine catalytique d'AgaD. Les acides aminés catalytiques de type acide-base et nucléophile sont indiqués par les numéros 160 et 155, respectivement. Le site de liaison au substrat est illustré en surface violette. B : Aperçu de la structure en surface. C : Vue de la structure de profil avec l'agaro-octaose (amarré dans le site de liaison par homologie avec la structure d'AgaB²²⁷, PDB 4ATF). D : Aperçu des chaînes latérales des acides aminés constitutifs du sous-site -1 et de leurs interactions à l'agaro-octaose. L'ellipse violette met en évidence le site de coupure où aura lieu l'attaque nucléophile. Les numéros au centre des cycles osidiques indique la localisation des sous-sites de liaison.

La famille 16 comporte un certain nombre de transglycosylases strictes. On y trouve notamment de nombreuses xyloglucanes-endotransglycosylases (XET)^{228,229} ainsi que les transglycosylases Crh 1/Crh 2^{69,62} respectivement impliquées dans le remodelage des parois cellulaires des plantes supérieures et des levures. A l'exception de la transglycosylase PtBgt16A de *Paecilomyces thermophila* récemment caractérisée²³⁰, aucune activité de transglycosylation n'a cependant été détectée au sein des autres GH16 répertoriées à ce jour comme actives sur les polysaccharides marins. Outre l'importance industrielle des agarases dans la dégradation de la biomasse des algues marines^{231,232}, l'obtention de transagarases artificielles par ingénierie pourrait également présenter un intérêt considérable dans la synthèse ou la réorganisation de polymères à base d'agarose.

En vue de tenter de convertir AgaD en trans-agarase, il convient tout d'abord d'identifier les résidus impliqués dans le contrôle de son activité hydrolytique. Il est par ailleurs possible que l'accès des molécules d'eau catalytiques au sous-site -1 soit contrôlé par des résidus spécifiques. Si la structure cristallographique de cette enzyme montre que le sous-site -1 est particulièrement bien exposé au solvant (bulk), il est cependant probable que la majeure partie d'entre elles ne participe pas à l'acte catalytique. Comme évoqué au chapitre premier, la mobilité importante des molécules d'eau issues du milieu peut en effet compromettre leur bon positionnement au point d'attaque nucléophile. Or, pour être hydrolytique, il est nécessaire qu'une molécule d'eau se stabilise en ce point en adoptant une géométrie d'attaque spécifique. L'internalisation préalable des molécules d'eau du milieu au sein de la structure protéique avant leur accès au sous-site -1 peut en revanche être nécessaire pour leur permettre, par l'établissement d'interactions avec les résidus environnants, de diminuer

leur énergie cinétique et augmenter ainsi leur chance d'adopter une géométrie fonctionnelle. Par analogie avec Tt β gly¹¹⁷, il est donc possible que les molécules d'eau catalytiques soient acheminées du milieu au sous-site -1 par le biais de canaux internes.

7.2 Organisation de l'eau dans la structure cristallographique d'AgaD

Au total 28 molécules d'eau internes peuvent être répertoriées au sein de la structure cristallographique d'AgaD (PDB 4ASM). Parmi celles-ci, 14 s'organisent en deux chaînes de molécules d'eau internes distinctes (*Figure 70*). Reliant la surface de la protéine au site -1, ces chaînes d'eau pourraient potentiellement jouer le rôle de canaux acheminant les molécules d'eau catalytiques jusqu'au point d'attaque nucléophile. L'une d'entre elles, nommé canal 1, est constituée de sept molécules d'eau s'étendant sur une longueur d'environ 13 Å entre la surface de la protéine et le site -1.

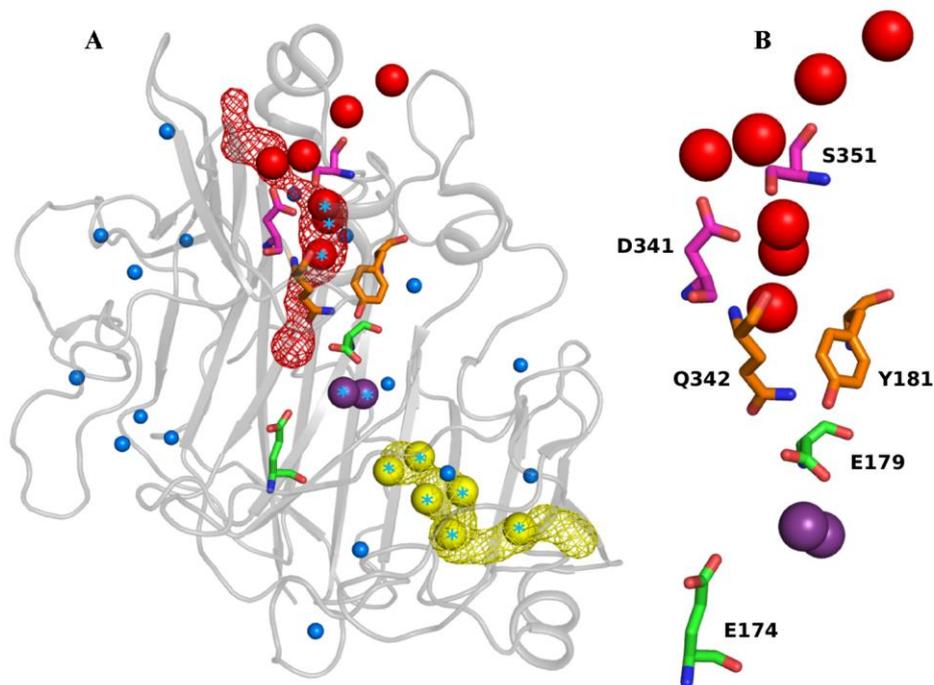


Figure 70. Canaux d'eau cristallographiques potentiels.

A: Structure cristallographique de la forme sauvage d'AgaD avec les atomes d'oxygène des molécules d'eau cristallographiques internes (sphères et étoiles bleues) à la structure protéique. Deux chaînes de molécules d'eau (sphères rouges et jaunes) connectent le milieu au site -1. Ces chaînes d'eau sont localisées au sein de tunnels représentés en grillage. La prédiction des tunnels a été réalisée par Johann Hendrickx (UFIP, Equipe de conception de protéines in-silico) à l'aide du

logiciel CAVER 3.0¹⁸³. B : Aperçu du canal d'eau potentiel étudié dans ce chapitre. Les résidus catalytiques sont illustrés en vert tandis que les paires de résidus semblant contrôler l'accès des molécules d'eau du milieu au canal puis au sous-site -1 sont représentés en bâtonnets magenta et orange. La molécule d'eau les plus proches de l'acide-base catalytique sont illustrées en sphères violettes.

La molécule d'eau du canal 1 la plus proche du sous-site -1 se situe à 7.3 Å du résidu catalytique acide-base (E179) et est séparée du site actif par les chaînes latérales des résidus Q142 et Y181. Par analogie avec les résidus W120 et H178 de Ttβgly, ces résidus semblent former un goulet d'étranglement susceptible de contrôler l'accès des molécules d'eau de ce canal vers le site -1. L'existence de ce canal est par ailleurs confirmée par les prédictions du logiciel CAVER¹⁸³ (Figure 71). Les prédictions du logiciel ont également qualifié l'existence de ce canal comme étant « la plus probable » parmi l'ensemble des autres canaux potentiels prédits au sein de la structure protéique d'AgaD.

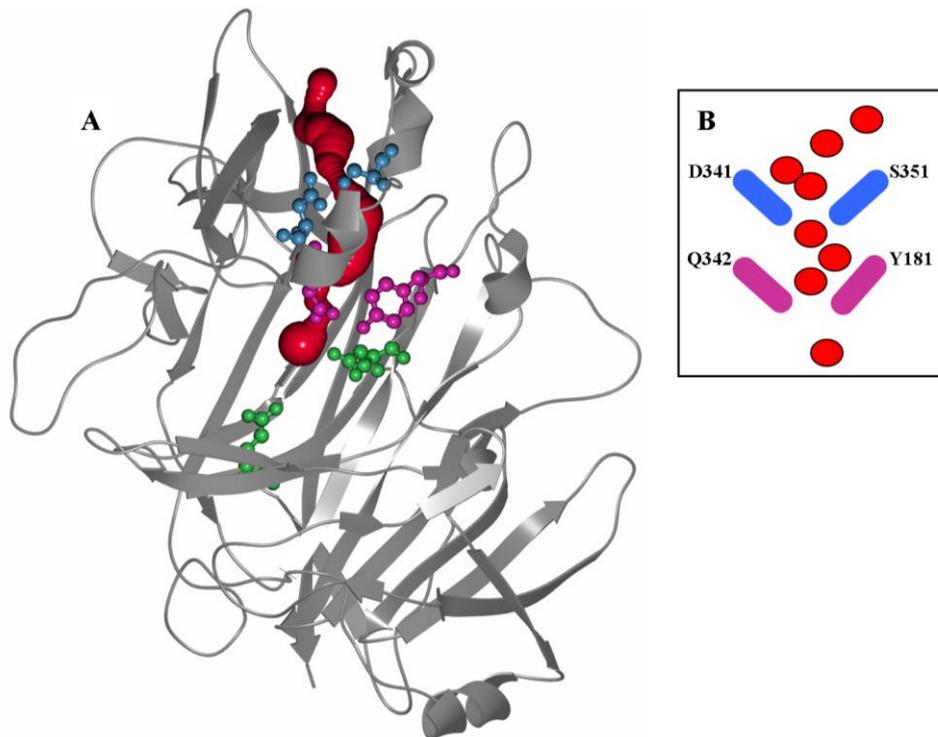


Figure 71. Canal d'eau potentiel fournissant les molécules d'eau catalytiques au site -1.

A : Prédiction de la trajectoire du canal par le logiciel CAVER 3.0. Les résidus susceptibles de contrôler la circulation des molécules d'eau dans ce canal sont illustrés en bleu et magenta. Les résidus catalytiques sont colorés en vert. B : Schéma simplifié (« pinball model ») illustrant en

comparaison le canal d'eau cristallographique correspondant (PDB 4ASM) et insistant sur le rôle des chaînes latérales des résidus potentiellement impliqués dans le contrôle de l'accès des molécules d'eau au sous-site -1 à partir du bulk.

Une seconde chaîne de sept molécules d'eau a également été identifiée à l'opposée du canal 1 par rapport au site -1. Ce canal d'eau potentiel, appelé canal 2, est séparé du site catalytique par un important goulet d'étranglement. Compte tenu de la distance séparant les deux dernières molécules d'eau les plus proches du site -1, il est peu probable qu'un passage aisé des molécules d'eau du milieu vers le site actif par ce canal soit possible. A l'inverse du canal 1, ce canal pourrait donc ne pas être impliqué dans l'approvisionnement des molécules d'eau catalytiques au site -1.

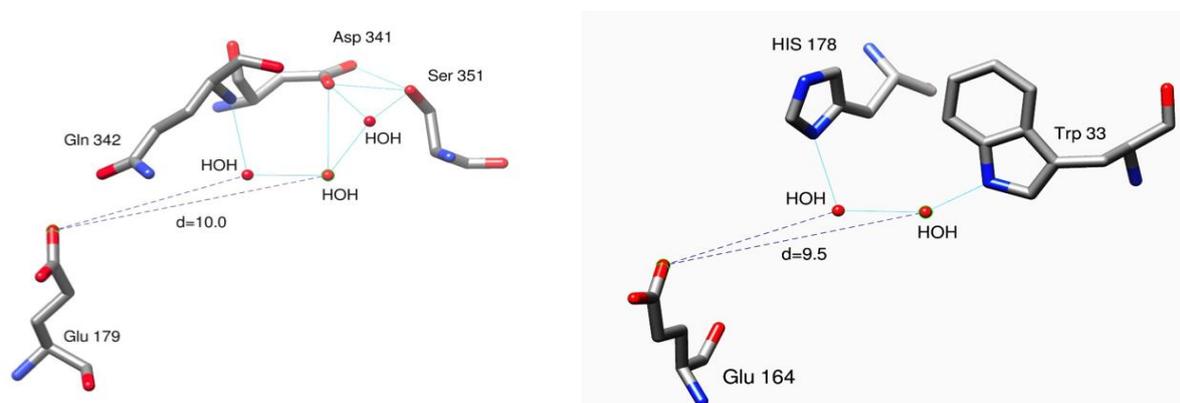


Figure 72. Disposition conservée des molécules d'eau cristallographiques en bordure du sous-site -1 chez Ttβgly (famille des GH1, PDB 1UG6) et AgaD (famille des GH16, PDB 4ASM).

Une comparaison entre ce canal et le canal 1 mis en évidence au sein de la structure cristallographique de Ttβgly (PDB 1UG6) montre par ailleurs que deux molécules d'eau cristallographiques trouvées dans ces canaux sont positionnées de manière équivalente par rapport à l'acide-base catalytique qui adopte également une géométrie très similaire (Figure 72) chez les deux enzymes (distance identique par rapport au carboxylate de l'acide-base catalytique). Cette observation est intéressante car elle suggère qu'une évolution convergente de ces deux enzymes ait pu avoir lieu pour aboutir à l'acquisition d'une activité hydrolytique commune en dépit de leur repliement très différent (α_8/β_8 pour la famille des GH1 et β -jelly-roll pour la famille des GH16).

7.3 Mutagenèse de résidus bordant le canal

Les mutations et l'analyse cinétique des variants ont été réalisées par Romain Irague, post-doctorant dans l'équipe Ingénierie Moléculaire et Glycobiologie de l'UFIP.

Le canal d'eau identifié chez AgaD est bordé de deux paires de résidus, respectivement Q342/Y181 et D341/S351 formant chacune un goulet d'étranglement aux extrémités de ce canal. En supposant que ce canal soit impliqué dans l'acheminement des molécules d'eau catalytiques, il serait donc probable que ces résidus puissent jouer un rôle dans le contrôle de leur passage du canal jusqu'au sous-site -1 et à posteriori affecter l'équilibre de la balance H/T. Une mutagenèse de ces quatre résidus a donc été entreprise afin de confirmer cette hypothèse. Les mutants correspondants ont été testés en présence de l'agarodécaose (LA-G)₅ en tant que substrat.

Dans le mécanisme catalytique d'AgaD, le résidu E174 réalise une attaque nucléophile sur ce composé pour aboutir à la formation d'un intermédiaire covalent glycosyl-enzyme avec départ d'un groupe partant de type agarotétraosyle ou agarohexaosyle. Dans une seconde étape, cet intermédiaire est quasi-exclusivement hydrolysé en néoagarotétraose (LA-G)₂ et néoagarohexaose (LA-G)₃ (Figure 73). La prédominance de l'activité d'hydrolyse implique une formation très limitée voire négligeable de produits de transglycosylation.

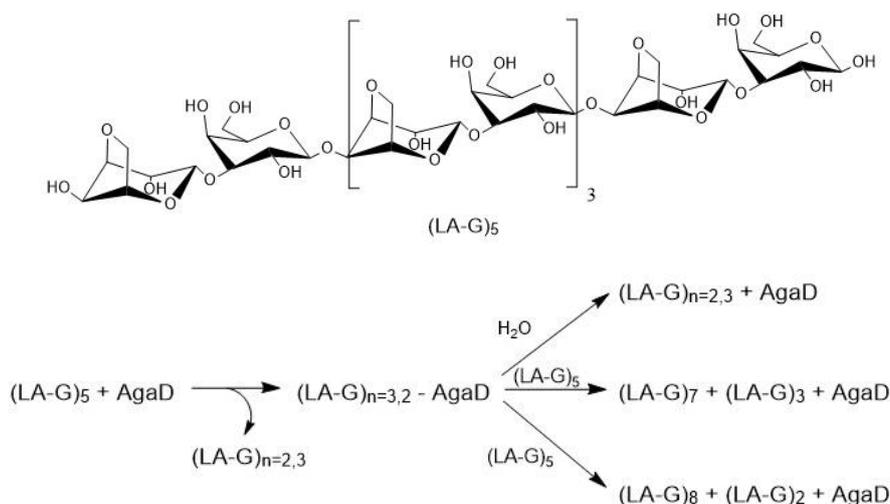


Figure 73. Mécanisme réactionnel de l'AgaD sauvage.

Le substrat naturel d'AgaD utilisé dans ces travaux est l'agarodécaose ou (LA-G)₅. Le mécanisme réactionnel de l'enzyme est présenté au dessous et indique les principaux produits formés.

La séparation du mélange réactionnel a montré que l'AgaD sauvage présente la capacité de former, en plus des produits d'hydrolyse (LA-G)₃ et (LA-G)₂, un produit de transglycosylation (LA-G)₇ au début de la réaction (*Tableau 5*). Jusqu'alors non mentionnés dans les précédents travaux publiés ¹⁸⁷, ces produits de transglycosylation sont formés à partir du néoagarodécaose qui joue à la fois le rôle de donneur et d'accepteur. Peu stables, ces produits subissent cependant une hydrolyse secondaire sur une incubation longue. Les données cinétiques confirment bien que la réaction majoritairement catalysée par l'AgaD sauvage est l'hydrolyse avec un ratio H/T (vitesse initiales) d'environ 4.24.

Vitesse initiales de formation des produits			Sommes et ratio		
(LA-G) ₂	(LA-G) ₃	(LA-G) ₇	V _H	V _T	V _H /V _T
0.173	0.179	0.037	0.157	0.037	4.24
± 0.005	± 0.019	± 0.004	± 0.028	± 0.004	

Tableau 5. Cinétique à l'équilibre d'AgaD en présence de (LA-G)₅.

La réaction a été catalysée par l'AgaD sauvage à 25°C en présence d'agarodécaose (10 mg/mL) comme substrat de la réaction. Les vitesses initiales d'hydrolyse (V_H) et de transglycosylation (V_T) sont respectivement obtenues via la somme des vitesses initiales de formation des produits (LA-G)₂ + (LA-G)₃ (pour l'hydrolyse) et (LA-G)₇ (pour la transglycosylation).

7.3.1 Mutagenèse des résidus Q342 et Y181

La mutagenèse des résidus Q342 et Y181 formant le goulet d'étranglement situé à l'interface avec le sous-site -1 a tout d'abord été effectuée. Au sein de la structure cristallographique, la chaîne latérale du résidu Y181 forme une liaison hydrogène avec celle du résidu acide-base catalytique (E179) et pourrait ainsi être impliqué dans sa stabilisation. En vue de le vérifier, le résidu Y181 a été remplacé par une phénylalanine. Le mutant correspondant s'est révélé être très peu actif (*Figure 74*).

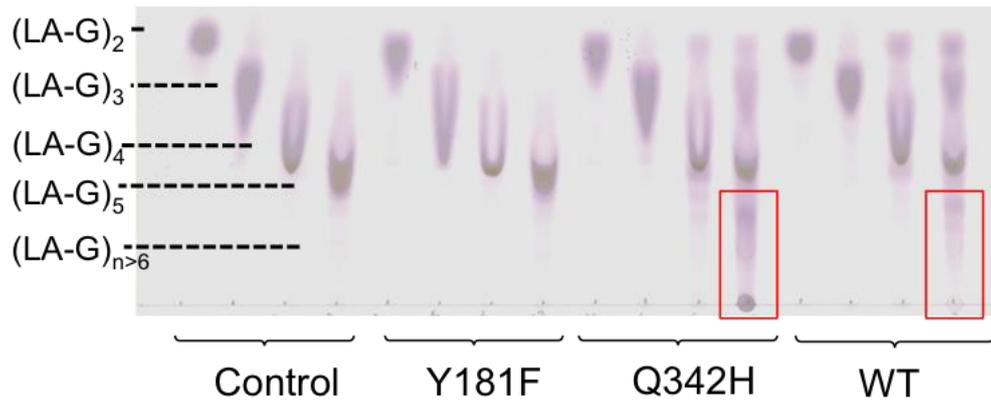


Figure 74. Test d'activité d'AgaD en présence d'agarodécaose et d'agarooctaose.

Seuls des produits d'hydrolyse sont obtenus lorsqu'AgaD est incubé en présence d'agarooctaose (LA-G)₄. En présence d'agarodécaose (LA-G)₅, l'enzyme forme un mélange de produits d'hydrolyse (LA-G)₂ et (LA-G)₃ ainsi que divers produits de transglycosylation (fractions encadrées de rouge).

Cette observation confirme ainsi l'hypothèse précédente : la suppression de la liaison hydrogène liée à la perte du groupement hydroxyle de la chaîne latérale de la tyrosine 181 pourrait déstabiliser celle du résidu acide-base. Compte tenu de la très faible performance catalytique du mutant Y181F, il est probable que cet effet puisse affecter la géométrie de l'état de transition précédant la formation de l'intermédiaire covalent. Afin de déterminer la substitution à réaliser sur le résidu Q342, un alignement multiple de 32 séquences d'agarases issues de la famille 16 et présentant au moins 20% d'identité de séquence avec AgaD a été réalisé. Si 91% des résidus observés à la position de l'alignement correspondant au résidu 342 se sont révélés être une glutamine, une histidine est cependant retrouvée dans environ 9% des séquences sélectionnées. La mutation Q342H diminue considérablement la vitesse de la réaction d'hydrolyse au profit celle de la réaction de transglycosylation (*Tableau 6*) qui devient alors la réaction majoritaire avec un rendement maximal de 90% (*Figure 76*).

Mutants V_i	WT	Q342H	D341L/ S351A	D341L/ S351F	D341L/ Q342H/ S351A	D341L/ Q342H/ S351F
	(LA-G) ₂	0.173 ± 0.005	0.164 ± 0.006	0.016 ± 0.001	0.011 ± 0.001	0.007 ± 0.001
(LA-G) ₃	0.179 ± 0.019	0.449 ± 0.029	0.033 ± 0.001	0.019 ± 0.001	0.023 ± 0.001	0.068 ± 0.005
(LA-G) ₇	0.037 ± 0.004	0.387 ± 0.009	0.022 ± 0.001	0.012 ± 0.001	0.021 ± 0.002	0.063 ± 0.006
(LA-G) ₈	nd	0.119 ± 0.004	0.007 ± 0.001	0.004 ± 5 10 ⁻⁴	0.005 ± 2 10 ⁻⁴	0.021 ± 0.001
(LA-G) ₉	nd	nd	nd	nd	0.002 ± 2 10 ⁻⁴	0.016 ± 0.001
V_H	0.157 ± 0.028	0.053 ± 0.048	0.010 ± 0.003	0.007 ± 0.002	0.002 ± 0.003	0.003 ± 0.013
V_T	0.037 ± 0.004	0.506 ± 0.014	0.029 ± 0.002	0.016 ± 0.002	0.029 ± 0.003	0.099 ± 0.008
V_T/V_H	0.24	9.54	2.90	2.28	14.5	33

Tableau 6. Effet des mutations sur la cinétique à l'équilibre d'AgaD en présence de (LA-G)₅.

La réaction a été catalysée par l'AgaD sauvage à 25°C en présence d'agarodécaose (10 mg/mL) comme substrat de la réaction. Les vitesses initiales d'hydrolyse (V_H) est la somme des vitesses initiales de formation des produits (LA-G)₂ + (LA-G)₃. La vitesse initiale de transglycosylation (V_T) est approximativement égale à la vitesse initiale de formation de (LA-G)₇. Ces vitesses sont exprimées en $\mu\text{mole}/\text{min}/\text{mg}$.

7.3.2 Mutagenèse des résidus D341 et S351

Le second goulet d'étranglement est formé par les résidus D341 et S351 à l'extrémité du canal débouchant sur la surface de la protéine. Comme suggéré précédemment, ces résidus pourraient contrôler l'accès de l'eau du milieu à l'intérieur du canal. L'analyse de l'alignement multiple évoqué plus haut a montré que ces deux résidus ne sont pas particulièrement bien conservés au sein de la famille des agarases, bien qu'au moins une des deux positions soit en général occupée par un résidu polaire. Etant donné qu'aucune substitution ne peut être clairement déterminée à partir de l'alignement, les résidus D341 et S351 ont été remplacés par deux résidus apolaires dans le but de perturber le réseau de liaisons hydrogènes connectant les molécules d'eau internes au canal en ce point.

Le choix des substitutions à réaliser s'est basé sur les prédictions du programme Rosetta²³³. Parmi toutes les paires d'acides aminés pouvant être envisagées, deux paires

de substitutions possibles ont été choisies sur la base de leur forte compatibilité structurale avec l'environnement proche des positions 341 et 351 au sein de la structure d'AgaD et de leur différence d'encombrement. Les double-mutants D341L/S351A et D341L/S351F ont ainsi été construits, exprimés, et testés. Il est intéressant de noter que ces deux mutants présentent une activité de transglycosylation fortement augmentée (Tableau 6) en dépit de la distance importante séparant les positions 341 et 351 du résidu acide-base catalytique ($\sim 13 \text{ \AA}$).

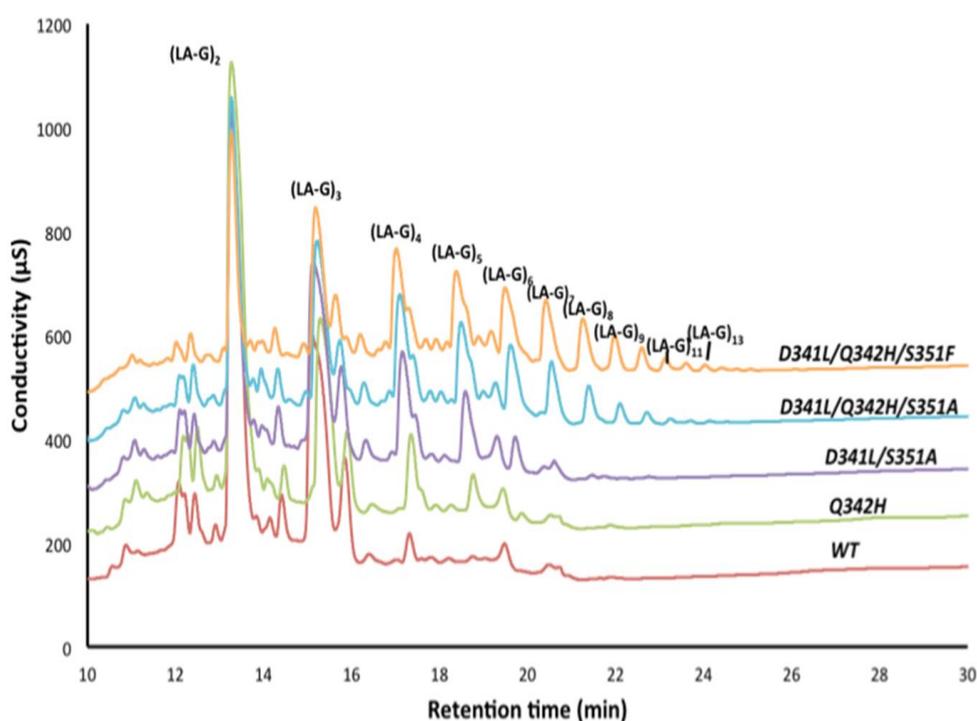


Figure 75. Séparation des produits de réaction par HPAEC-PAD après 24h d'incubation à 25°C en présence d'agarodécaose (10mg/ml). Cette expérience a été réalisée par Diane Jouanneau, Station Biologique de Roscoff, UMR-CNRS 8227.

A l'exception du substrat utilisé (agarodécaose), l'ensemble des autres pics se réfèrent à un composé formé au cours de la réaction. Les produits d'hydrolyse sont l'agarotétraose (LA-G)₂ et l'agarohexaose (LA-G)₃, tandis que les réactions de disproportionation (transglycosylation) forment de nombreux oligoagaroses de tailles variables (LA-G)_{n>6}.

Lorsque la mutation Q342H y est combinée, un effet additif peut être observé : les triple-mutants correspondants (D341L/S351A/Q342H et D341L/S351F/Q342H) présentent une activité de transglycosylation très fortement augmentée au détriment de l'hydrolyse avec un ratio H/T < 0.03. Lorsqu'il sont incubés sur plusieurs heures en

présence de (LA-G)₅, ces triple-mutants sont capables de former un nombre important de produits allant du (LA-G)₂ au (LA-G)₃ (Figure 75). La formation de produits de transglycosylation de hauts poids moléculaires par réactions de disproportionation est d'autant plus efficace que l'activité hydrolytique est faible limitant l'hydrolyse secondaire des produits de transglycosylation. Il est intéressant de constater que ces mutations combinées diminuent plus fortement l'hydrolyse (réduite d'un facteur 50 par rapport à l'AgaD sauvage) qu'elles n'augmentent la transglycosylation (augmentée d'un facteur 2.7 seulement). A l'inverse, le mutant Q342H affiche la plus forte augmentation d'activité de transglycosylation (Figure 76). En revanche, moins de produits de disproportionation sont formés en raison de la persistance d'une activité hydrolytique plus importante que celles mesurée chez les triple-mutants.

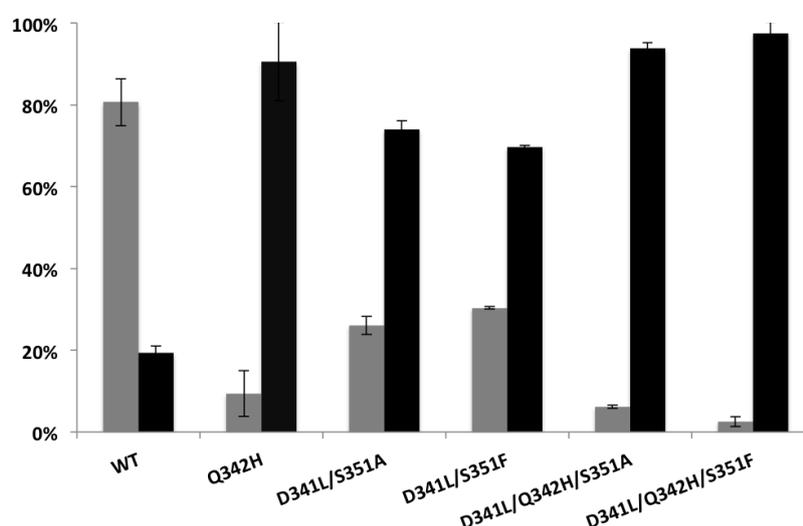


Figure 76. Balance des vitesses initiales d'hydrolyse (V_H) et de transglycosylation (V_T).

Ces vitesses initiales « résultantes » sont obtenues à partir de la somme des vitesses initiales de formation de chaque produit issu d'une même réaction donnée (hydrolyse ou transglycosylation). En gris clair et noir, respectivement : le rapport $V_H/(V_H + V_T)$ et $V_T/(V_H + V_T)$.

7.3.3 Conclusions

L'effet remarquable de ces mutations sur l'activité d'AgaD montre que les résidus D341, S351, et Q342 sont impliqués dans le contrôle de l'équilibre de la balance H/T. L'élucidation des bases moléculaires gouvernant cet effet reste cependant difficile à expliquer. En supposant que le canal 1 fournisse les molécules d'eau catalytique au site

catalytique, il est possible que la forte diminution d'activité hydrolytique chez ces mutants puisse s'expliquer par une modification de la dynamique de l'eau interne à ce canal qui pourrait alors compromettre l'accès des molécules d'eau catalytiques au point d'attaque nucléophile. Avant de vérifier cette hypothèse, il convient tout d'abord de s'assurer que ces mutations ne modifient pas l'orientation ou la dynamique des résidus impliqués dans catalyse. Ces résidus doivent en effet être stabilisés dans une géométrie optimale en vue de stabiliser l'architecture de l'état de transition initiant la réaction hydrolytique. Les états de transition des réactions d'hydrolyse et de transglycosylation étant structurellement très proches, il paraîtrait logique qu'une perturbation de la dynamique des chaînes latérales des résidus de la machinerie catalytique affecte les deux réactions.

L'accepteur osidique intervenant dans la réaction de transglycosylation peut cependant être aisément stabilisé par des interactions avec les résidus de première et seconde couche tandis que la stabilisation de la molécule d'eau catalytique au point d'attaque nucléophile ne dépend en revanche que de son interaction avec le résidu acide-base catalytique et le cycle osidique du substrat donneur. Il est donc probable qu'une modification de la dynamique des résidus de la machinerie catalytique aurait très probablement des conséquences d'autant plus importantes sur la stabilisation de l'état de transition de la réaction d'hydrolyse que sur celui de la réaction de transglycosylation et pourrait ainsi être suffisante pour expliquer les effets observés chez les mutants. A l'exception de la position 342, les mutations introduites sur les positions 341 et 351 sont distantes de plus de 13 Å du résidu acide-base catalytique.

De nombreux travaux ont par ailleurs révélés que l'introduction de mutations sur des positions éloignées du site catalytique pouvait avoir une influence considérable sur la dynamique des résidus du site catalytique chez certaines enzymes^{234,235,236}. Il est donc possible que les mutations introduites en positions 341 et 351 puissent, à distance, affecter la dynamique de la machinerie catalytique en influençant celle des résidus alentour par effet de mouvements corrélés. Il conviendra donc de comparer dans un premier temps la dynamique de la structure protéique des mutants à celle de l'AgaD sauvage afin de s'assurer que les mutations introduites ne déstabilisent pas l'architecture globale de l'enzyme.

7.4 Analyse de la dynamique de la protéine

Contrairement au chapitre précédent, il n'a ici pas été possible de simuler l'intermédiaire glycosyl-enzyme d'AgaD. En effet, la partie glycosyle (agarotétraosyle) contient des unités d'anhydro-galactose dont la paramétrisation n'a à ce jour pas été réalisée. Par ailleurs, aucune structure d'un intermédiaire covalent d'AgaD, ni même d'un homologue de cette enzyme n'est encore disponible. Il a donc été choisi de ne simuler que la forme libre d'AgaD et de ses mutants. Les résultats des analyses des trajectoires correspondantes sont ainsi présentés ci-dessous.

7.4.1 Stabilité globale et locale du squelette peptidique

Le suivi du RMSD de la structure protéique de l'ensemble des mutants ainsi que de l'AgaD sauvage montre que celle-ci reste stable durant la simulation. En général, les mutants présentent néanmoins des valeurs de RMSD plus élevées par rapport à AgaD sauvage (*Figure 77*). Ce résultat laisse supposer que l'introduction des mutations a légèrement déstabilisé le squelette peptidique de ces derniers. Il est d'ailleurs cohérent d'observer un tel effet puisqu'il est probable que la suppression des liaisons hydrogènes formées entre les résidus D341 et S351 et entre les résidus Q342 et E179 par mutagenèse doit contribuer à déstabiliser la structure globale. L'effet déstabilisant des mutations est particulièrement prononcé sur l'ensemble des résidus situés à 10 Å autour du résidu acide-base catalytique (*Figure 77, B*). Ce résultat suggère que les mutations déstabilisent principalement la poche de liaison au substrat allant du sous-site -2 au sous-site +2. Cet effet pourrait donc avoir des conséquences sur la stabilisation du substrat au sein du site catalytique ainsi que sur celle des résidus de la machinerie catalytique.

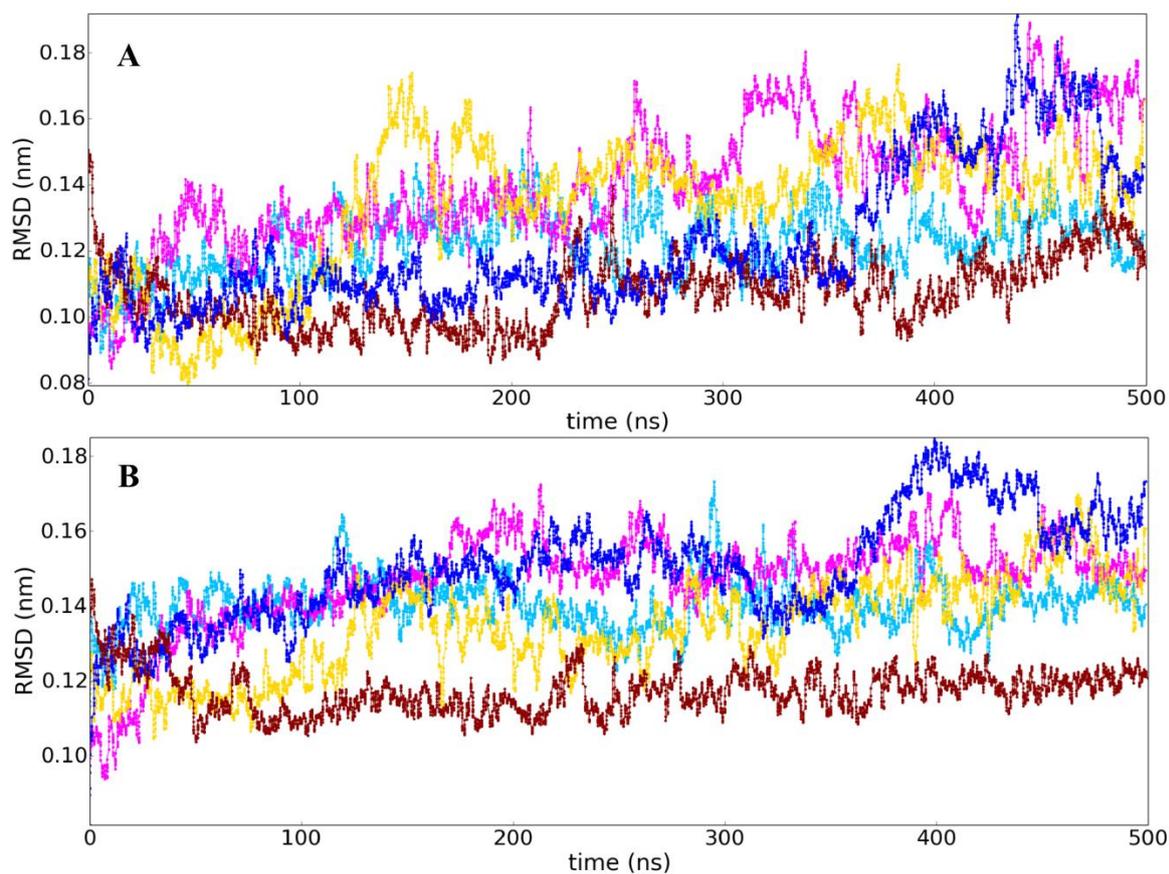


Figure 77. RMSD calculé sur le squelette peptidique (atomes $C\alpha$).

La déviation des coordonnées des atomes $C\alpha$ a été moyennée sur l'ensemble des résidus de la protéine (A) ou sur l'ensemble des résidus situés à 10 Å de l'acide-base catalytique (B). En rouge : *AgaD* sauvage, en jaune : le mutant D341L/S351A, en cyan : le mutant Q342H, en bleu : le mutant D341L/S351F/Q342H, en magenta : le mutant D341L/S351A/Q342H.

La stabilité de l'hélice- α portant les résidus D341, S351, et Q342 est également affectée par les mutations (Figure 78). Il est aussi intéressant de constater que sa mobilité est plus fortement augmentée en présence de la triple-mutation D341L/S351F/Q342H en comparaison aux autres mutants. Cet effet propre à ce mutant pourrait peut être en partie expliquer les résultats cinétiques associés.

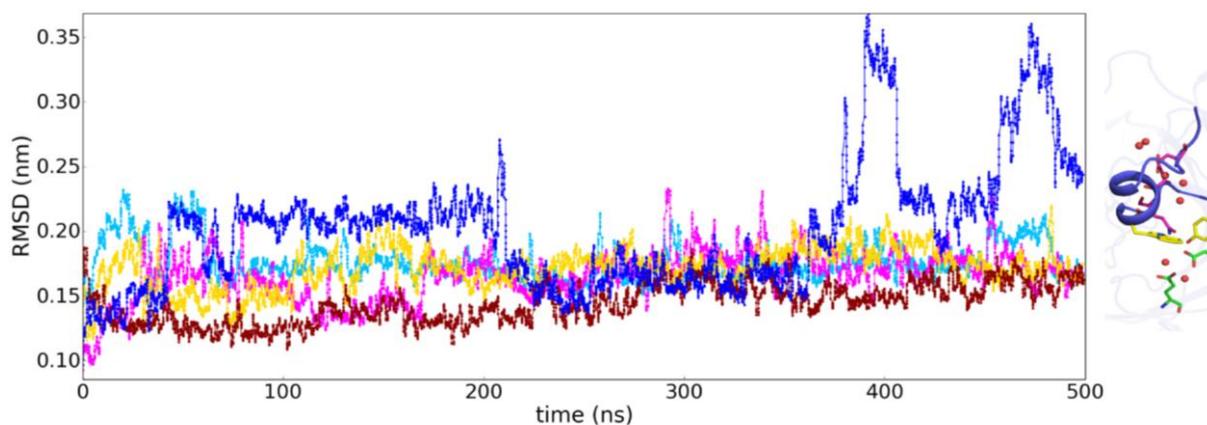


Figure 78. RMSD calculé sur le squelette peptidique de la région mutée.

La déviation des coordonnées des atomes C α a été moyennée sur l'ensemble des résidus formant et encadrant l'hélice- α portant les résidus mutés D341, S351, et Q342 (résidus 343 à 352). La région correspondante (allant du résidu 341 au 352) est illustrée à droite sur la figure. Le code couleur est le même que dans la figure 77.

Comme décrit dans le chapitre précédent, un calcul de RMSD a également été effectué sur l'ensemble des régions protéiques de chaque mutant présentant une différence d'exposition au solvant interne avec l'enzyme sauvage dans le but d'évaluer l'influence des mutagenèses sur la stabilité locale du squelette peptidique (Figure 79). Les écarts moyens de RMSD calculés sur les régions sélectionnées montrent que les changements conformationnels les plus importants sont principalement observés sur des boucles externes situées à l'opposé de la région mutée. Compte tenu de leur distance importante vis-à-vis des résidus catalytiques, il est difficile d'affirmer si ces régions non-structurées sont importantes pour la catalyse. D'autre part, la comparaison des écarts moyens de RMSD sur ces régions entre les mutants ne montre pas de corrélation avec les résultats cinétiques correspondants. Un changement conformationnel non-négligeable (écart de RMSD de l'ordre d'1 Å) du brin β portant le résidu nucléophile catalytique peut cependant être observé chez le mutant Q342H, et dans une moindre mesure chez les mutants D341L/S351A/Q342H et D341L/S351F/Q342H. Il est donc possible que la mobilité plus importante de ce brin β puisse affecter celle du squelette peptidique du résidu nucléophile catalytique et en conséquence impacter l'activité de ces mutants. La région mutée présente également un écart moyen de RMSD du même ordre de grandeur chez le mutant D341L/S351F/Q342H. Bien qu'il ne soit pas exclu qu'ils puissent influencer la catalyse, les changements conformationnels affectant ces

deux régions ne sont cependant pas observés chez tous les mutants. Il est donc probable que d'autres facteurs soient responsables de l'inversion de la balance H/T chez les mutants.

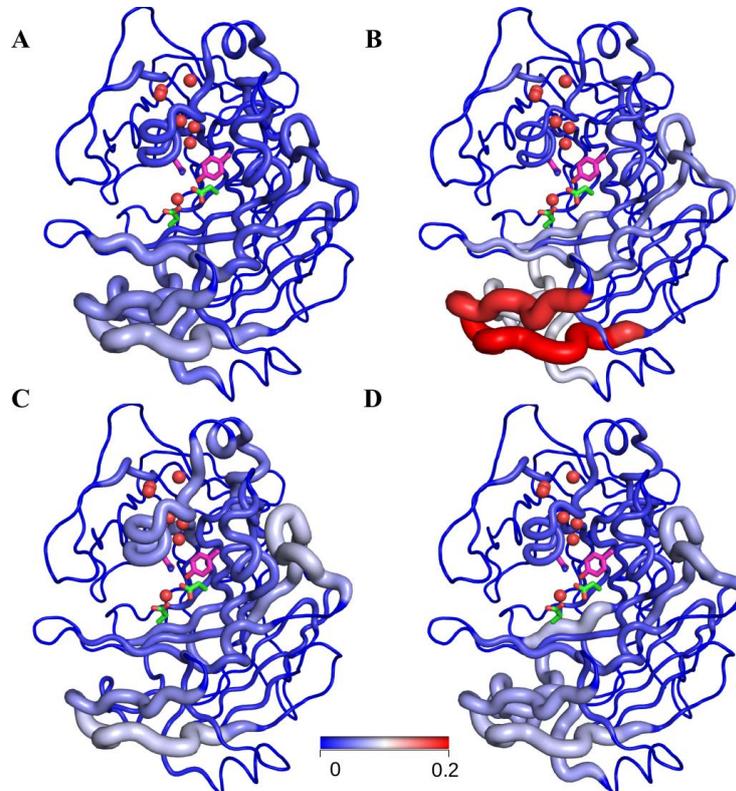


Figure 79. Projection structurale des écarts moyens de RMSD (en nm) calculés sur le squelette peptidique des régions de la protéine montrant des différences de contact avec l'eau enfouie entre l'AgaD sauvage et chaque mutant.

L'écart moyen de RMSD pour une région donnée est exprimé comme la moyenne des différences $RMSD_{mutant} - RMSD_{wt}$ calculées sur cette région à chaque cliché de trajectoire. En A : le mutant D341L/S351A, en B : le mutant D341L/S351A/Q342H, en C : le mutant D341L/S351F/Q342H, en D : le mutant Q342H.

7.4.2 Flexibilité du squelette peptidique

Le calcul des fluctuations du squelette peptidique d'AgaD sauvage et de ses mutants montre que les mutations introduites augmentent la flexibilité de certaines boucles localisées en surface de la protéine (Figure 80). La distance également importante de ces boucles par rapport au site actif rend cependant peu probable leur implication dans la

catalyse. Le squelette de l'hélice- α portant les résidus mutés subit également une augmentation significative de sa flexibilité en présence de la triple-mutation D341L/S351F/Q342H. Cette observation conforte les résultats précédents qui montraient également que cette région subissait une forte hausse de RMSD chez ce mutant. Bien que des variations de RMSF puissent être observées entre les différents mutants, il est difficile d'établir une corrélation entre les changements de flexibilité moyenne observés sur ces régions et les résultats cinétiques. De manière générale, la flexibilité du squelette peptidique de la poche catalytique n'est également pas significativement perturbée par les mutations.

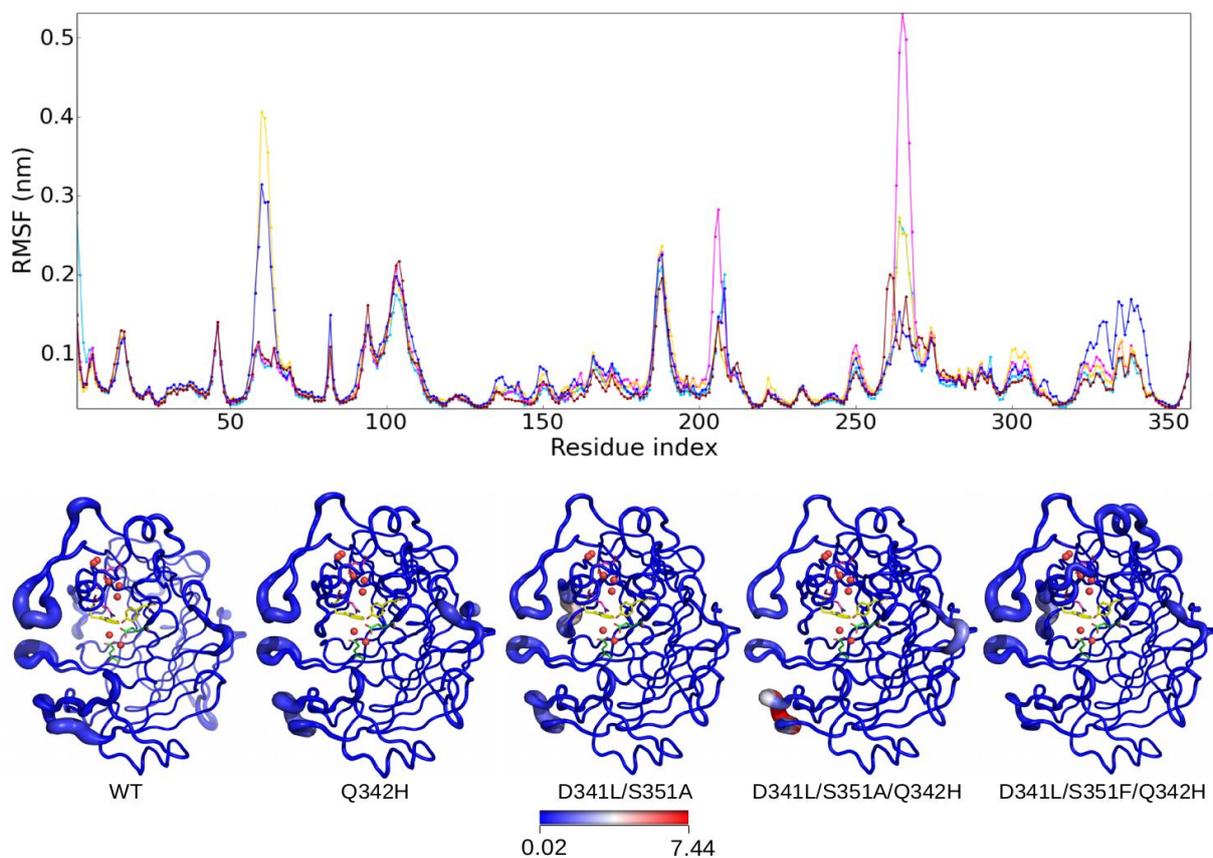


Figure 80. RMSF calculé sur le squelette peptidique et projection structurale des valeurs calculées sur les atomes du squelette peptidique.

En rouge : AgaD sauvage, en jaune : le mutant D341L/S351A, en cyan : le mutant Q342H, en bleu : le mutant D341L/S351F/Q342H, en magenta : le mutant D341L/S351A/Q342H. Les valeurs de facteur-B (en nm^2) calculées à partir des RMSF sont projetées sur le squelette peptidique de la conformation dynamique moyenne de d'AgaD déterminée sur 500 ns de trajectoire de dynamique moléculaire.

7.4.3 Flexibilité de la chaîne latérale du résidu acide-base catalytique

Un échantillonnage des angles de torsion χ_1 et χ_3 de la chaîne latérale du résidu acide-base catalytique (E179) a été réalisé afin de vérifier si les mutations introduites modifient sa dynamique (*Figure 81*). La mutation Q342H, présente chez les mutants affichant les plus fortes activités de transglycosylation, augmente la mobilité conformationnelle de la chaîne latérale de ce dernier. Cet effet est d'ailleurs marqué par l'apparition d'une nouvelle population de conformères. Ce résultat semble donc confirmer l'hypothèse du rôle probable du résidu Q342 dans la stabilisation de la chaîne latérale de l'acide-base catalytique.

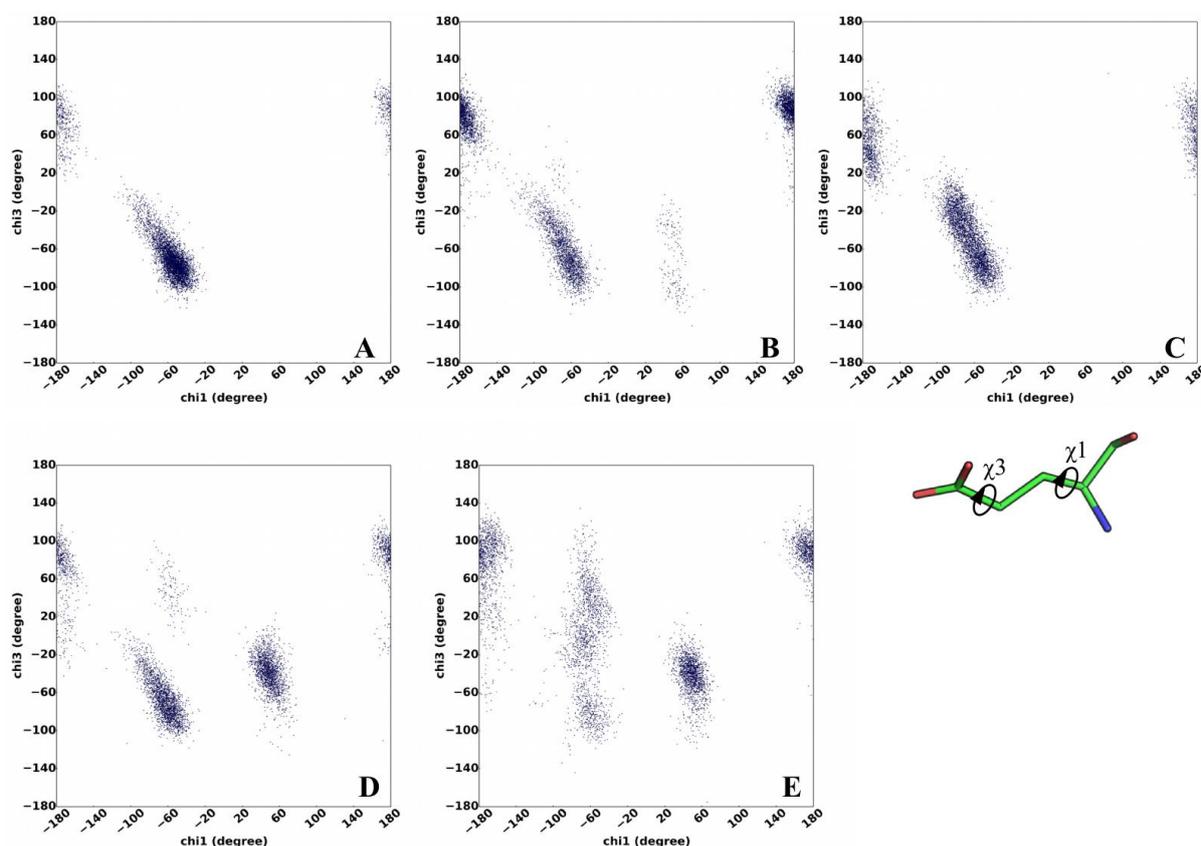


Figure 81. Angles de torsion (χ_1, χ_3) de la chaîne latérale du résidu acide-base catalytique (E179). En A: l'AgaD sauvage, en B: le mutant Q342H, en C: le mutant D341L/S351A, en D: le mutant D341L/S351A/Q342H, en E: le mutant D341L/S351F/Q342H.

La mobilité de cette dernière est également amplifiée lorsque les double-mutations D341L/S351A et D341L/S351F sont combinées à la mutation Q342H. Il est probable que l'augmentation de l'espace conformationnel exploré par la chaîne latérale du résidu

E179 chez les mutants porteurs de la mutation Q342H puisse être liée à une augmentation générale de la dynamique de la structure protéique qui déstabiliserait le résidu E179. Les précédents calculs de RMSD ont en effet montré qu'une augmentation de la mobilité de la protéine dans son ensemble est par ailleurs systématiquement observée chez les mutants porteurs de la mutation Q342H. En revanche, la double mutation D341L/S351A seule ne modifie que très faiblement la mobilité de la chaîne latérale du résidu E179. Au vu de la distance importante des résidus D341 et S351 à l'acide-base catalytique, il paraît à première vue logique que leur modification n'affecte pas significativement la mobilité de ce dernier. Compte tenu de l'augmentation importante de l'activité de transglycosylation du mutant D341L/S351A, ce résultat suggère ainsi que la mobilité de l'acide-base catalytique ne semble pas être le facteur principal à considérer pour expliquer les effets des mutations introduites chez AgaD à ce stade de la catalyse.

7.4.4 Flexibilité des chaînes latérales d'autres résidus du site -1

Essentielle à la catalyse, la stabilisation des résidus constituant la machinerie catalytique nécessite qu'ils forment des interactions avec des résidus proches (*Figure 82*). L'analyse de la trajectoire de l'AgaD sauvage a révélée que le résidu acide-base catalytique établissait des interactions polaires avec les résidus de première couche D176 et Y181. L'analyse d'un alignement multiple de 32 séquences de GH16 présentant au moins 20% d'identité de séquence avec AgaD a montré que les résidus D176 et Y181 sont respectivement retrouvés dans 100% et 93% des séquences sélectionnées.

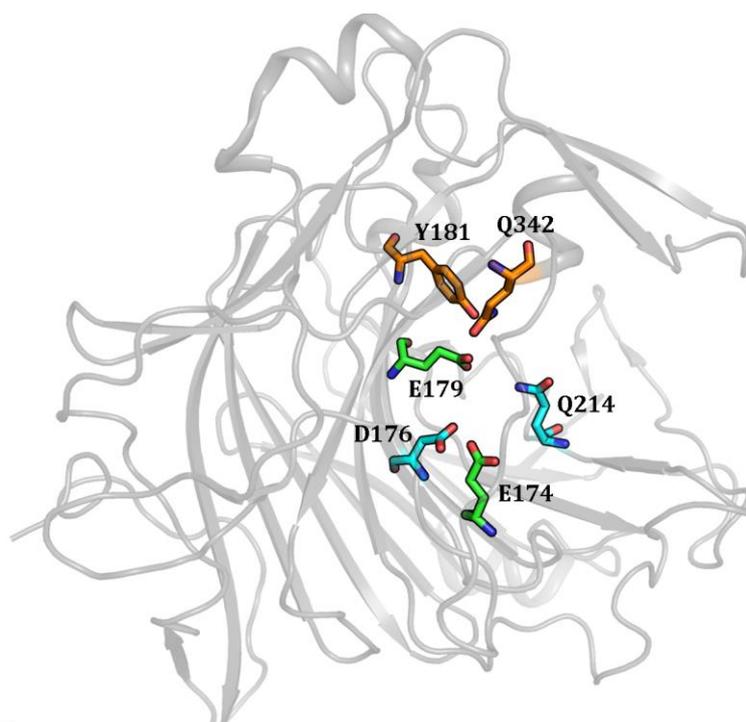


Figure 82. Localisation des résidus étudiés sur la structure cristallographique.

Les atomes de carbone des résidus catalytiques sont colorés en vert, ceux des résidus formant le goulet d'étranglement en sortie du canal 1 à l'interface avec le sous-site -1 sont colorés en orange.

Il a d'ailleurs été montré précédemment que la mutation Y181F entraînait l'inactivation de l'enzyme et que cet effet pouvait s'expliquer par la suppression de la liaison hydrogène établie entre le groupement hydroxyle de la chaîne latérale de ce résidu et celle de l'acide-base catalytique dans la structure cristallographique d'AgaD et qui pourrait être essentielle à la stabilisation de ce dernier. Il est intéressant de constater que le résidu Y181 subit des changements conformationnels analogues à l'acide-base catalytique en présence des mutations (*Figure 83*). Ce résultat laisse supposer que les chaînes latérales de ces deux résidus se déplacent de manière corrélée l'une par rapport à l'autre.

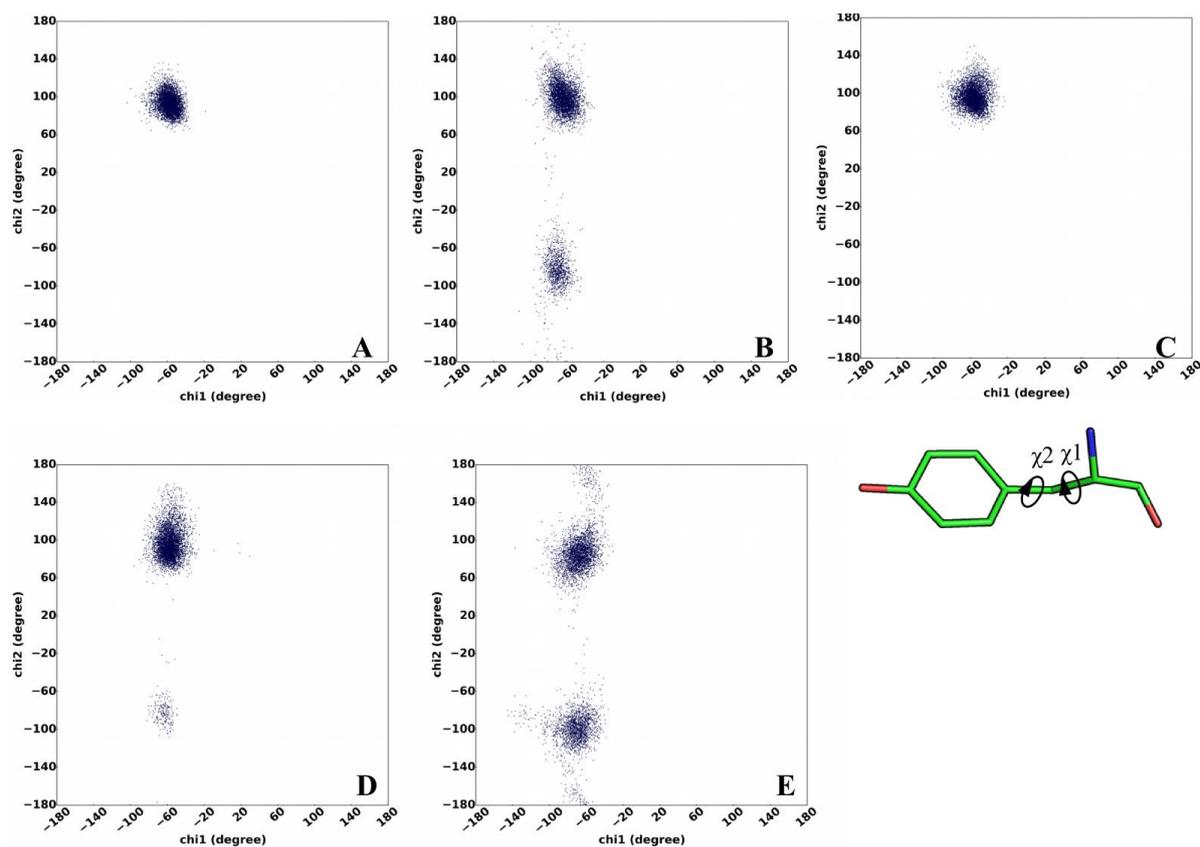


Figure 83. Angles de torsion (χ_1 , χ_2) de la chaîne latérale du résidu Y181.

En A : l'Agad sauvage, en B : le mutant Q342H, en C : le mutant D341L/S351A, en D : le mutant D341L/S351A/Q342H, en E : le mutant D341L/S351F/Q342H.

Le suivi des angles de torsion de la chaîne latérale du résidu D176 montre que la dynamique de celle-ci est plus fortement perturbée chez l'ensemble des mutants (Figure 84). L'augmentation de la dynamique de la chaîne latérale de ce résidu semble également être corrélée à l'augmentation de l'activité de transglycosylation. Au vu de l'ensemble de ces résultats et compte tenu de la forte conservation des résidus Y181 et D176 au sein des séquences d'homologues d'Agad et de leur proximité à l'acide-base catalytique, il est très probable que les résidus Y181 et D176 puisse potentiellement réguler la dynamique conformationnelle de dernier et plus largement intervenir dans la stabilisation de l'état de transition.

Deux autres résidus, Q342 et Q214, établissent également des interactions polaires avec le résidu acide-base catalytique. Si le résidu Q214 ne voit pas sa dynamique conformationnelle significativement modifiée en présence de la mutation Q342H, celle-ci est très fortement perturbée chez le double mutant D341L/S351A (Figures 85 et 86).

On peut donc conclure que la mutation Q342H atténue les effets de la double mutation D341L/S351A chez le triple mutant D341L/S351A/Q342H et stabilise en conséquence le résidu Q214. En supposant que le résidu Q214 soit plus spécifiquement impliqué dans la stabilisation de l'état de transition de la réaction d'hydrolyse que dans celui de la réaction de transglycosylation, la forte déstabilisation de la chaîne latérale de ce résidu chez le double mutant D341L/S351A pourrait alors expliquer les effets de la double mutation D341L/S351A sur l'activité d'AgaD.

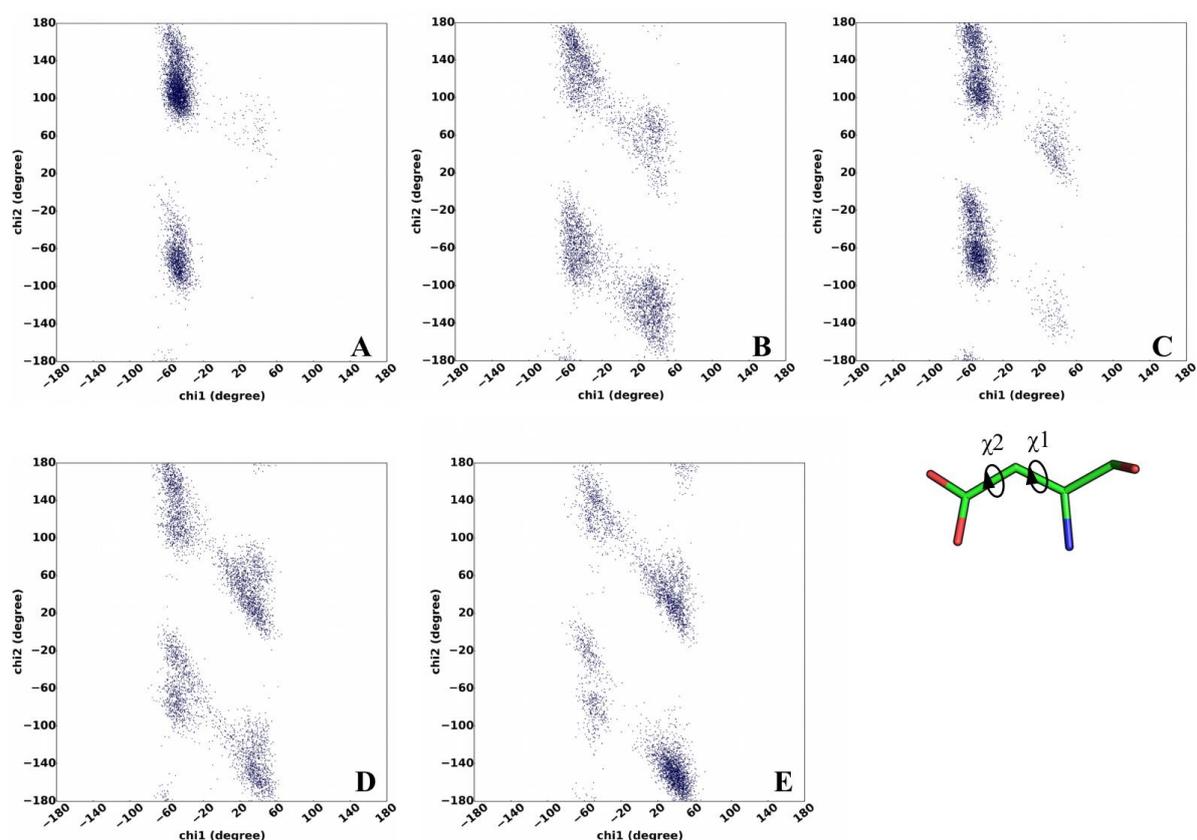


Figure 84. Angles de torsion (χ_1, χ_2) de la chaîne latérale du résidu D176.

En A: l'AgaD sauvage, en B: le mutant Q342H, en C: le mutant D341L/S351A, en D: le mutant D341L/S351A/Q342H, en E: le mutant D341L/S351F/Q342H.

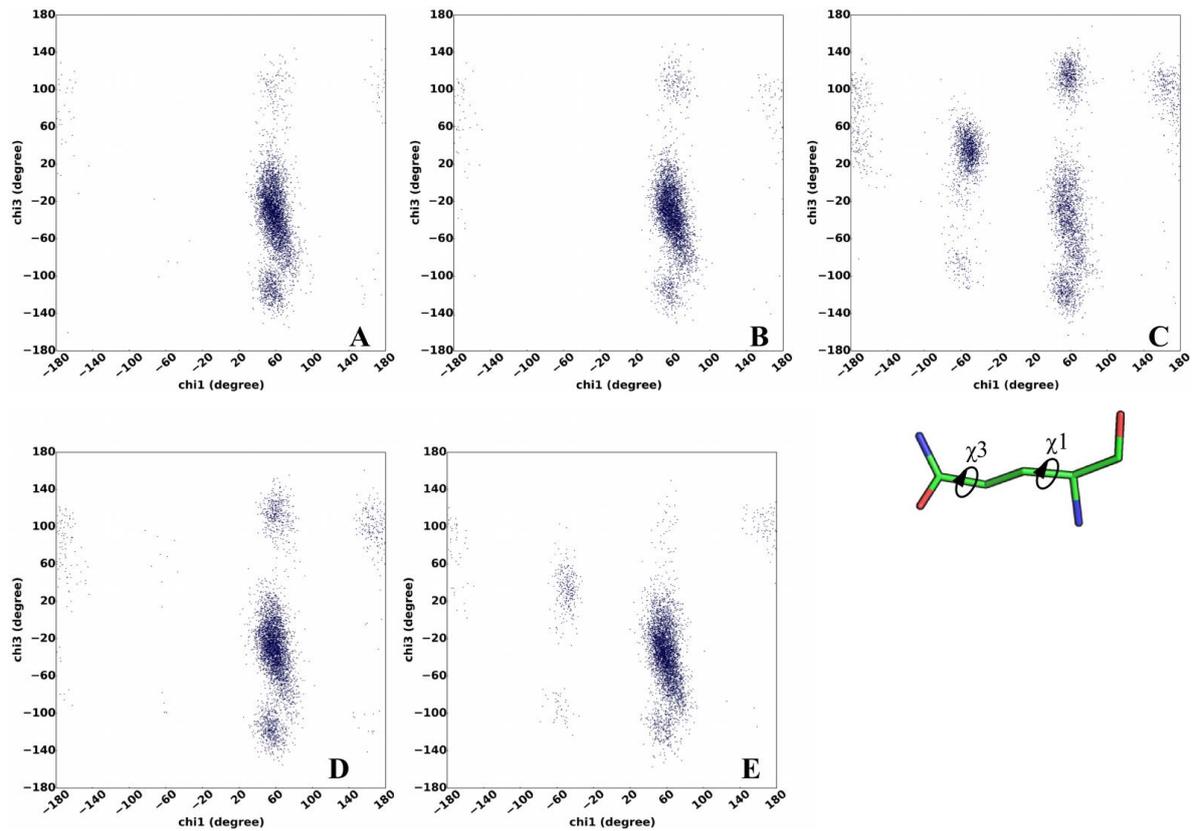


Figure 85. Angles de torsion (χ_1, χ_3) de la chaîne latérale du résidu Q214.

En A : l'AgaD sauvage, en B : le mutant Q342H, en C : le mutant D341L/S351A, en D : le mutant D341L/S351A/Q342H, en E : le mutant D341L/S351F/Q342H.

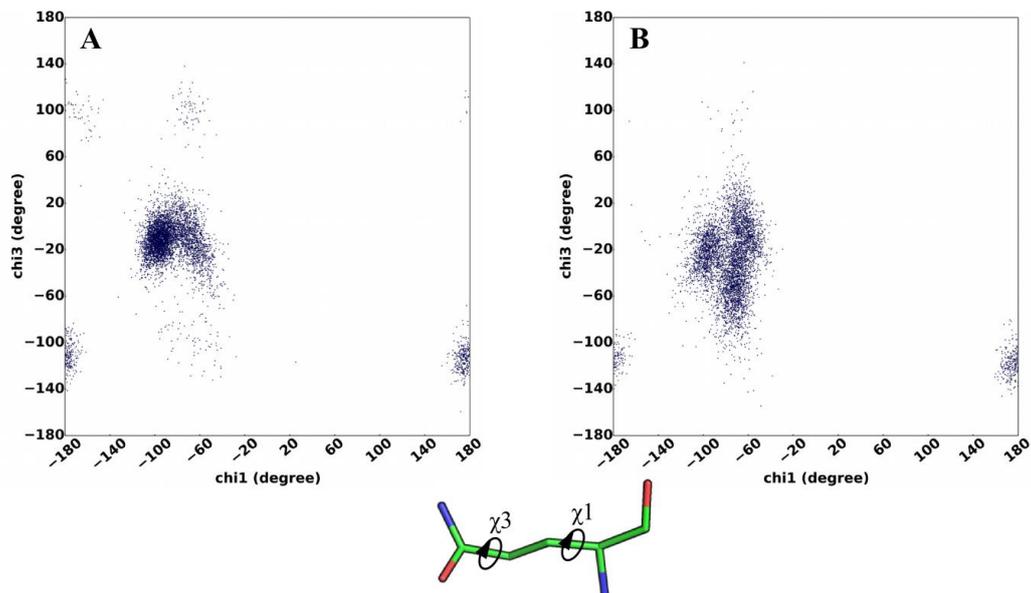


Figure 86. Angles de torsion (χ_1, χ_3) de la chaîne latérale du résidu Q342.

En A et B, respectivement : chez l'AgaD sauvage et le double-mutant D341L/S351A.

Mutants Résidu	Q342H	D341L/ S351A	D341L/ Q342H/ S351A	D341L/ Q342H/ S351F
	D176	+++	++	+++
E179	++	+	+++	+++
Y181	+++	0	++	+++
Q214	0	+++	0	++
Q342	-	++	-	-

Tableau 7. Effets des mutations sur la dynamique conformationnelle des résidus étudiés précédemment.

L'effet propre à une mutation donnée est évalué par rapport à l'enzyme sauvage. L'importance de l'effet d'une mutation donnée est quantifiée par un nombre croissant de « + ». 0 indique que la mutation n'a pas d'effet sur la dynamique conformationnelle du résidu. « - » signifie l'absence de comparaison.

Au vu des résultats précédents, il est tentant de supposer qu'une modification de la dynamique conformationnelle des résidus interagissant avec l'acide-base catalytique puisse altérer leur capacités à établir des liaisons hydrogène avec ce dernier. Comme l'augmentation de la dynamique des résidus D176 et Y181 semble corrélérer avec l'augmentation de l'activité de transglycosylation (Tableau 7), seul le réseau de liaisons hydrogène impliquant ces deux résidus et l'acide-base catalytique a été caractérisé.

7.4.5 Caractérisation des liaisons H stabilisant l'acide-base catalytique

L'analyse du réseau d'interactions montre en réalité que le résidu Y181 interagit principalement avec le résidu E179 au sein de l'enzyme sauvage *via* des liaisons hydrogène pontées par des molécules d'eau. Ce dernier est également très peu stabilisé par le résidu D176. De manière générale, l'introduction des mutations perturbe significativement le réseau de liaisons hydrogène établies entre ces résidus et l'acide-base catalytique (Figure 87). Contrairement à toute attente, ce dernier établit un nombre plus important de liaisons hydrogène en présence des mutations qui déstabilisent le plus fortement sa chaîne latérale. Dans le cas des interactions par liaisons hydrogène directes, ce résultat suggère que des mouvements corrélés pourraient avoir lieu entre la chaîne latérale du résidu E179 et celles des résidus D176/Y181. Sous l'influence des

mutations, les chaînes latérales de ces derniers adopteraient au cours de leur trajectoire des conformations « proches » de celle de l'acide-base catalytique et pourraient ainsi former plus fréquemment des liaisons hydrogène avec cette dernière. Le réseau d'interactions à longue distance entre les résidus D176/Y181 et E179 impliquant la formation d'une liaison hydrogène pontée par d'une ou deux molécule(s) d'eau est également considérablement affecté par les mutations. Ces modifications traduisent ainsi une perturbation de l'organisation des molécules d'eau internes au site catalytique de l'enzyme.

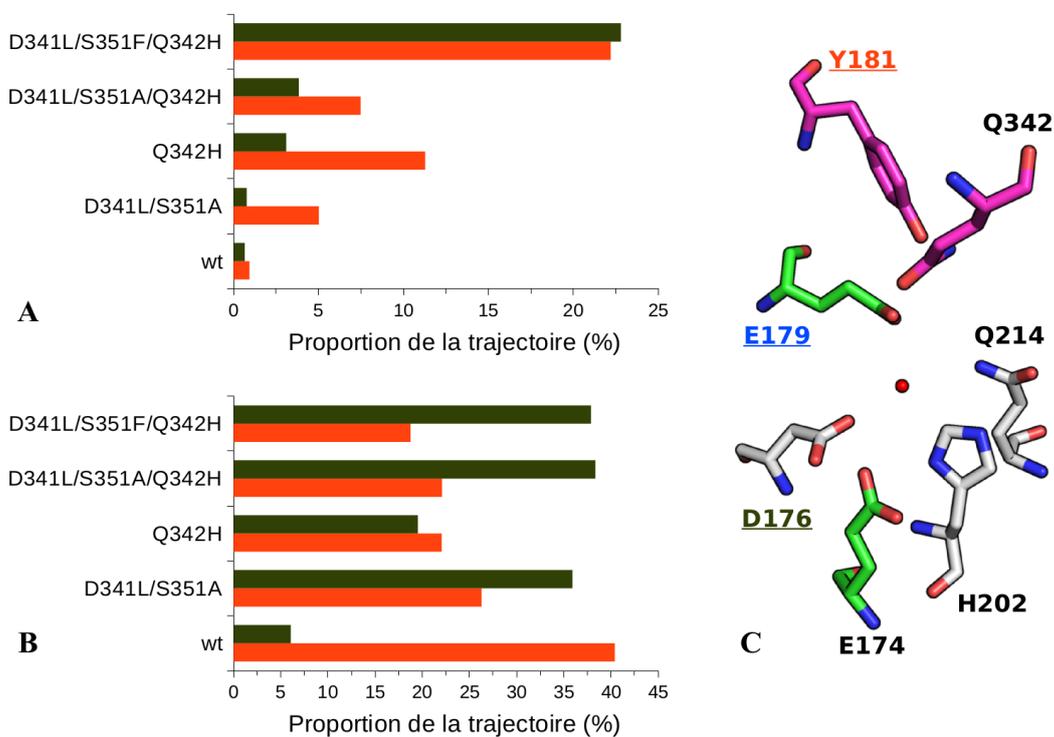


Figure 87. Liaisons hydrogène stabilisant le résidu acide-base catalytique.

Proportion de clichés de la trajectoire pour lesquels au moins une liaison hydrogène est établie entre les résidus E179 et D176 (verdâtre) et Y181 (orange) par contact direct résidu-résidu (A) ou relayées par une ou deux molécules d'eau intermédiaires (B). En C : Les résidus catalytiques sont illustrés en vert, ceux participant à la formation d'un goulet d'étranglement en sortie du canal d'eau étudié à l'interface avec le sous-site -1 sont colorés en magenta. D'autres résidus du sous-site -1 liant l'acide-base catalytique et dont la dynamique conformationnelle est modifiée par les mutations sont illustrés en gris.

7.4.6 Conclusions

Les résultats précédents montrent que les mutations impliquant le résidu Q342 perturbent significativement la dynamique conformationnelle de la protéine et plus précisément celle de l'acide-base catalytique ainsi que des résidus en interaction avec celui-ci. La perturbation du réseau de liaisons hydrogènes dans lequel sont impliqués ces résidus par l'augmentation de la dynamique de leurs chaînes latérales pourraient très probablement affecter la stabilisation de l'état de transition initiant la réaction d'hydrolyse.

D'autre part, la réorganisation spatiale du réseau de molécules d'eau impliquées dans la stabilisation de ces chaînes latérales pourrait également avoir une influence non-négligeable sur la catalyse. De nombreux travaux ont en effet soulevé l'importance de l'agencement des molécules d'eau au sein des sites actifs d'enzymes dans la pré-organisation des interactions électrostatiques indispensables à la catalyse ²³⁷. Il est donc possible que ces effets puissent expliquer les résultats expérimentaux obtenus.

Comme évoqué précédemment, il est également possible que la simple mutation du résidu Q342 puisse perturber la stabilisation de l'état de transition de la réaction d'hydrolyse. L'étude de la structure cristallographique d'AgaB, un homologue d'AgaD complexé au galactose, montre en effet que la glutamine correspondante stabilise par une liaison hydrogène l'atome d'oxygène O4 du sucre donneur ¹⁷.

Si l'introduction de la double-mutation D341L/S351A perturbe également fortement le réseau de liaisons hydrogène stabilisant l'acide-base catalytique, elle ne modifie pas de manière significative la dynamique conformationnelle de sa chaîne latérale en comparaison à l'enzyme sauvage. Il est donc possible que l'établissement de nouvelles interactions entre ce résidu et des résidus proches puisse compenser ces effets et permettre à ce dernier de conserver une population de conformères analogue à celle rencontrée chez l'enzyme sauvage. Bien que leur implication dans la catalyse n'ait pas été démontrée à ce jour, il n'est cependant pas exclu que la modification de la dynamique conformationnelle des résidus D176, Q214 (*Figure 85*), et Q342 (*Figure 86*)

chez le double-mutant puisse déstabiliser l'état de transition de la réaction d'hydrolyse et ainsi expliquer les résultats expérimentaux obtenus.

Comme chez les autres mutants, la réorganisation du réseau de liaisons hydrogène impliquant des molécules d'eau du sous-site -1 pourrait également avoir un impact significatif sur la catalyse. Cet effet reflète de manière générale une modification de la dynamique de l'eau interne au sous-site -1 qui pourrait être liée à une modification de la dynamique des molécules d'eau internes au canal d'eau cristallographique mis en évidence précédemment. En vue de le vérifier, une analyse de la trajectoire des molécules d'eau interne à la protéine et au canal putatif a été effectuée.

7.5 Analyse de la dynamique de l'eau interne à la protéine

Contrairement au chapitre précédent, l'analyse des trajectoires des molécules d'eau internes à la protéine ne comprendra pas l'identification et l'étude des molécules d'eau potentiellement catalytiques. Définies dans le chapitre premier comme l'ensemble des molécules d'eau dont l'atome d'oxygène se situe à une distance respective de 3 et 4 Å des résidus acide-base et nucléophile catalytiques au sein de l'intermédiaire covalent glycosyl-enzyme, elles n'ont pas pu être identifiées ici du fait que les simulations ont été effectuées sur l'enzyme libre et non sur le complexe de Michaelis ou l'intermédiaire covalent glycosyl-enzyme.

7.5.1 Diffusion de l'eau interne au sein de la protéine

De manière encore plus spectaculaire que chez Tt β gly, le confinement de l'eau interne au sein de la structure d'AgaD perturbe sa diffusion au-delà de 5 à 10 ns de trajectoire (*Figure 88*). Les fluctuations importantes du déplacement quadratique des molécules d'eau internes ne permettent donc pas de calculer une vitesse de diffusion unique sur l'intégralité de la trajectoire et ainsi d'évaluer l'effet des mutations sur cette vitesse.

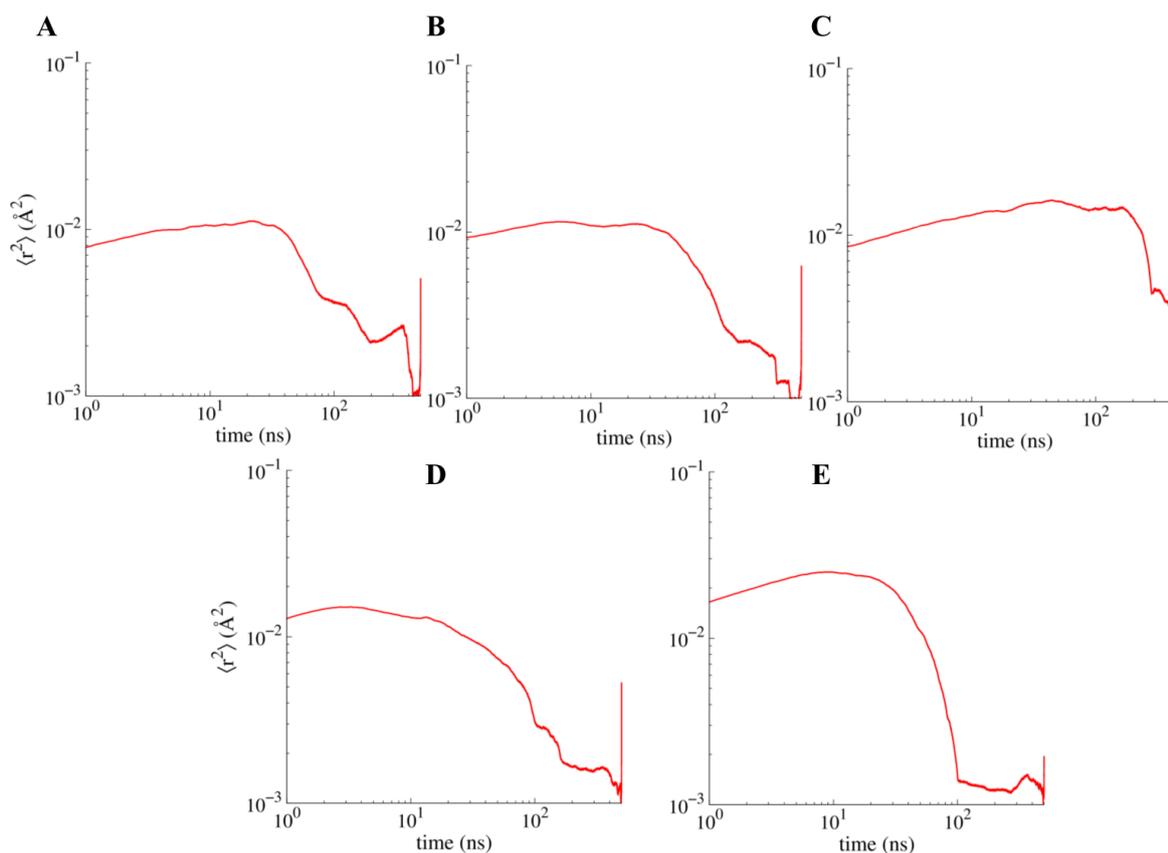


Figure 88. Déplacement quadratique moyen des molécules d'eau internes à la structure protéique d'AgaD.

En A: l'AgaD sauvage, B: mutant D341L/S351A, C: mutant Q342H, D: mutant D341L/S351A/Q342H et E: mutant D341L/S351F/Q342H.

La distribution des valeurs de temps mis par les molécules d'eau internes pour parcourir une distance d'au minimum 3 Å au sein de la protéine ne montre pas de différences significatives entre l'enzyme sauvage et les mutants (*Figure 89*). Tout comme pour Ttβgly, on peut également conclure que les mutations introduites ne semblent pas significativement altérer la diffusion de l'eau à l'intérieur de la protéine.

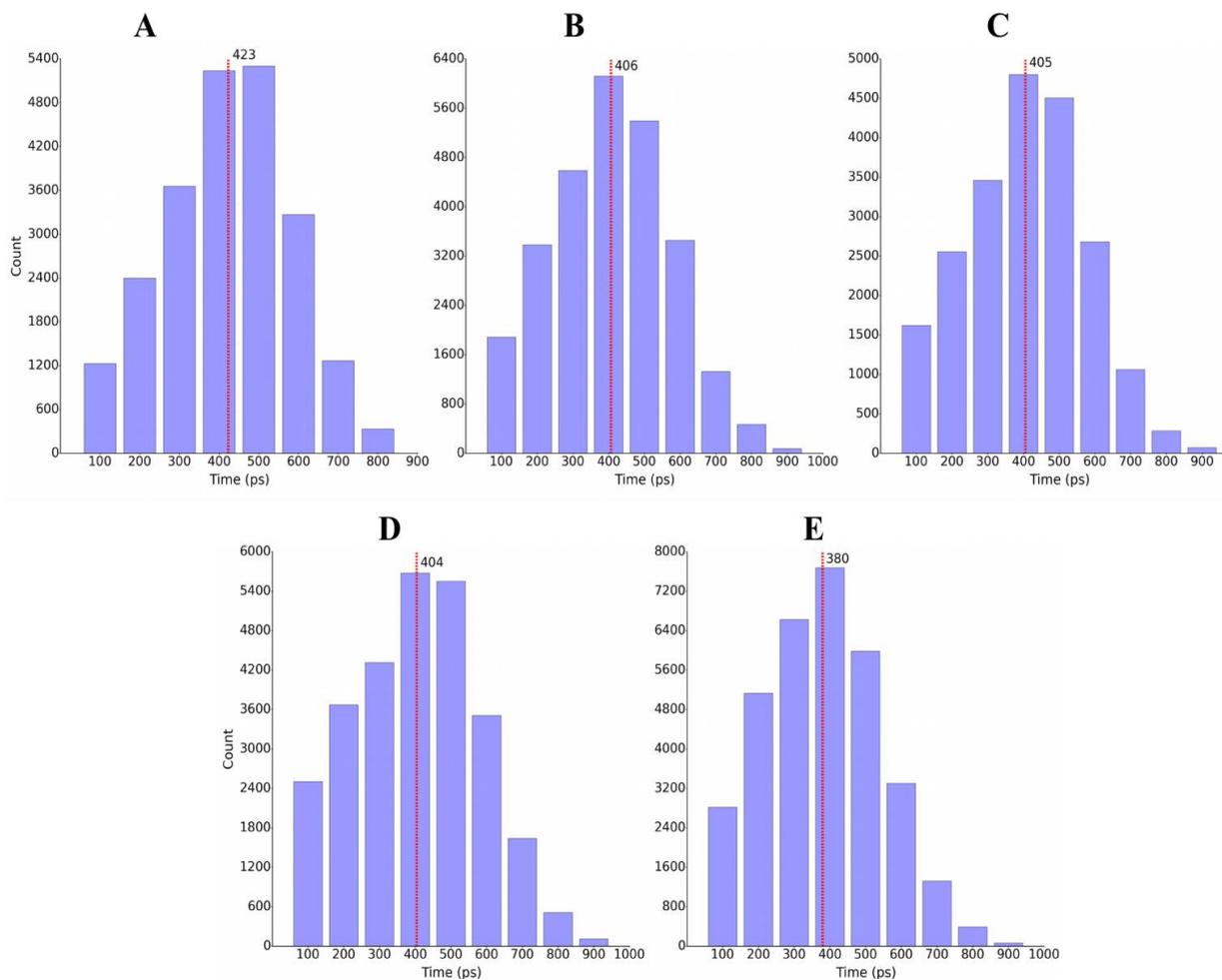


Figure 89. Distribution du temps mis par les molécules d'eau internes à AgaD pour parcourir une distance d'au minimum 3 Å au sein de la protéine.

En A: l'AgaD sauvage, B: mutant D341L/S351A, C: mutant Q342H, D: mutant D341L/S351A/Q342H et E: mutant D341L/S351F/Q342H. La valeur moyenne est indiquée par une ligne verticale rouge.

7.5.2 Identification des canaux et poches d'eau internes à la protéine

Par analogie avec Ttβgly, l'ensemble des réseaux de molécules d'eau internes à la protéine ont été caractérisés et clustérisés afin d'identifier les canaux et les poches d'eau enfouies au sein de la structure protéique. L'observation des trajectoires des molécules d'eau enfouies montre quelques différences notables entre l'enzyme sauvage et les mutants (Figure 90). L'accessibilité de l'eau au sein de nombreuses cavités internes à la structure protéique s'avère notamment être plus élevée chez les triple-mutants, c'est-à-dire ceux qui présentent la plus forte activité de transglycosylation (Figure 91).

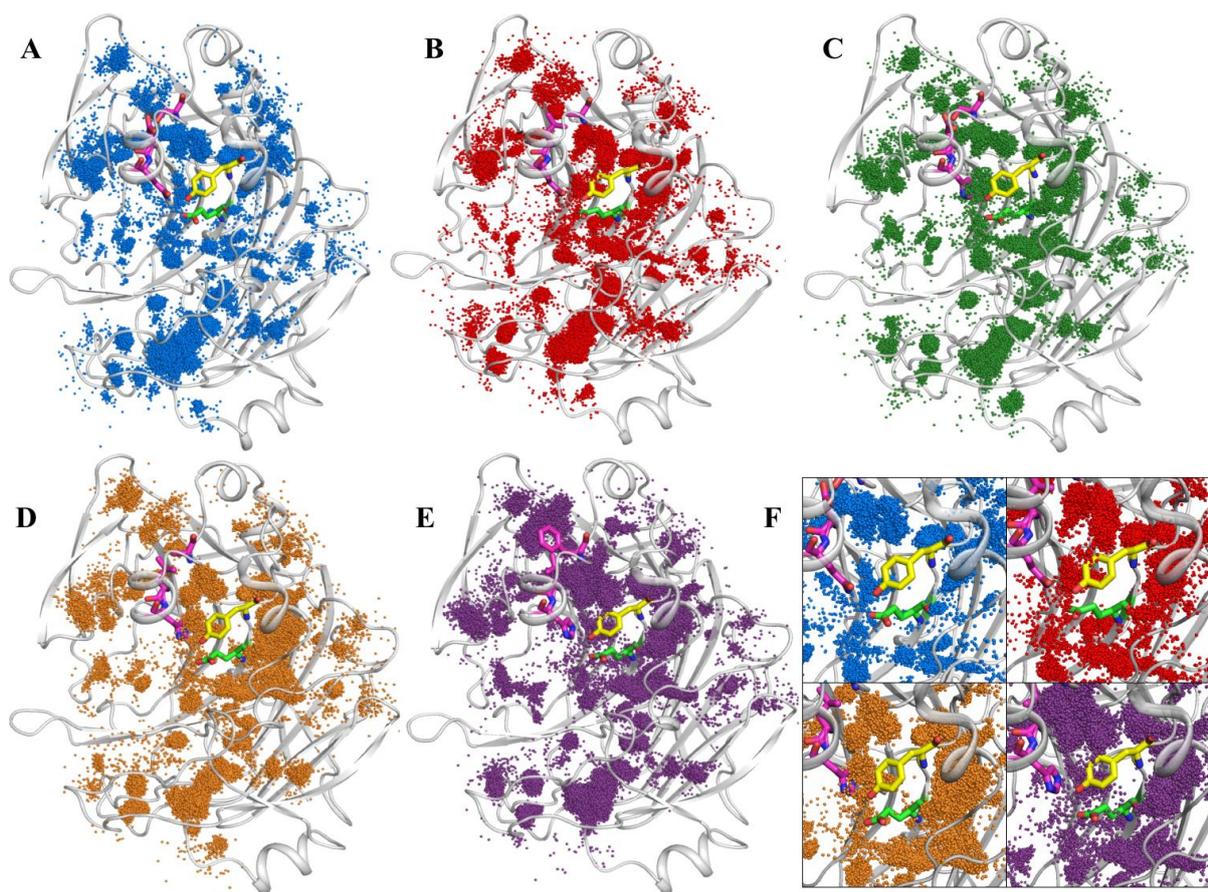


Figure 90. Ensemble des poches et canaux d'eau potentiels isolés sur 500 ns de trajectoire de dynamique moléculaire.

Les résidus acide-base catalytique et Y181 sont illustrés en vert et jaune respectivement. Les résidus aux positions mutées sont colorés en magenta.

En A : l'AgaD sauvage, en B : le mutant D341L/S351A, en C : le mutant Q342H, en D : le mutant D341L/S351A/Q342H, en E : D341L/S351F/Q342H. L'encadré F met en valeur les régions montrant les plus fortes modifications de solvations entre les différents mutants.

Ce résultat est cohérent avec les mesures de RMSD évoquées précédemment : il est logique que la plus forte mobilité du squelette peptidique des mutants puisse perturber la dynamique de l'eau interne en contact avec ce dernier.

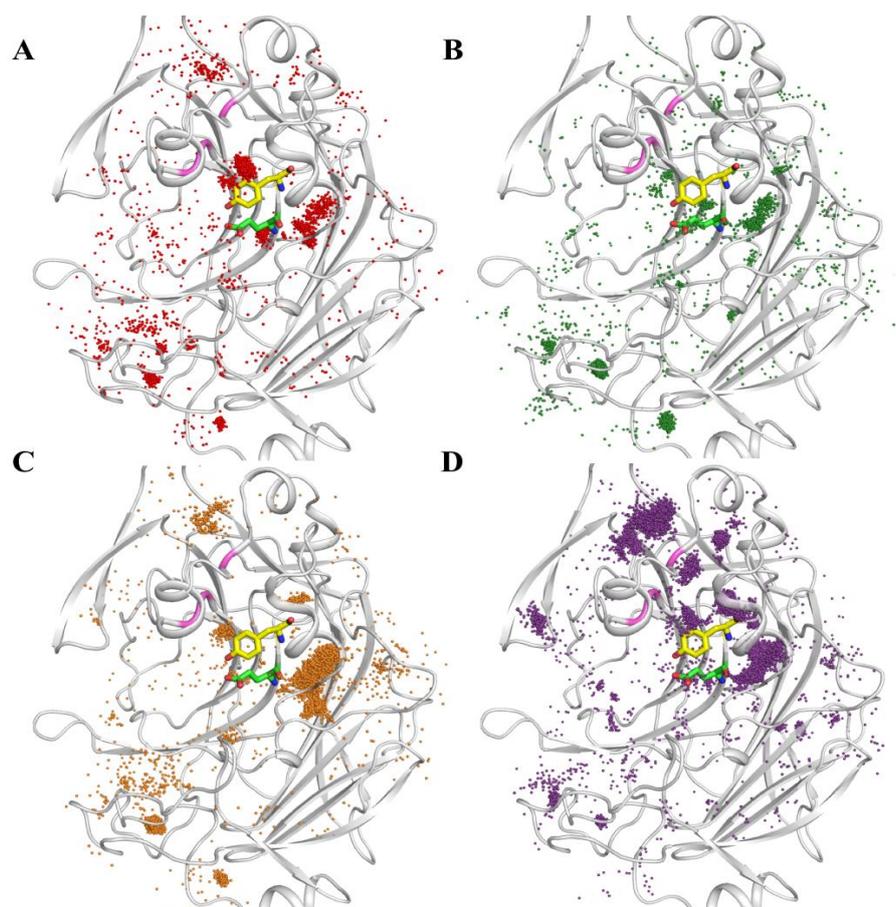


Figure 91. Ensemble des positions spécifiquement explorées par les molécules d'eau internes chez les mutants.

Ces positions ont spécifiquement été isolées chez les mutants dans le sens où elles ne sont pas retrouvées dans la trajectoire de l'eau interne chez l'enzyme sauvage. Le code couleur est identique à celui de la figure 90. En A : le mutant D341L/S351A, B : le mutant Q342H, C : le mutant D341L/S351A/Q342H et D : le mutant D341L/S351F/Q342H.

Par ailleurs, la mise en évidence d'un réseau de molécules d'eau courant le long de la chaîne d'eau précédemment observée au sein de la structure cristallographique d'AgaD confirme bien l'existence d'un canal potentiel à cet endroit. Ce réseau de molécules d'eau s'apparente cependant plus à une poche ou un réservoir qu'à un véritable canal chez l'enzyme sauvage et le mutant D341L/S351A (*Figure 92*). Chez ces derniers, les molécules d'eau constituant ce canal adoptent en effet une trajectoire limitée autour de trois molécules d'eau (2329, 2330, et 2554) de la structure cristallographique d'AgaD au voisinage desquelles elles forment une poche d'eau délimitée par les paires de résidus D341/S351 et Q342/Y181. Un relargage de molécules d'eau de la poche jusqu'au voisinage de l'acide-base catalytique semble en revanche être engendré par la mutation

Q342H et montre ainsi que la dynamique des molécules d'eau internes au canal est considérablement modifiée chez l'ensemble des mutants porteurs de cette mutation.

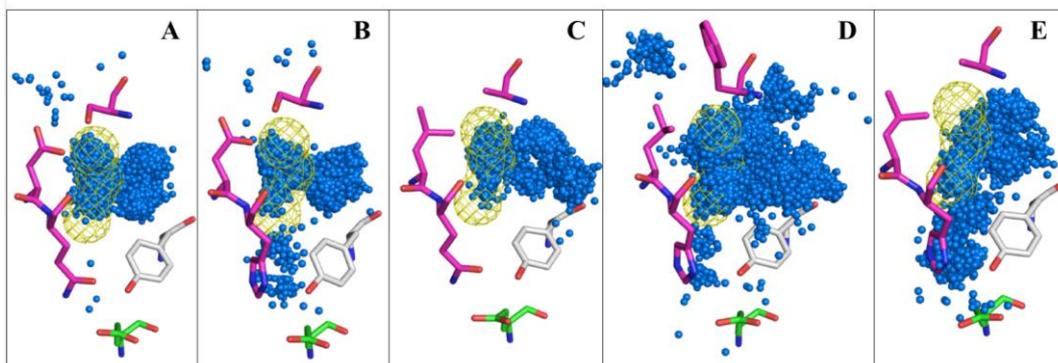


Figure 92. Trajectoire du canal d'eau potentiel mis en évidence par simulation de dynamique moléculaire.

En A : l'AgaD sauvage, en B : le mutant Q342H, en C : le mutant D341L/S351A, en D : le mutant D341L/S351F/Q342H, en E : le mutant D341L/S351A/Q342H. Les atomes d'oxygènes des molécules d'eau formant ce canal potentiel sont illustrés sous la forme de sphères bleues. Les molécules d'eau retrouvées dans la structure cristallographiques d'AgaD (PDB 4ASM) localisées entre les deux paires de résidus formant les goulets d'étranglement délimitant l'accès à la surface de la protéine et au sous-site -1 sont illustrées en maillage jaune. Les résidus acide-base catalytique et Y181 sont respectivement représentés en bâtonnets vert et gris. Les résidus aux positions mutés sont colorés en magenta.

Il est possible que l'accessibilité accrue des molécules d'eau du canal au sous-site -1 chez ces mutants puisse contribuer à perturber le réseau d'interactions dans lequel sont impliqués les résidus du sous-site -1 et justifier la formation d'un nombre plus important de liaisons hydrogène entre ces résidus et les molécules d'eau environnantes. En revanche, cette hypothèse ne permet pas d'expliquer la réorganisation du réseau de liaisons hydrogène observée chez le double-mutant D341L/S351A puisque les molécules d'eau présentes au sein de la poche semblent y rester confinées. Compte tenu des différences observées entre les trajectoires des molécules d'eau constituant ce canal chez l'enzyme sauvage et ces mutants ainsi que de son rôle fonctionnel probable dans l'activité d'AgaD, l'ensemble des analyses suivantes seront consacrées exclusivement à l'étude de la dynamique de l'eau interne à ce dernier.

7.5.3 Purges du canal d'eau caractérisé

Par analogie avec le chapitre précédent, le temps de purge se rapporte à la durée de vie d'un groupe de molécules d'eau échantillonné sur un cliché de simulation donné dans la trajectoire du canal. La comparaison des valeurs de temps de purge sur l'ensemble de la trajectoire de dynamique moléculaire montre des différences significatives entre l'AgaD sauvage et ses mutants (*Figure 93*). L'introduction des mutations diminue considérablement le temps de purge de l'eau interne au canal et accélère ainsi sa dynamique. Ce résultat est cohérent avec l'allure des trajectoires de molécules d'eau observées chez les mutants porteurs de la mutation Q342H : l'accessibilité importante de l'eau du canal au sous-site -1 chez ces mutants pourrait favoriser et accélérer leur départ du canal vers le milieu.

Bien que les molécules d'eau trouvées dans la trajectoire du canal d'eau cristallographique semblent rester confinées dans la cavité délimitée par les paires de résidus Q342/Y181 et D341/S351 chez le mutant D341L/S351A, il est possible que l'augmentation de la flexibilité de la structure protéique provoquée par cette mutation puisse leur permettre de s'en échapper plus rapidement.

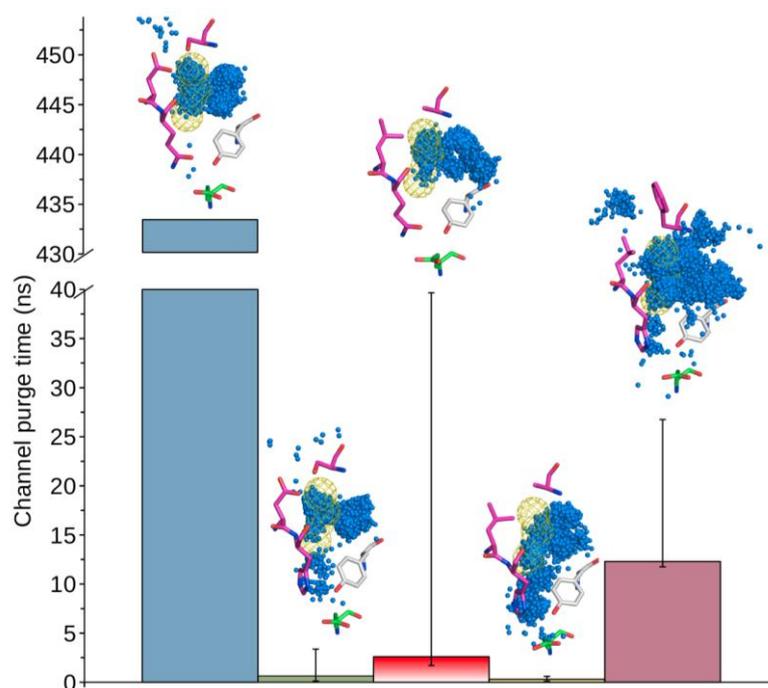


Figure 93. Temps de purge de l'eau au sein du canal étudié.

Les diagrammes bâtons affichent les valeurs médianes des temps de purge de l'eau du canal caractérisé chez l'AgaD sauvage (bleu) et les mutants Q342H (vert), D341L/S351A (rouge), D341L/S351A/Q342H (jaune) et D341L/S351F/Q342H (violet). Les barres inférieures et supérieures indiquent respectivement les 1^{er} et 3^{ème} quartiles. La trajectoire de molécules d'eau propres à chaque canal est illustrée à titre indicatif.

7.5.4 Résidence de l'eau au sein du canal

Le calcul des temps de résidence moyens des molécules d'eau internes au canal confirme que les mutations accélèrent leur dynamique (Figure 94). Chez les mutants, le temps de résidence moyen de ces dernières au sein de la cavité située à l'interface entre les résidus mutés est significativement réduit en comparaison à l'enzyme sauvage. Cette observation appuie donc les résultats précédents qui montraient une diminution importante des temps de purge de l'eau interne à cette cavité. L'augmentation de la dynamique de la structure protéique autour du canal provoquée par les mutations pourrait ainsi être à l'origine d'une fuite accrue des molécules d'eau en son sein vers la surface ou le site actif.

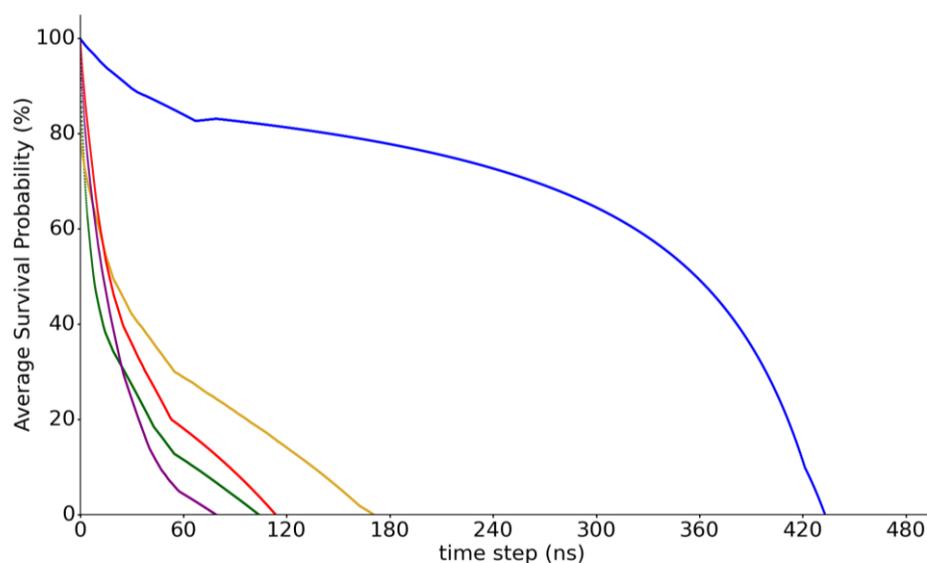


Figure 94. Courbes de survie de l'eau au sein du canal étudié.

Le point d'intersection des courbes avec l'axe des abscisses indique le temps de résidence moyen des molécules d'eau au sein du canal. Le code couleur est identique à celui de la figure précédente.

7.5.5 Conclusions

L'ensemble des résultats précédents montre que les résidus 341, 342, et 351 contrôlent la dynamique des molécules d'eau internes à la cavité qu'ils délimitent. Le résidu Q342 joue en outre un rôle crucial dans le contrôle de l'accès de l'eau de cette cavité au site -1. Au regard des résultats expérimentaux et théoriques, une accélération de la dynamique des molécules d'eau internes à cette poche ainsi qu'une augmentation de leur accessibilité au sous-site -1 semble de manière générale bien corrélée avec une diminution importante de l'activité hydrolytique. Cette corrélation peut paraître à première vue contre-intuitive puisqu'il serait logique qu'une accessibilité plus importante des molécules d'eau au site catalytique puisse favoriser la réaction d'hydrolyse. Il est cependant probable que l'augmentation de la dynamique des molécules d'eau dans cette cavité puisse accroître la dynamique locale des résidus alentour et de l'eau catalytique.

Si l'étude du double-mutant D341L/S351A a montré que la dynamique de la chaîne latérale du résidu acide-base catalytique ne permettait pas toujours d'expliquer l'effet des mutations sur l'activité d'AgaD, la stabilisation du réseau de liaisons hydrogène impliquant les résidus et les molécules d'eau du sous-site -1 pourrait être en revanche crucial pour le positionnement correct de la molécule d'eau catalytique et la formation de l'état de transition correspondant. Contrairement à la molécule d'eau catalytique, le positionnement de l'accepteur de la réaction de transglycosylation ne doit pas être en revanche significativement altéré par les mutations. En effet, ce dernier dispose d'un grand nombre de groupements hydroxyles lui permettant d'établir de nombreuses interactions avec la protéine au niveau des sous-sites +1, +2, +3 et +4. Comme la molécule d'eau catalytique n'est stabilisée que par deux interactions très spécifiques, ceci expliquerait donc pourquoi une perturbation significative du réseau de liaisons hydrogène au sein du sous-site -1 affecterait préférentiellement la réaction d'hydrolyse au profit de la transglycosylation.

- Chapitre V -
Généralisation de
l'étude au GH de la
famille 33

8 Généralisation de l'étude aux enzymes de la famille des GH33

8.1 Sialidases et trans-sialidases de la famille 33

Aussi connues sous le nom de sialidases ou neuraminidases, ces enzymes catalysent l'hydrolyse d'un résidu d'acide sialique (très souvent l'acide N-acétyl-neuraminique ou Neu5Ac) lié le plus souvent en position $\alpha(2,3)$ ou $\alpha(2,6)$ à un sucre lui-même fréquemment associé à diverses macromolécules, incluant des polysaccharides, glycoprotéines, ou glycolipides. Comptant plus de 50 formes structurales distinctes, les acides sialiques occupent une place importante dans les processus de reconnaissance biologique²³⁹. Les sialidases sont réparties en quatre familles distinctes selon la classification CAZy²⁹ incluant les exo-sialidases regroupées sous les familles GH33 (sialidases bactériennes et eucaryotiques), les GH83 (neuraminidases et hémagglutinines) et GH34 (neuraminidases virales), et les endo-sialidases représentées par la famille GH53 (β -1,4-galactanases). En dépit d'une faible conservation de séquences entre ces différentes familles, l'ensemble de ces enzymes partagent un domaine catalytique très similaire caractérisé par une architecture et une machinerie catalytique commune²³⁹. Ce projet se focalise exclusivement sur la famille des GH33 et plus précisément sur les sialidases et trans-sialidases trypanosomales.

8.2 Sialidase et trans-sialidase des trypanosomes *T. rangeli* et *T. cruzi*.

La trans-sialidase du parasite *Trypanosoma cruzi* (ou TcTS) catalyse le transfert de résidus d'acides sialiques conjugués sur un sucre donneur sur des groupements β -galactosyl de divers accepteurs^{240, 241, 242}. Il a par ailleurs été montré que ces trans-sialidases jouaient un rôle essentiel dans le processus de virulence de nombreux trypanosomes pathogènes tels *T. cruzi* et *T. brucei*, respectivement responsables de la maladie de Chagas^{240, 241, 243} et de la maladie du sommeil^{195, 196} sur les continents sud américain et africain. Cette réaction peut présenter un intérêt industriel²⁴⁴ si le sucre accepteur est à forte valeur ajoutée. Par exemple, le transfert de l'acide sialique du sialyl-lactose sur la lactosamine peut être catalysé par TcTS pour former un produit à forte valeur ajoutée, la sialyl-lactosamine, qui est un des constituants du lait maternel.

La sialidase de *Trypanosoma rangeli* (ou TrSA) s'avère en revanche incapable de transférer l'acide sialique porté par un sucre donneur sur un sucre accepteur. A l'inverse de TcTS, cette enzyme catalyse l'hydrolyse de l'acide sialique²⁴⁵. En dépit de leurs activités catalytiques distinctes, la détermination des structures 3D d'une isoforme de TrSA et TcTS a montrée que les deux enzymes partageaient un repliement analogue²¹². Les structures 3D révèlent en effet une architecture commune formée de l'association d'un domaine catalytique N-terminal et d'un domaine C-terminal de type lectine connectés entre eux par une hélice- α (Figure 95). Avec une conservation de 70 % des résidus sur l'ensemble du domaine catalytique des deux séquences, ces protéines homologues constituent un modèle de choix pour l'étude des bases moléculaires gouvernant l'équilibre H/T.

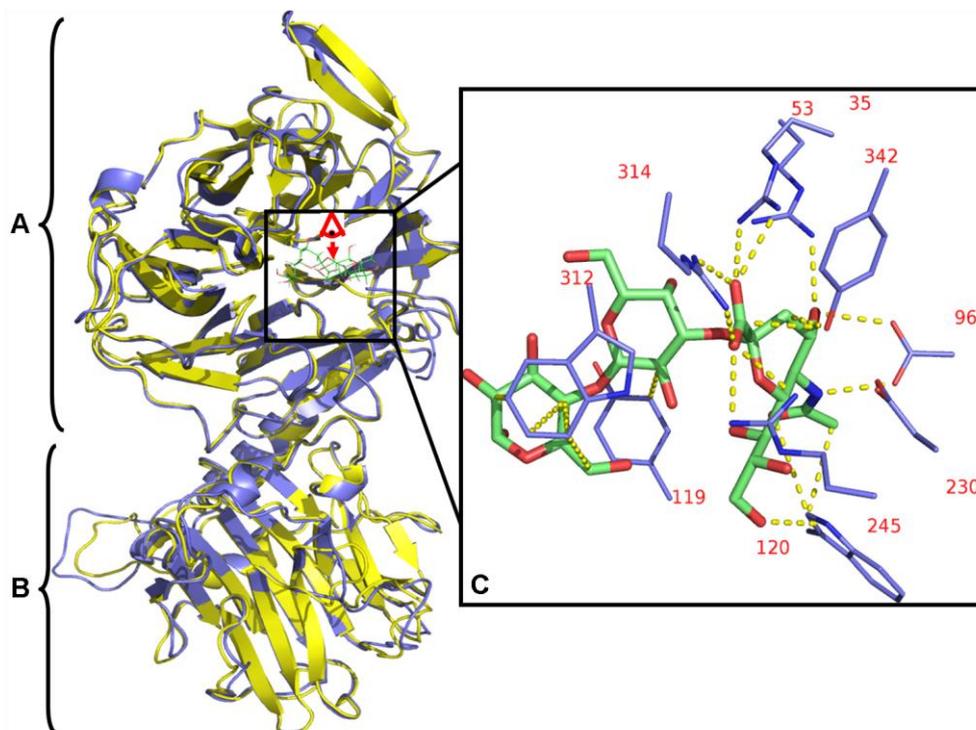


Figure 95. Structure 3D de TcTS²⁴² et TrSA²¹² complexées au sialyl-lactose.

Les structures des deux enzymes ont été alignées afin de mettre en évidence leurs similarités. La structure de TcTS (en jaune) et TrSA (en violet) sont issues de la PDB, d'identifiants respectifs : 1S0I et 1N1S. Les domaines catalytique et lectine sont respectivement identifiés par les lettres A et B. Le sialyl-lactose est illustré en bâtonnets. L'encadré C présente une vue de dessus de l'ensemble des interactions stabilisant ce substrat au sein du site catalytique de TcTS. Les résidus formant la triade catalytique sont D59 (acide-base), Y342 (nucléophile), et E230 (base catalytique). Le site accepteur (+1), exclusivement présent chez TcTS, est essentiellement formé par les résidus W312 et Y119.

Malgré la très forte similarité entre les deux enzymes, les molécules d'eau présentes au sein du site catalytique de TcTS ne réagissent que très peu avec le substrat donneur pour conduire à la réaction d'hydrolyse. De précédentes analyses de simulations de dynamique moléculaire ²⁴⁶ ont suggéré que ce phénomène pouvait s'expliquer par une accessibilité réduite de l'eau au site catalytique de TcTS en comparaison à TrSA. Dans le cadre de cette étude, il a été proposé que le changement conformationnel d'une boucle lors de la formation du complexe de Michaëlis entre TcTS et le substrat donneur contribuerait à exclure une partie des molécules d'eau du site catalytique. Cette boucle porte le résidu W312 qui est impliqué dans la formation du site accepteur avec le résidu Y119. Initiée par l'entrée du substrat donneur au sein du sous-site -1, la formation d'un sandwich hydrophobe impliquant les résidus W312/Y119 ^{242, 247} et le substrat serait à l'origine de l'effet de désolvatation évoqué plus haut. Ces résultats suggèrent donc que la dynamique de la protéine jouerait un rôle crucial dans la catalyse et que les différences observées au cours de cette étude permettraient au moins en partie d'expliquer les différences d'activités entre ces deux enzymes. En dépit du rôle essentiel des résidus W312 et Y119 dans la catalyse de la trans-sialidation chez TcTS ²⁴⁸, la réversion du résidu Y119 par celui trouvé chez TrSA ne permet pas d'inverser la balance H/T. En effet, bien que la réversion S119Y chez TrSA diminue l'activité d'hydrolyse de près de 50 % par rapport à l'enzyme sauvage ^{249, 112}, elle ne permet pas en contrepartie de restaurer une activité trans-sialidase comparable à celle de TcTS. Ces résultats suggèrent ainsi que d'autres résidus doivent intervenir dans le contrôle de l'équilibre entre ces deux activités.

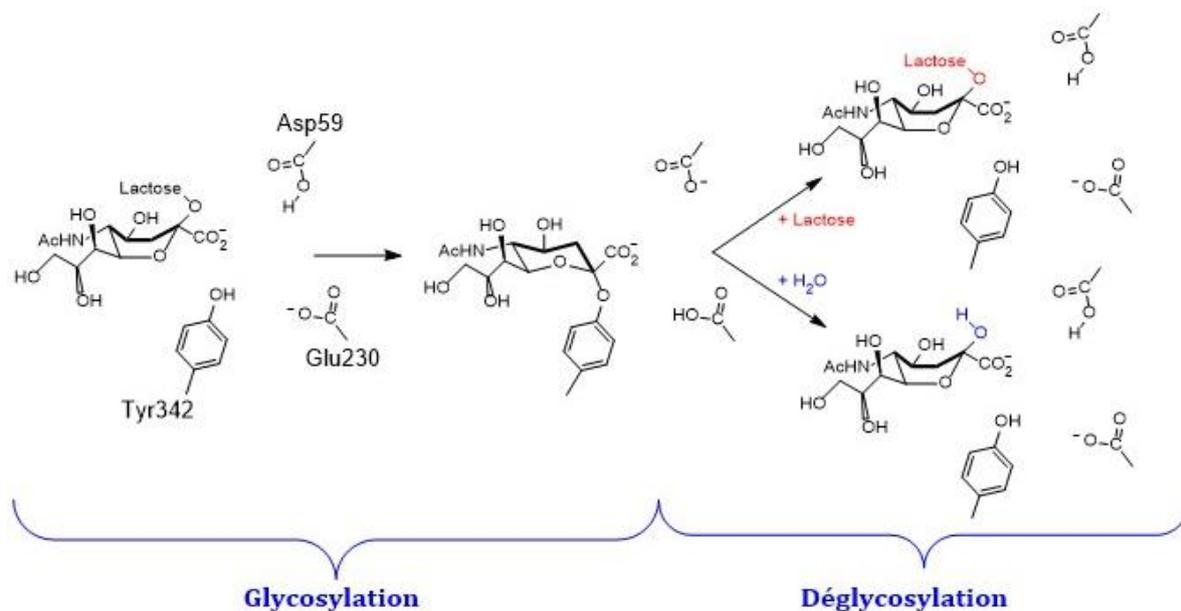


Figure 96. Mécanisme réactionnel de TcTS^{250,211,251}.

Le substrat donneur représenté dans le site de liaison est le 3'-sialyl-lactose (3'SL). La réaction de trans-sialidation, majoritaire chez cette enzyme, est réalisée en présence de lactose pour former du 1'-sialyl-lactose. Une réaction d'hydrolyse du 3'SL, très minoritaire, peut également avoir lieu.

8.3 Ingénierie de la sialidase de *T. rangeli* en trans-sialidase

De nombreuses expériences de mutagenèse ont été réalisées afin de tenter d'identifier l'ensemble des résidus contrôlant la balance H/T chez ces deux enzymes. Les travaux menés par Paris et coll.²⁵² ont par ailleurs mis en évidence six autres résidus du domaine catalytique impliqués dans le contrôle de la balance H/T et localisés en bordure du sous-site -1. Leur substitution chez TrSA par ceux trouvés chez TcTS a ainsi permis de restaurer 10 % de l'activité de trans-sialidation de TcTS chez TrSA. Une étude théorique a également suggérée l'addition de quatre nouvelles mutations à celles évoquées précédemment afin d'accroître l'activité de trans-sialidation chez TrSA²⁵³. D'autres travaux de mutagenèse effectués sur cette même enzyme ont également permis de caractériser un ensemble de résidus distants du site catalytique (motif₁₉₇-VTNKKKQ-₂₀₃ chez TcTS) impliqués dans le contrôle de la balance entre les deux activités²⁵⁴. Par analogie avec le cas Ttβgly, ces résultats confirment la contribution de résidus distants du site actif sur l'activité de transglycosylation.

Compte tenu de la répartition spatiale quasi-identique des résidus catalytiques chez TrSA et TcTS après superposition de leurs structures 3D ²⁴⁵, la restitution de l'activité de trans-sialidation de TcTS chez TrSA pourrait nécessiter la mutagenèse de résidus distants du site catalytique qui pourraient être impliqués dans l'acheminement des molécules d'eau catalytiques ou dans leur positionnement au point d'attaque nucléophile. Jers et coll. ²⁵⁴ ont par ailleurs émis l'hypothèse que l'augmentation de l'activité de trans-sialidation observée chez TrSA à l'issue de la substitution du motif ¹⁹⁷-IADMGGR-²⁰³ par le motif ¹⁹⁷-VTNKKKQ-²⁰³ trouvé chez TcTS pouvait s'expliquer par l'introduction des charges additionnelles des trois lysines du motif ¹⁹⁷-VTNKKKQ-²⁰³. Il a été supposé que l'introduction de ces charges supplémentaires pouvait engendrer une modification locale du réseau de liaisons hydrogène relayées par des molécules d'eau internes connectant cette région en surface au site catalytique. En accord avec les résultats des deux chapitres précédents, une telle hypothèse suggère donc que la dynamique des réseaux de molécules d'eau internes à la protéine pourrait jouer un rôle dans le contrôle de la balance H/T chez les GH33. Si plusieurs études de la dynamique de la structure protéique de TcTS et TrSA ont déjà été publiées, aucun travail portant sur la dynamique de l'eau interne à ces enzymes n'a été publié à ce jour. Par analogie avec les deux chapitres précédents, des simulations de dynamique moléculaire de TcTS et TrSA ont donc été réalisées en vue de caractériser la dynamique de leurs molécules d'eau internes.

8.4 Analyse de la dynamique des structures protéiques de TcTS et TrSA

Dans la continuité des chapitres précédents, une analyse préalable de la dynamique de la structure protéique des formes sauvage de TcTS et TrSA a été effectuée afin de comparer les stabilités de ces deux enzymes. En raison de l'absence de paramètres disponibles pour le résidu sialyl-tyrosine, il a été décidé de simuler les complexes de Michaëlis de chaque enzyme avec le sialyl-lactose à la place des intermédiaires covalents sialyl-enzyme. Les résultats des analyses des trajectoires correspondantes sont présentés ci-dessous.

8.4.1 Stabilité et flexibilité du squelette peptidique

Le suivi du RMSD du squelette peptidique des deux protéines montre que la structure protéique de TrSA est plus stable que celle de TcTS (*Figure 97*). Ces différences de dynamique entre TcTS et TrSA ont déjà été décrites dans la littérature et pourraient jouer un rôle important dans le choix du type de réaction catalysée. Bien que la poche de liaison au substrat chez TrSA présente une dynamique un peu plus élevée que celle de TcTS, les différences de RMSD restent cependant peu significatives ($< 0.5 \text{ \AA}$). Ce résultat suggère ainsi que la dynamique de la région charnière entre les deux domaines de ces protéines pourrait en partie être responsable des différences de stabilité observées.

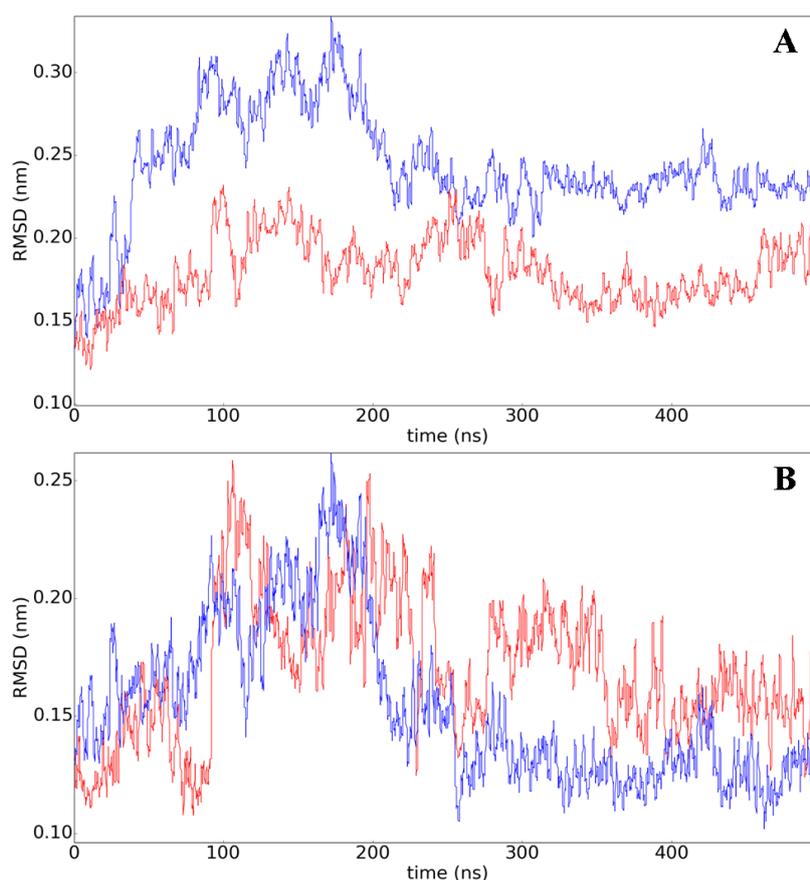


Figure 97. RMSD calculé sur le squelette peptidique (atomes $C\alpha$) de TcTS et TrSA.

La déviation des coordonnées des atomes de carbone $C\alpha$ a été moyennée sur l'ensemble des résidus de la protéine (A) ou sur l'ensemble des résidus situés à 10 \AA de l'acide-base catalytique (B). L'évolution du RMSD de TcTS et TrSA est respectivement indiquée en bleu et rouge.

A l'exception de certaines boucles situées en surface du domaine lectine, ces deux enzymes ne présentent pas de différences importantes en termes de flexibilité moyenne (*Figure 98*). Le fait qu'aucune différence notable de stabilité et de flexibilité ne peut être mise en évidence entre les domaines catalytiques de ces deux enzymes n'est pas surprenant considérant leur architecture très similaires (70 % d'identité de séquence). Des différences de flexibilité locales, impliquant certains résidus spécifiques, ont en revanche déjà été décrites dans la littérature.

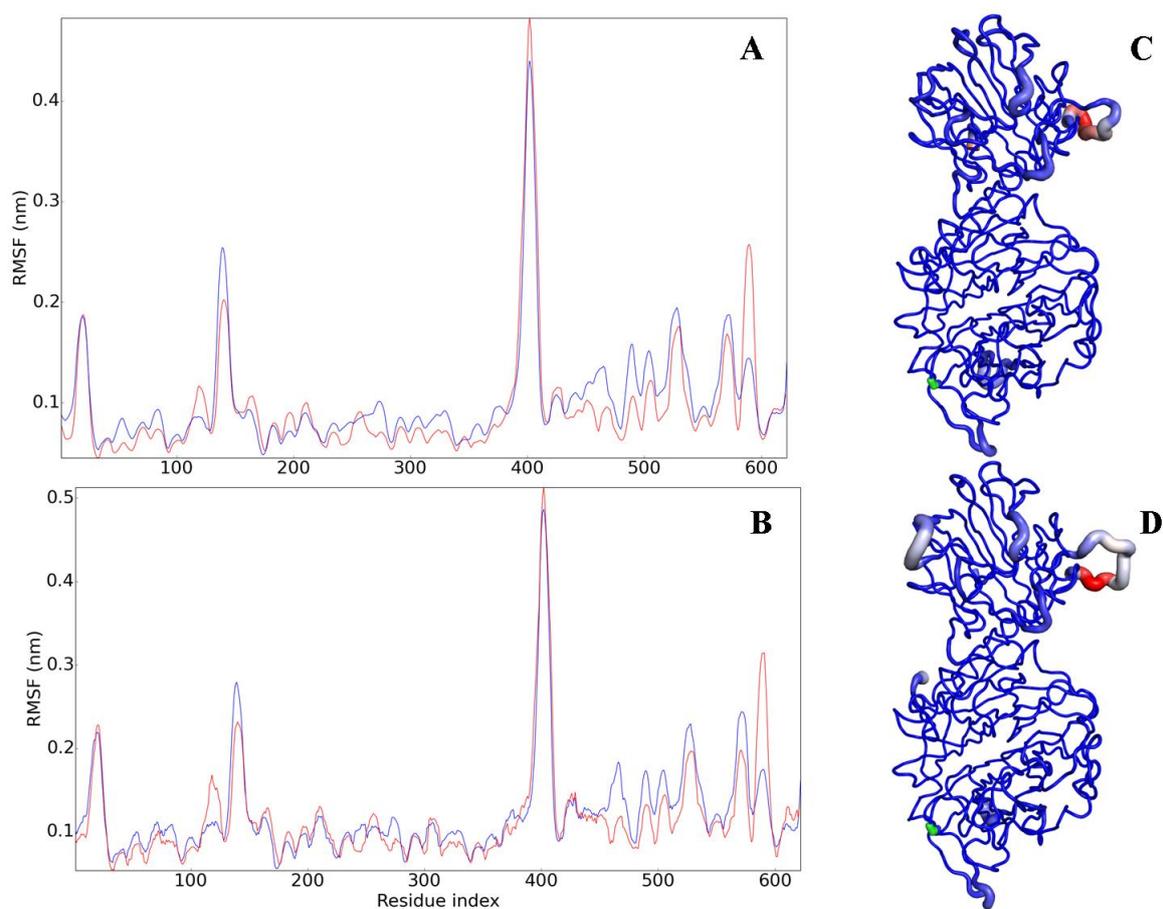


Figure 98. RMSF calculé sur le squelette peptidique et l'ensemble de la protéine

La valeur du RMSF par résidu projeté sur la séquence de TcTS et TrSA est respectivement indiquée en bleu et rouge. A : RMSF moyenné sur les atomes $\text{C}\alpha$ du squelette peptidique. B : RMSF moyenné sur l'ensemble des atomes de chaque résidu. Les valeurs de facteur-B calculées à partir des valeurs de RMSF sont projetées sur le squelette peptidique de la conformation dynamique moyenne de TcTS (C) et TrSA (D) déterminée sur 500 ns de trajectoire de dynamique moléculaire. Le diamètre du squelette peptidique est d'autant plus large et sa couleur d'autant plus chaude que les valeurs de facteur B sont importantes. Le carbone $\text{C}\alpha$ du résidu catalytique acide-base est indiqué en vert.

8.4.2 Flexibilité des chaînes latérales de la triade catalytique

L'échantillonnage des angles de torsion des résidus acide-base et glutamate catalytiques montre que ces deux résidus adoptent au cours de leurs trajectoires des conformations significativement différentes entre TcTS et TrSA (*Figure 99*). Ces différences doivent également avoir une influence non-négligeable sur l'activité de ces deux enzymes puisqu'elles montrent que la machinerie catalytique n'est pas stabilisée de la même manière chez TcTS et TrSA. La plus forte dynamique de la chaîne latérale du résidu acide-base catalytique chez TcTS en comparaison à TrSA pourrait notamment expliquer pourquoi la réaction d'hydrolyse est défavorisée chez cette enzyme au profit de la transglycosylation. Comme déjà discuté dans les chapitres précédents, la stabilisation de la molécule d'eau catalytique est essentielle pour la formation de l'état de transition associé à la réaction d'hydrolyse. Ainsi, il est très probable que l'augmentation de la dynamique de la chaîne latérale du résidu acide-base catalytique chez TcTS puisse compromettre le bon positionnement de la molécule d'eau catalytique au niveau du point d'attaque nucléophile et ainsi justifier la faible activité d'hydrolyse chez cette enzyme.

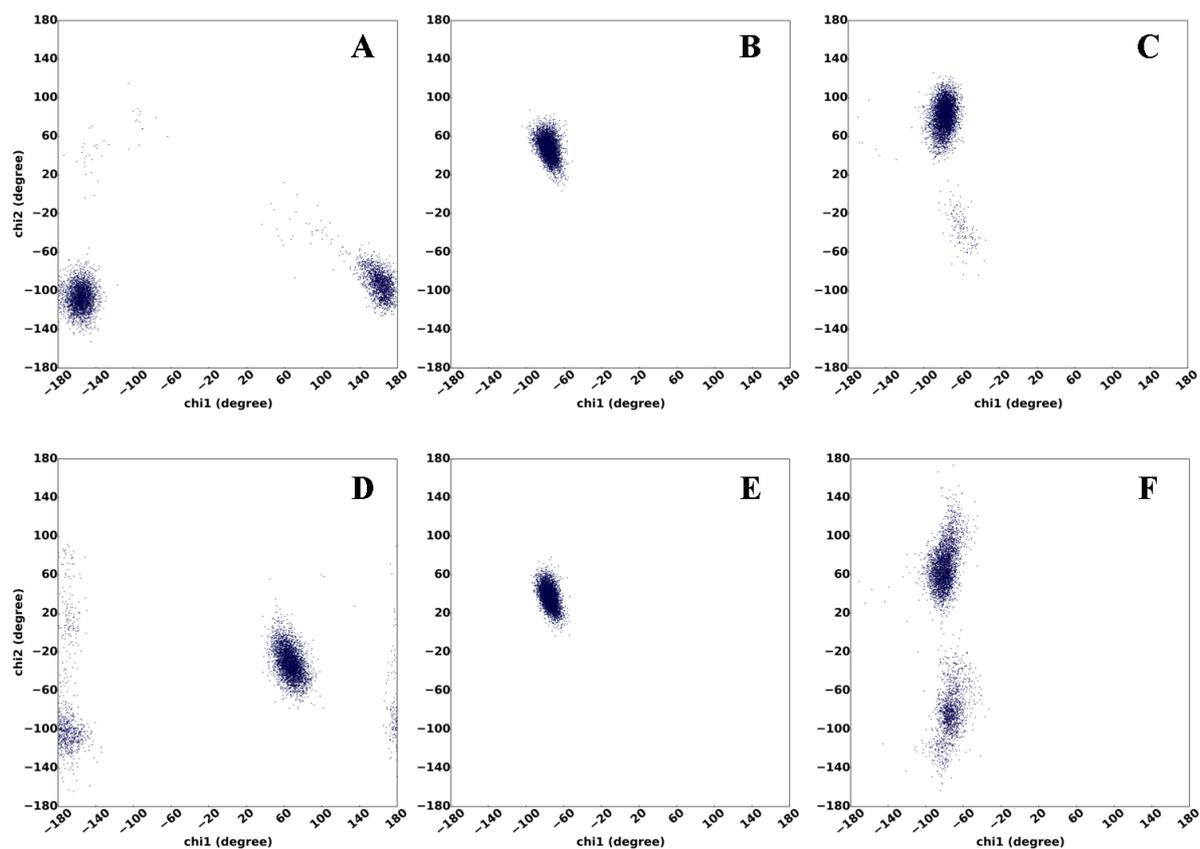


Figure 99. Angles de torsion de la chaîne latérale des résidus catalytiques chez TcTS (A, B, C) et TrSA (D, E, F).

La projection des angles de torsion (χ_1 , χ_2) du résidu catalytique acide-base de TcTS (D59) et TrSA (D60) est respectivement représentée en A et D. La projection des angles de torsion (χ_1 , χ_2) du résidu catalytique nucléophile de TcTS (Y342) et TrSA (Y343) est respectivement représentée en B et E. La projection des angles de torsion (χ_1 , χ_3) de la base catalytique de TcTS (E230) et TrSA (E231) est respectivement représentée en C et F.

Ces résultats appuient donc les conclusions des travaux de Bueren-Calabuig²⁵⁵ qui mettaient en évidence que les états de transition associés à la réaction d'hydrolyse et de transglycosylation n'étaient pas stabilisés de manière identique entre les deux enzymes. Les différences de flexibilité des chaînes latérales de ces deux résidus pourraient donc en grande partie expliquer pourquoi ces deux enzymes, bien que structurellement très proches, présentent des activités différentes. Cependant, une simulation des intermédiaires covalent sialyl-enzymes reste nécessaire pour vérifier ces hypothèses.

8.5 Analyse de la dynamique de l'eau interne à TcTS et TrSA

8.5.1 Identification des canaux et poches d'eau internes à la protéine

Les analyses des trajectoires de dynamique moléculaire générées à partir des structures cristallographiques de TrSA et TcTS complexées au sialyl-lactose ont révélé la présence de cinq canaux d'eau dont quatre sont internes au domaine catalytique (*Figure 100*). Tous débouchent dans le site catalytique chez TrSA contre trois chez TcTS. Il est par ailleurs intéressant de constater que certains canaux d'eau internes à TcTS présentent de nombreuses interruptions en comparaison aux canaux « homologues » chez TrSA.

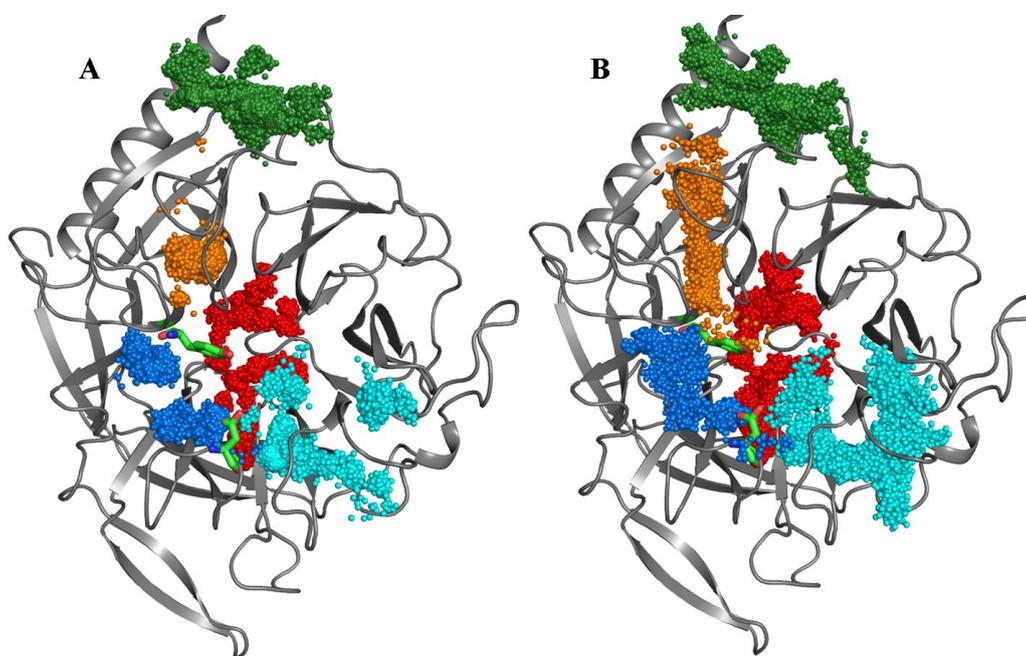


Figure 100. Ensemble des canaux d'eau potentiels isolés sur 500 ns de trajectoire de dynamique moléculaire au sein du domaine catalytique de TcTS (A) et TrSA (B).

Le domaine catalytique est représenté en cartoon gris. Les résidus acide-base et nucléophile catalytiques sont illustrés en bâtonnets. Les atomes d'oxygènes des molécules d'eau internes à un canal spécifique sont représentés en sphères d'une couleur donnée.

Cette fragmentation des canaux d'eau chez TcTS traduit un déficit en molécules d'eau internes qui pourrait sans doute expliquer les différences de stabilité observées précédemment. En effet, comme évoqué dans le chapitre premier, les molécules d'eau enfouies au sein des structures protéiques sont souvent impliquées dans des réseaux

d'interactions polaires avec les résidus alentours et jouent ainsi un rôle essentiel dans la stabilisation de l'architecture des protéines. Il est donc tentant de supposer qu'un déficit en molécules d'eau à l'intérieur de la structure protéique de TcTS puisse déstabiliser cette dernière ; justifiant ainsi sa plus grande flexibilité en comparaison à celle de TrSA.

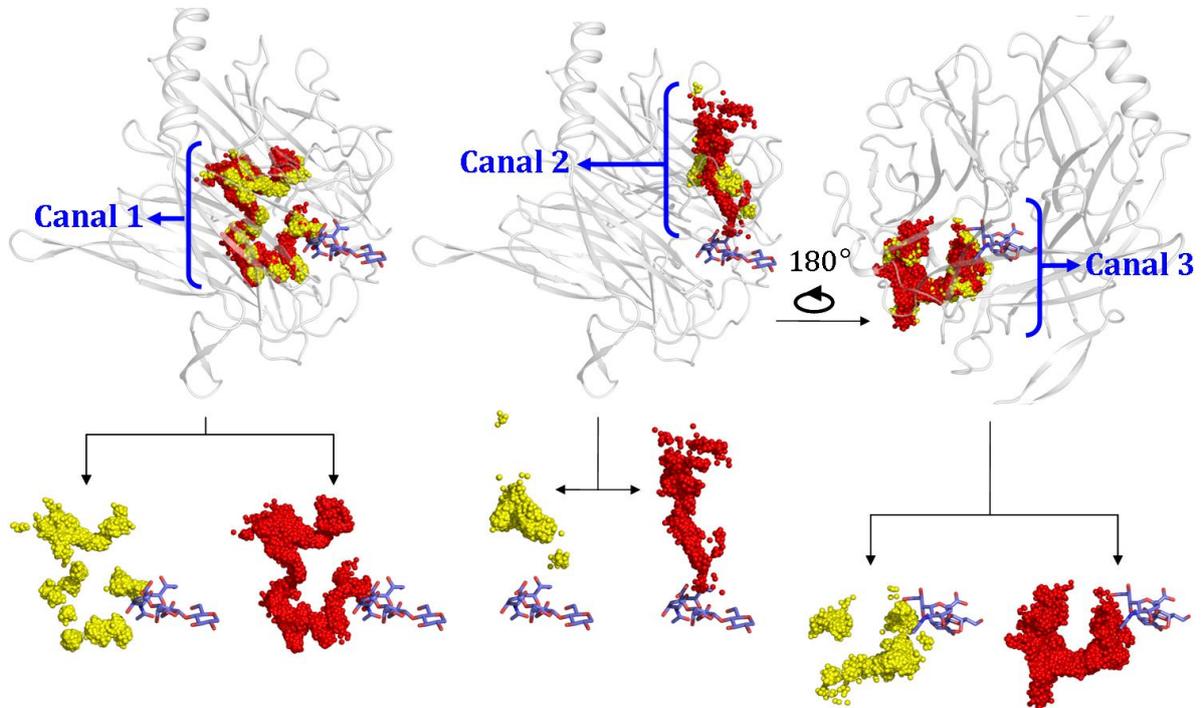


Figure 101. Comparaison de trois canaux du domaine catalytique de TcTS et TrSA.

Le domaine catalytique est représenté en cartoon. Les atomes d'oxygènes des molécules d'eau internes aux canaux chez TcTS et TrSA sont respectivement représentés en sphères de couleur jaune et rouge.

Parmi les quatre canaux d'eau potentiels caractérisés au sein du domaine catalytique, trois d'entre eux (dénommés canaux 1, 2, 3) sont bordés par des résidus répertoriés dans la littérature comme impliqués dans le contrôle de la balance entre hydrolyse et trans-sialidation (*Figures 101 et 102*). Une étude plus approfondie, présentée ci-dessous, sera donc limitée à ces trois canaux exclusivement.

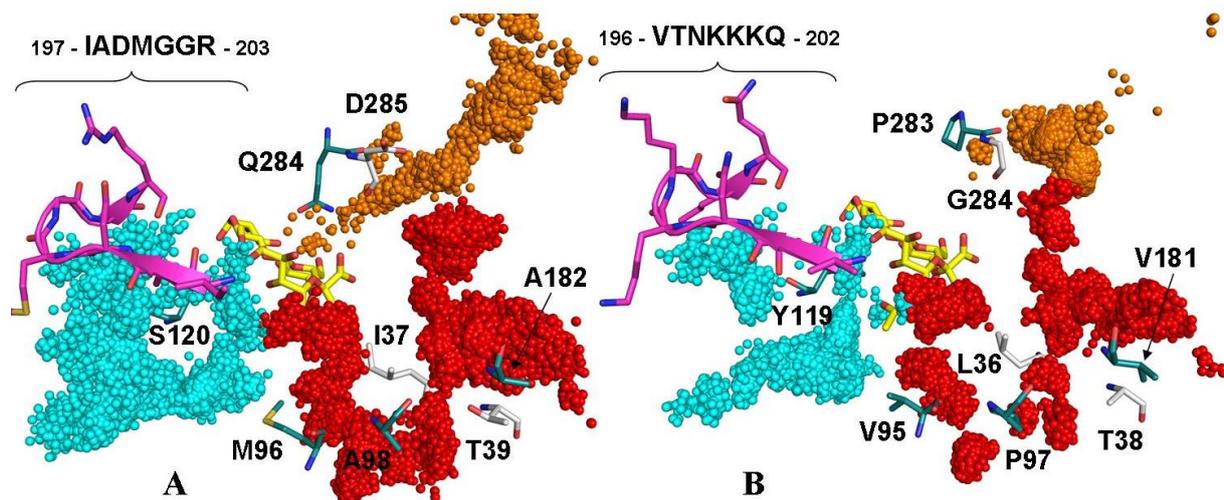


Figure 102. Résidus fonctionnels bordant les canaux d'eau internes à TrSA (A) et TcTS (B).

Les oxygènes des molécules d'eau internes aux canaux d'eau 1, 2, et 3 sont respectivement représentés sous la forme de sphères rouge, orange, et cyan. Les atomes de carbone du substrat donneur (3'-sialyl-lactose) sont illustrés en bâtonnet jaunes. Les atomes de carbone des résidus testés dans les travaux de Paris et coll.²⁵² et Jers et coll.²⁵⁴ sont respectivement illustrés en bâtonnets bleu anthracite et magenta. Les atomes de carbone de trois des cinq résidus prédits comme impliqués dans la balance H/T dans les travaux de Pierdominici-Sottile et coll.²⁵³ sont colorés en gris.

8.5.2 Etude du canal 1

Le plus grand canal, nommé ici canal 1, est notamment interrompu au niveau d'un résidu P95 chez TcTS (*Figure 103*). Muté en alanine chez TrSA, ce résidu est connu pour jouer un rôle dans l'équilibre de la balance entre activité d'hydrolyse et trans-sialidation. Cette proline tient son rôle fonctionnel en raison de la courbure qu'elle provoque sur le squelette peptidique chez TcTS. Cet effet a une influence importante sur la catalyse puisqu'il induit un déplacement de la chaîne latérale du résidu D96 impliqué dans la stabilisation du substrat. Les simulations ont montré que la chaîne latérale de la proline semble par ailleurs exercer une contrainte stérique dans la trajectoire du canal à proximité du site catalytique pouvant ainsi expliquer la fragmentation de celui-ci. Il est d'ailleurs intéressant de noter que trois autres positions mutées (M96, A98, A182) par Paris et coll.²⁵² et connues pour être impliquées dans le contrôle de la balance H/T sont situées en bordure de ce même canal. Des expériences de mutagenèse ciblées sur

d'autres résidus en bordure du canal restent encore à réaliser afin de prouver un rôle potentiel de ce canal dans l'activité de TrSA.

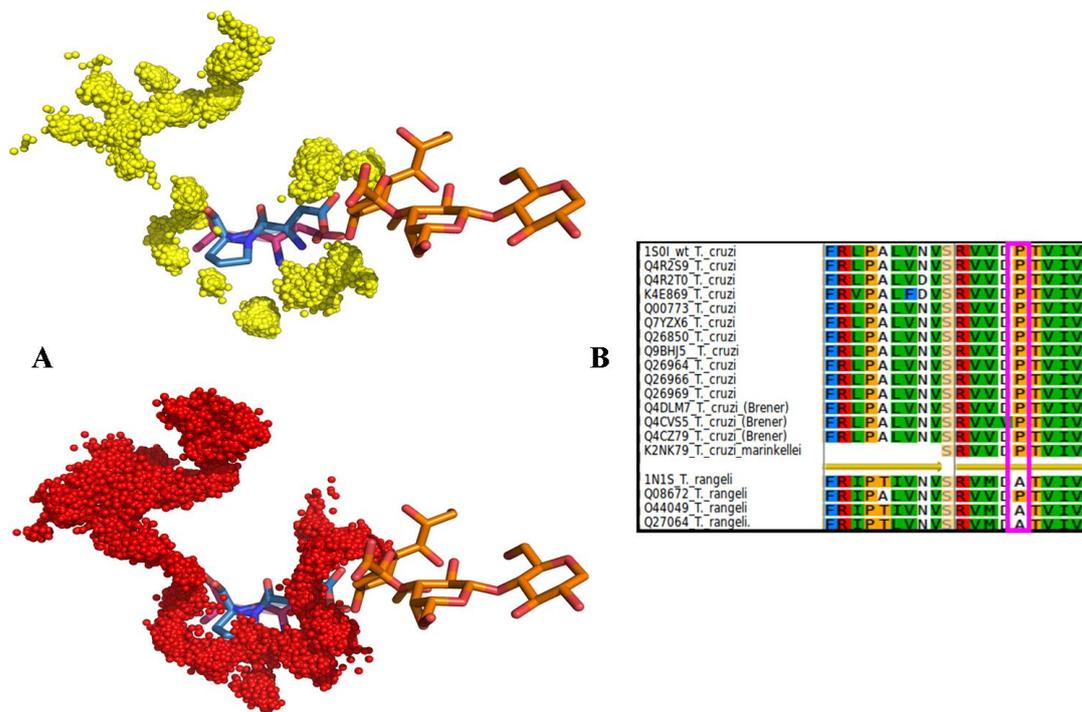


Figure 103. Aperçu du canal 1 et localisation du résidu 96.

Les atomes d'oxygènes des molécules d'eau internes au canal 1 chez TcTS et TrSA sont respectivement représentés en sphères de couleur jaune et rouge. Le résidu 96 est illustré en bâtonnets. En B : Un alignement multiple de séquence de trans-sialidases et sialidases de *T. cruzi* et *T. rangeli* souligne la conservation de ce résidu dont la position est encadrée en magenta. Les atomes de carbone du substrat donneur (3'-sialyl-lactose) est représenté en bâtonnets oranges.

En comparaison aux autres canaux, ce canal est particulièrement remarquable puisqu'il connecte le sous-site -1 au cœur hydrophobe du domaine catalytique et rassemble dans sa trajectoire un nombre important de molécules d'eau cristallographiques. Un alignement structural des 15 structures de (trans)sialidases de *T. cruzi* et *T. rangeli* répertoriées dans la PDB montre qu'environ 51 % de l'ensemble des molécules d'eau cristallographiques prises sur l'ensemble de ces structures sont localisées dans la trajectoire de ce canal. Ces résultats confirment par ailleurs les observations faites par Paris et coll. qui ont suspecté la présence de ce canal en mettant en évidence une chaîne de molécules d'eau cristallographiques dans sa trajectoire.

8.5.3 Etude du canal 2

Les molécules d'eau internes au canal 2 décrivent des trajectoires très différentes entre TcTS et TrSA. Chez TcTS, ces dernières restent en effet en majeure partie confinées au sein d'une poche à mi-chemin entre la surface du domaine catalytique et le sous-site -1. Chez TrSA, l'eau circule en revanche sur toute la longueur ($\sim 17-18 \text{ \AA}$) du canal de la surface du domaine catalytique jusqu'au sous-site -1. Si ce canal débouche bien dans le site actif chez TrSA, son implication dans l'approvisionnement des molécules d'eau catalytiques paraît cependant peu probable puisque les molécules d'eau situées à l'extrémité du canal débouchant dans le sous-site -1 restent trop éloignées du résidu acide-base catalytique.

Les fortes différences observées dans la trajectoire de l'eau du canal 2 chez ces deux enzymes pourraient être dues à la nature de certains acides aminés bordant ce canal. Le motif ³⁰⁵HPLN-₃₀₈, très conservée sur l'ensemble des séquences de sialidases de *T. rangeli*, comporte une histidine (H305) dont la chaîne latérale épouse parfaitement le canal et établit de nombreuses interactions polaires avec l'eau. Cette histidine est également impliquée dans un réseau de liaisons hydrogène avec les résidus T276, S288, R252 et D285 chez TrSA (*Figure 104*). Le résidu D285 établit un grand nombre de liaisons hydrogène avec la chaîne latérale de l'histidine 305 (H304 chez TcTS) et semble même contrôler l'orientation de cette dernière.

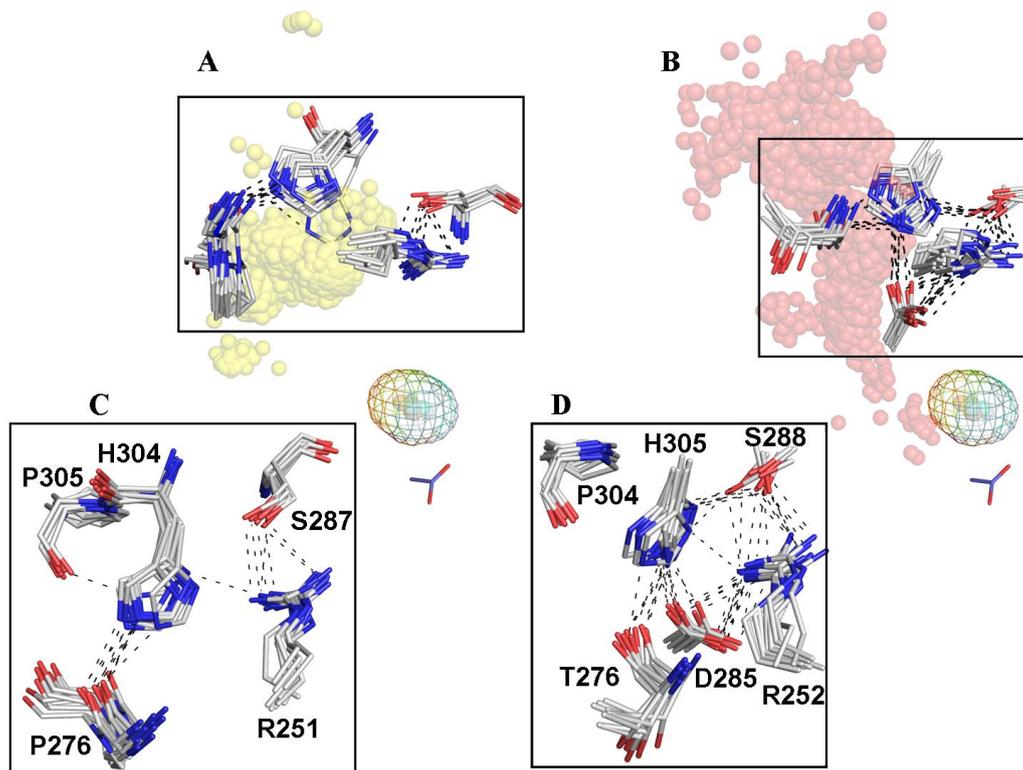


Figure 104. Aperçu du canal 2 et du réseau de liaisons hydrogène impliquant H (304/305).

Les atomes d'oxygène des molécules d'eau internes au canal 2 chez TcTS (A) et TrSA (B) sont respectivement représentés en sphères de couleur jaune et rouge. Les résidus impliqués dans la formation du réseau de liaisons hydrogène chez TcTS (C) et TrSA (D) sont illustrés en bâtonnets et sont issus de la superposition des clichés de trajectoire pris à 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 ns. Les liaisons hydrogène potentielles sont illustrées en noir. Le résidu catalytique acide-base est représenté en bâtonnets dont les atomes de carbone sont colorés en violet. Les positions occupées par l'atome de carbone anomérique du substrat donneur sont représentées en sphères entourées d'un maillage.

Le remplacement du résidu D285 par G dans la séquence de la TcTS sauvage provoque en effet un déplacement de la chaîne latérale du résidu H304 dans la trajectoire du canal caractérisé chez TrSA. Il est par ailleurs intéressant de noter que la substitution D285G avait déjà été testée et contribue à augmenter l'activité de trans-ialidation de TrSA²⁵². La réorientation de la chaîne latérale de cette histidine chez TcTS pourrait également être assistée par le résidu P276, muté en N chez TrSA. Cette proline fait partie d'une boucle portant deux autres prolines chez TcTS (motif₂₇₅-GPSPKS-280) contre une seule chez TrSA (motif₂₇₆-TNSPTS-281). Le cycle de P276 entraîne le déplacement de cette boucle permettant au carbonyle du résidu précédent (G275 chez

TcTS) de se placer à une distance optimale de la chaîne latérale du résidu H304 pour établir un réseau de liaisons hydrogène avec cette dernière (*Figure 105*). Le résidu P276 pourrait donc être essentiel pour assister le changement de conformation de la chaîne latérale de cette histidine. Il est très probable que cette réorientation de la chaîne latérale de cette histidine puisse empêcher la formation du canal observé chez TrSA et du même coup forcer les molécules d'eau à emprunter un autre chemin chez TcTS.

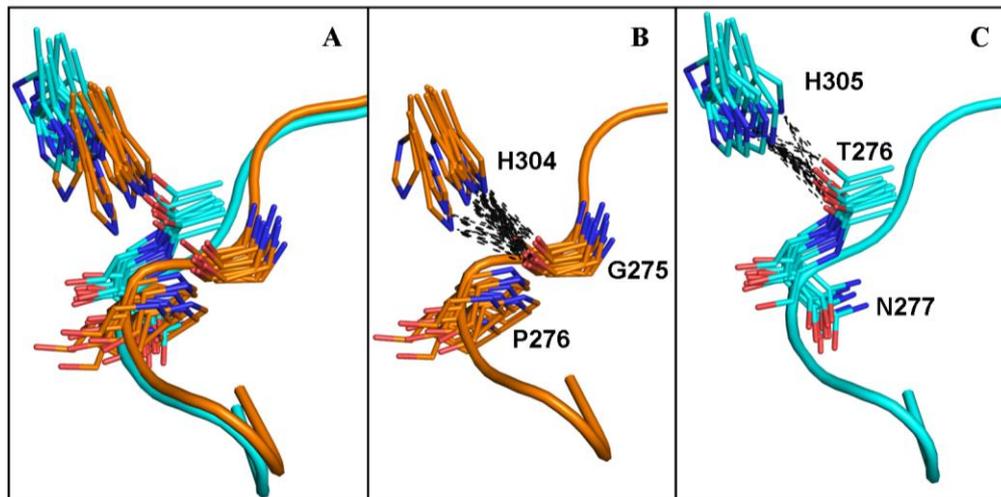


Figure 105. Impact du résidu 276/277 chez TcTS/TrSA sur la stabilisation de H (304/305).

Une superposition de la boucle portant le résidu 276 chez TcTS et 277 chez TrSA est visible en A. La boucle est représentée en cartoon orange chez TcTS et cyan chez TrSA. Les conformations des chaînes latérales représentées en bâtonnets sont issues de la superposition des clichés de trajectoire pris à 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 ns chez TcTS (B) et TrSA (C). Les liaisons hydrogène sont illustrées en noir.

D'autre part, l'observation des trajectoires des chaînes latérales des résidus bordant l'extrémité du canal débouchant sur le sous-site -1 montre que trois d'entre eux, respectivement W312, R314, et A341 (muté en G chez TrSA), contribuent chez TcTS à former un goulet d'étranglement en sortie du canal observé chez TrSA. Chez TcTS, l'encombrement stérique de la chaîne latérale du résidu A341 entraîne le déplacement de celle de l'arginine 314 dans la trajectoire du canal mis en évidence chez TrSA (*Figure 106*). La chaîne latérale du résidu W312, proche de celle de l'arginine 314, semble également former avec celle-ci un goulet d'étranglement qui pourrait empêcher la progression des molécules d'eau du canal 2 dans le sous-site -1 chez TcTS. Chez TrSA, le passage des molécules du canal vers le sous-site -1 semble en revanche favorisé par la

présence du résidu Q284 (muté en P chez TcTS). La chaîne latérale du résidu Q284 exerce en effet une gêne stérique sur la chaîne latérale du résidu W312 de sorte à ce que celle-ci s'écarte de la sortie du canal ¹¹². L'ensemble de ces phénomènes pourrait donc également permettre d'expliquer en partie pourquoi les molécules d'eau internes à ce canal présentent des trajectoires si différentes entre TcTS et TrSA.

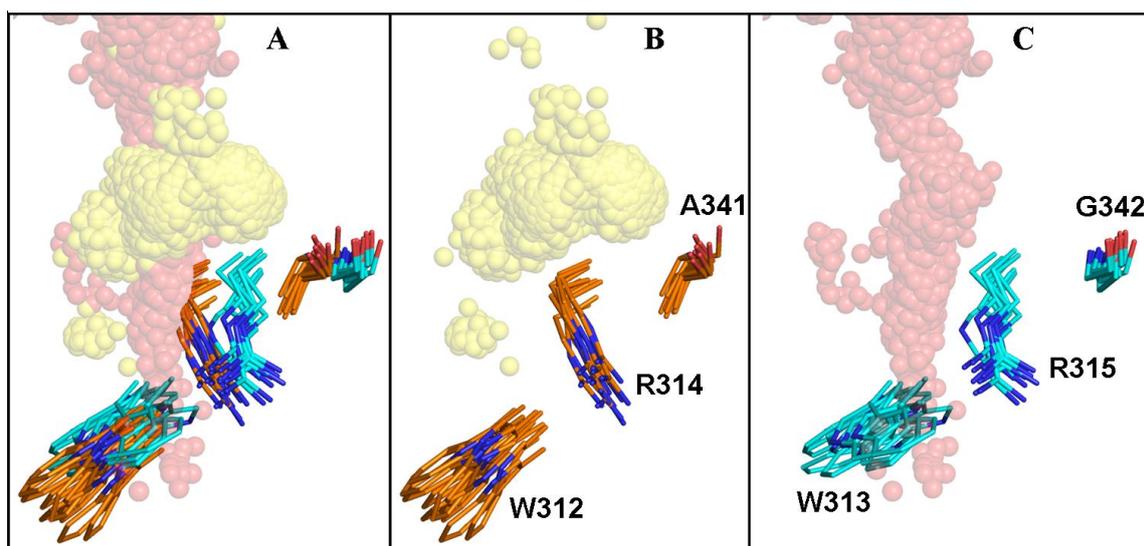


Figure 106. Conformation du résidu R314/315 en sortie du canal 2 chez TcTS/TrSA.

Les atomes d'oxygènes des molécules d'eau du canal isolé chez TcTS et TrSA sont respectivement colorés en jaune et rouge. Les conformations des chaînes latérales représentées en bâtonnets sont issues de la superposition des clichés de trajectoire pris à 0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 ns chez TcTS (B) et TrSA (C). En A : superposition des encadrés B et C.

8.5.4 Etude du canal 3

Le canal 3 connecte également le milieu au site catalytique. Mesurant environ 16 à 18 Å de long, ce canal présente deux extrémités débouchant à la surface du domaine catalytique (*Figure 107*). Par analogie avec les canaux 1 et 2, ce canal présente un aspect fragmenté chez TcTS associé à un fort déficit en molécules d'eau en comparaison à celui mis en évidence chez TrSA. Les extrémités du canal localisées en surface du domaine catalytique sont par ailleurs bordées par des boucles regroupant l'ensemble des résidus mis en évidence par Jers et coll. ²⁵⁴ comme impliqués dans le contrôle de la balance H/T chez TrSA. Chez TcTS, on retrouve donc parmi ces boucles celle porteuse du motif ¹⁹⁷-VTNKKKQ-²⁰³ dont les chaînes latérales aliphatiques des lysines (ainsi que celle du

résidu R127) semblent obstruer les deux extrémités du canal en surface. L'encombrement stérique de ces chaînes latérales pourrait potentiellement diminuer l'accès des molécules d'eau du milieu à l'intérieur du canal et justifier en partie son contenu plus faible en eau chez TcTS.

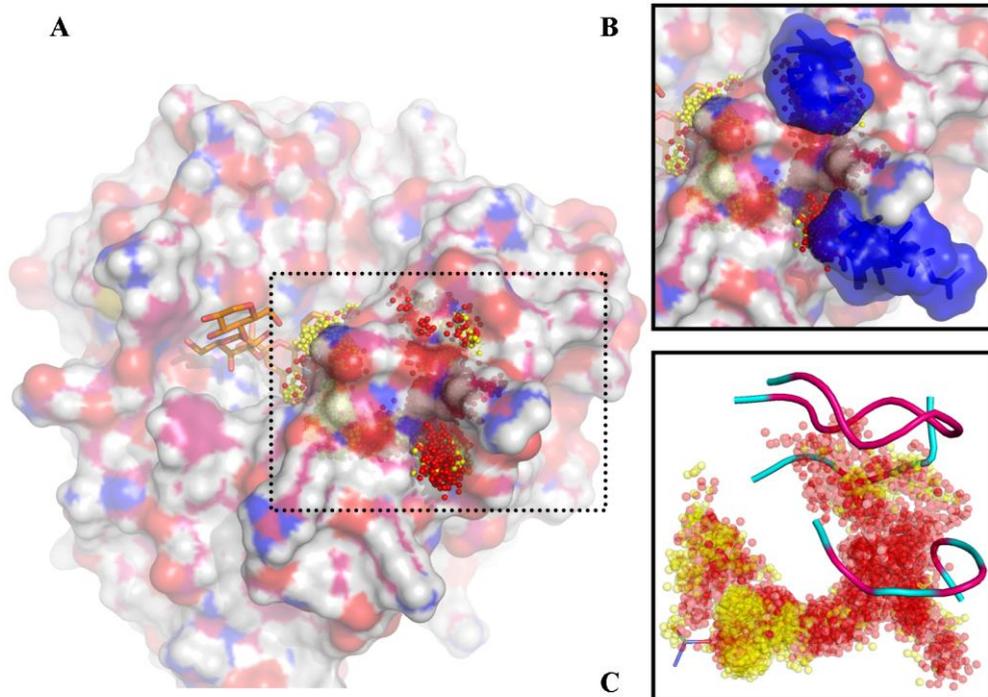


Figure 107. Localisation du canal 3 et des résidus vicinaux répertoriés dans la littérature.

Les atomes d'oxygènes des molécules d'eau internes au canal 3 chez TcTS et TrSA sont superposés et sont respectivement représentés en sphères de couleur jaune et rouge. A : Aperçu du canal au sein du domaine catalytique de TrSA. Le substrat donneur (ici le 3'-sialyl-lactose) est représenté en bâtonnets. B : Aperçu de l'extrémité du canal débouchant sur la surface du domaine. Les résidus lysine du motif VTNKKKQ ainsi que le résidu R127 présents exclusivement chez TcTS sont illustrés en surface et bâtonnets bleu foncé. C : Localisation des résidus mutés par Jers et coll.²⁵⁴ dont le squelette peptidique de TrSA est coloré en magenta.

La fragmentation du canal chez TcTS est provoqué par la formation d'un réseau de liaisons hydrogène entre trois résidus, respectivement H123, Q174, et Y113 (*Figure 108*). La chaîne latérale du résidu Y113 semble jouer un rôle essentiel dans la formation de ce réseau puisque son interaction avec la chaîne latérale du résidu Q174 permet à cette dernière de contacter à son tour celle du résidu H123 et de l'orienter dans la trajectoire du canal. L'encombrement stérique de la chaîne latérale de cette histidine

provoque ainsi une fragmentation du canal par effet d'exclusion des molécules d'eau autour de son noyau imidazole. Le résidu Y113 chez TcTS est remplacé par une phénylalanine chez TrSA. L'absence de groupement hydroxyle sur le noyau de la phénylalanine rend impossible toute formation de liaisons hydrogène avec la glutamine. En conséquence, la conformation de chaîne latérale de l'histidine responsable de la fragmentation du canal chez TcTS est elle-même affectée et tend à s'écarter de la trajectoire du canal. Cet effet permet ainsi aux molécules d'eau de circuler sur toute la trajectoire du canal chez TrSA. Au vu des conséquences importantes qu'elle semble induire sur la dynamique de l'eau interne à ce canal, la mutation Y113F pourrait être intéressante à tester chez TcTS.

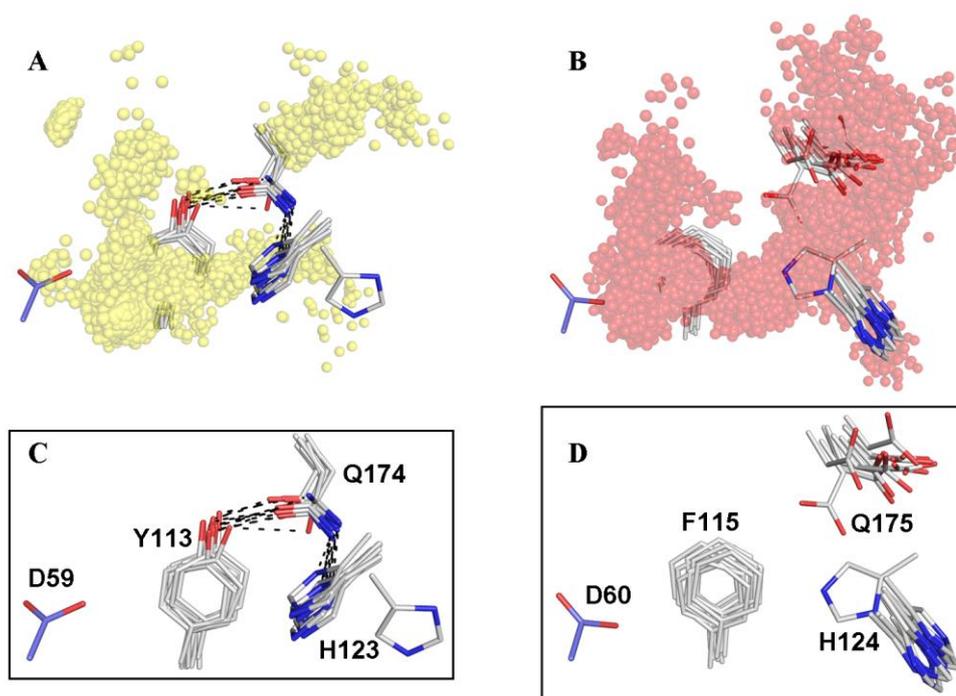


Figure 108. Aperçu du canal 3 et du réseau de liaisons hydrogène impliquant H (123/124).

Les atomes de carbone du résidu acide-base catalytique sont représentés en bâtonnets violets. Les liaisons hydrogènes sont illustrées en noir. A : *T. cruzi*, B : *T. rangeli*.

8.5.5 Diffusion de l'eau interne au domaine catalytique

Afin de quantifier la dynamique des molécules d'eau internes aux canaux mis en évidence, un calcul du déplacement quadratique de l'eau interne au domaine catalytique

a été réalisé sur l'ensemble de la trajectoire. L'allure des profils de déplacements des atomes d'oxygènes de molécules d'eau montre des différences significatives entre TcTS et TrSA (Figure 109).

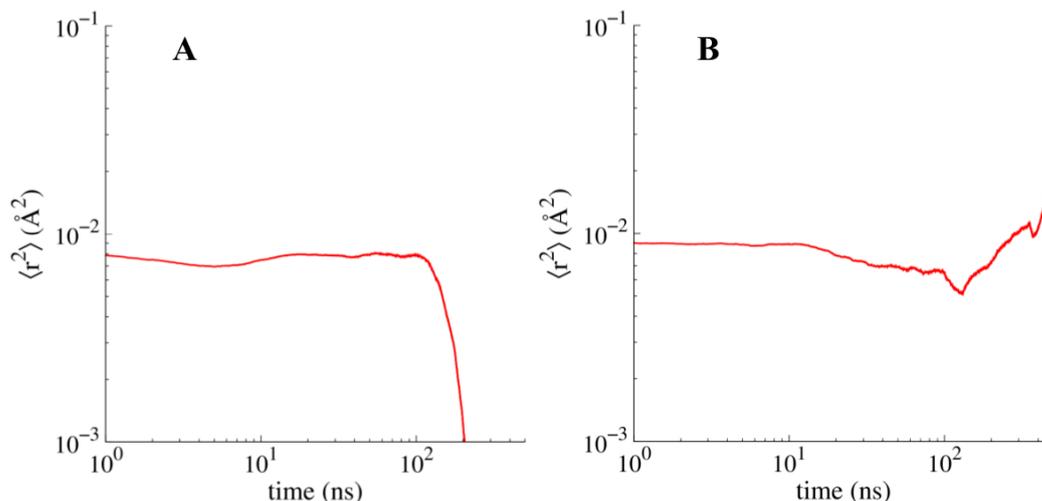


Figure 109. Déplacement quadratique moyen des molécules d'eau internes au domaine catalytique de TrSA et TcTS. Les résultats obtenus chez TrSA et TcTS sont représentés en A et B, respectivement.

Chez les deux enzymes, le faible déplacement quadratique moyen des molécules d'eau au cours des cent premières nanosecondes de la trajectoire traduit une cinétique de diffusion lente. Cet effet pourrait être dû à l'internalisation des molécules d'eau internes au sein des cavités hydrophobes du domaine catalytique. Ces cavités hébergent par ailleurs la quasi-totalité des molécules d'eau constituant les différents canaux d'eau mis en évidence dans cette étude. Il a par ailleurs été montré précédemment que le canal 1 abritait la majeure partie des molécules d'eau internes au domaine catalytique au cours des simulations. Ce canal s'étend sur presque toute la longueur du domaine et, de ce fait, est le plus enfoui de tous les canaux caractérisés. Au vu de la présence d'un grand nombre de résidus apolaires en bordure de ce canal, il paraîtrait donc logique que la majeure partie des molécules d'eau qui le constitue reste confinées en son sein et voient ainsi leur vitesse de diffusion réduite en comparaison en molécules d'eau de la poche de liaison au substrat, plus fortement exposées au milieu.

Chez TcTS, la très forte décroissance du déplacement quadratique moyen des molécules d'eau interne au domaine catalytique au-delà de 100 ns de trajectoire montre que la diffusion des molécules d'eau internes est plus fortement ralentie en comparaison

à TrSA. Ce ralentissement suggère que les chaînes latérales des résidus enfouis en contact avec les molécules d'eau internes pourraient contraindre leur progression au sein du domaine catalytique. Cette hypothèse est appuyée par l'aspect fragmenté des canaux caractérisés chez TcTS qui montre que la circulation de l'eau interne au domaine catalytique de TcTS semble plus contrainte que chez TrSA. Il a par ailleurs été montré précédemment que ce phénomène était lié à une modification de la dynamique des chaînes latérales de résidus spécifiques bordant ces mêmes canaux. Ces résultats sont également confortés par les travaux de Demir et Roitberg²⁴⁶ qui montraient que la présence du substrat donneur au sein du site catalytique de TcTS réduisait considérablement sa flexibilité ainsi que son accessibilité à l'eau du milieu en comparaison à TrSA. De tels effets pourraient expliquer l'aspect fragmenté des canaux chez TcTS et à fortiori les différences de diffusions observées.

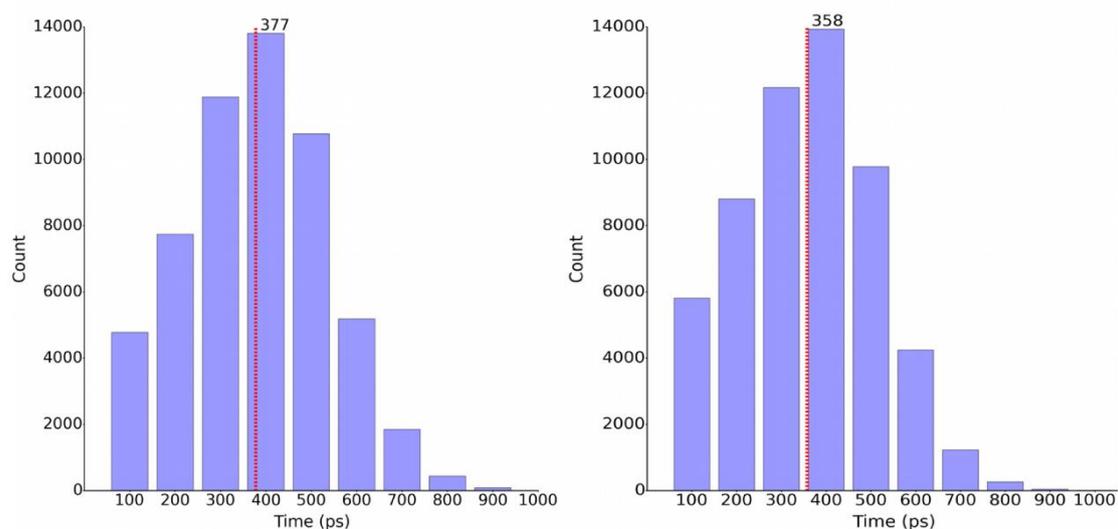


Figure 110. Distribution du temps mis par les molécules d'eau internes à TcTS et TrSA pour parcourir une distance d'au minimum 3 Å au sein de la protéine.

En A: TcTS, B: TrSA. La valeur moyenne est indiquée par une ligne verticale rouge.

Le temps moyen nécessaire aux molécules d'eau interne pour parcourir une distance d'au minimum 3 Å au sein du domaine catalytique reste cependant sensiblement identique chez les deux enzymes (Figure 110). Ce résultat montre donc que la capacité des molécules d'eau internes à diffuser au-delà d'une courte distance ne semble en moyenne pas affectée par les différences liées à la dynamique des deux protéines.

8.6 Conclusions

Au vu de la forte accessibilité au solvant des sites catalytiques des deux enzymes, et bien qu'il ait été montré que le site actif de TcTS semblait moins exposé à l'eau du milieu que celui de TrSA, il est difficile de dire si un de ces canaux pourrait être impliqué dans l'approvisionnement de molécules d'eau catalytique au sous-site -1. Par analogie avec Ttβgly, la simulation des intermédiaires covalents sialyl-enzyme permettrait d'isoler les molécules d'eau potentiellement catalytiques et d'identifier leurs origines respectives. Les différences de trajectoire des molécules d'eau internes aux canaux 1, 2, et 3 entre TcTS et TrSA traduisent cependant une modification générale de la dynamique de l'eau au sein du domaine catalytique. Puisque la dynamique des molécules d'eau en contact avec la protéine influence très souvent la dynamique de cette dernière, il est possible que ces effets aient des conséquences significatives sur la dynamique de la protéine et puissent expliquer les différences observés précédemment entre TcTS et TrSA. Chez TcTS, il est également probable que la fragmentation et le déficit en eau des canaux puissent affecter les réseaux d'interactions électrostatiques structurant le domaine catalytique ainsi que celles connectant des régions distantes du domaine au site catalytique. Les modifications observées sur la dynamique de l'eau interne au domaine catalytique rejoignent donc l'hypothèse suggérée par Jers et coll.²⁵⁴ Une caractérisation complète du réseau de liaisons hydrogène stabilisant les résidus de la machinerie catalytique serait nécessaire pour vérifier si ces modifications sont responsables des différences de stabilité des chaînes latérales des résidus catalytiques D59 et E230 chez TcTS. La validation de cette hypothèse permettrait alors de confirmer le rôle fonctionnel de la dynamique de l'eau interne aux canaux dans la catalyse de ces deux enzymes.

8.7 Identification de nouveaux résidus contrôlant la balance H/T

A ce jour, à peine plus de 10 % de l'activité de trans-sialidation de TcTS a pu être restitué chez TrSA *via* l'incorporation de 16 mutations²⁵⁷ parmi l'ensemble de celles proposées par les études menées par Paris et coll.²⁵², Jers et coll.²⁵⁴, et Pierdominici-Sottile et coll.²⁵³ Les résultats de ces travaux suggèrent ainsi que d'autres résidus doivent influencer la balance H/T chez TrSA et laissent ainsi entrevoir de nombreuses autres possibilités de mutagenèses. Dans le but de guider le choix de nouvelles

mutations à réaliser, un alignement multiple de deux jeux de séquences homologues de trois sialidases de *T. rangeli* et de cinq trans-sialidases de *T. cruzi* a été réalisé (Figure 111).

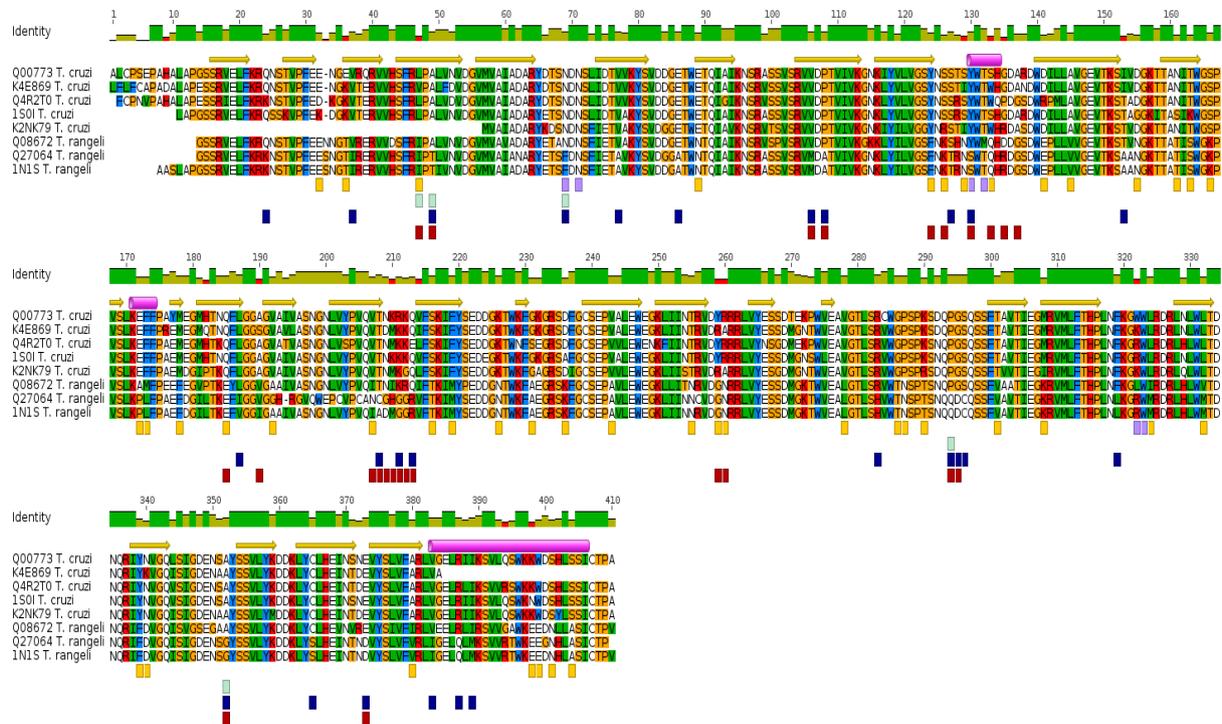


Figure 111. Alignement multiple de séquences de sialidases et trans-sialidases de *T. rangeli* et *T. cruzi* (issues de la base de données UniProtKB).

Les éléments de structure secondaire de la structure protéique de TcTS (PDB 1S01) sont représentés au dessus de l'alignement. Les résidus sont colorés selon leur classe physico-chimique. Sous l'alignement, les rectangles jaunes indiquent les positions divergentes à plus de 90 % entre l'ensemble des séquences de trans-sialidase de *T. cruzi* et celles de sialidases de *T. rangeli*. Les rectangles rouges indiquent les positions testées par mutagenèse au sein des travaux de Paris et coll.²⁵² et Jers et coll.²⁵⁴ Les rectangles violets et cyan indiquent les positions mutées par Amélie Saumonneau (UFIP, Equipe d'ingénierie moléculaire et glyco-biologie) et celles suggérées par les prédictions de Pierdominici-Sottile et coll.²⁵³ Les rectangles bleus signalent les positions identiques entre l'ensemble des séquences de trans-sialidases de *T. cruzi* et la séquence de la sialidase de *T. rangeli* Q08672. L'alignement a été réalisé via le logiciel MUSCLE¹⁹⁰ et visualisé à l'aide du logiciel Geneious.

8.7.1 Recherche de résidus fonctionnels en bordure des canaux d'eau

Cet alignement multiple a permis d'identifier un total de 44 positions divergentes à plus de 90 % d'identité de séquence entre l'ensemble des séquences de *T. cruzi* et celles de *T. rangeli* alignées (*Tableau 8*). Il est possible que la réversion des 44 aminoacides correspondants sur la séquence de TrSA par ceux trouvés chez TcTS puisse permettre de restaurer chez TrSA une activité de trans-sialidation comparable à celle de TcTS. Si le choix d'un seuil de divergence élevé est intéressant pour limiter le nombre de positions à tester par mutagenèse, il est cependant possible que le seuil choisi (90 %) pour l'identification des positions divergentes entre ces deux groupes de séquences puisse être trop stringent.

En effet, sur l'ensemble des mutations connues pour augmenter l'activité de trans-sialidation chez TrSA (*Figure 112*), seules deux (I37L, G249Y) des six proposées par Paris et coll.²⁵² et une (I197V) des sept suggérées par Jers et coll.²⁵⁴ font partie des 44 substitutions identifiées. Bien que sept positions identifiées parmi celles-ci aient déjà été testées par mutagenèse, 37 positions restent encore à être examinées. Afin de limiter le nombre de mutations à tester, seules les positions divergentes à plus de 90 % entre l'ensemble des séquences de sialidases de *T. rangeli* et de trans-sialidases de *T. cruzi* et dont la chaîne latérale du résidu correspondant est en interaction (distante de moins de 3,5 Å de l'atome d'oxygène d'une molécule d'eau) avec des molécules d'eau internes aux canaux identifiés ont été considérées parmi les 37 positions initiales.

# PDB	TcTS MSA	TrSA MSA	d. D59	références	MD TcTS	MD TrSA
22	E, k, d	E	16.49	-	-	-
25	E, K	T	24.0	-	-	-
36	L, v	I	7.36	Pierdominici-Sotille et coll.	1 : 9521	1 : 13818
78	E	N	27.05	-	-	-
113	Y	F	6.38	Paris / Jers et coll.	3 : 11959	3 : 22404
115	S, r	K	12.47	-	3 : 391	3 : 835
118	S, l	N	9.6	-	3 : 217, 4 : 2	4 : 1692, 3 : 59
122	S,W	Q	9.35	Paris / Jers et coll.	3 : 1927	3 : 3702, 4 : 2009, 2 : 1
130	D, r	E	15.73	-	-	-
134	A	V	17.89	-	-	-
144	D, g	N	41.18	-	-	-
150	N, s	T	30.16	-	-	-
152	T, k	S	27.56	-	-	-
155	S	K	22.89	-	-	-
161	E	P, a	23.53	-	-	-
162	F	L, m	22.28	-	-	-
167	M	F	19.6	-	-	-
174	Q	E	8.87	Paris / Jers et coll.	3 : 13364	3 : 49714
181	V	A, g	18.46	-	1 : 6612	1 : 11368
196	V	l, n	15.08	Paris / Jers et coll.	-	-
205	S	T	15.39	-	-	-
208	F	M	18.06	-	-	-
215	K	N	26.85	-	-	-
220	G, s	A	23.47	-	-	-
225	D, a	K	21.43	-	-	5 : 4053
232	V	A	17.5	-	1 : 34462	1 : 17994
244	T	N	17.81	-	-	-
248	Y, R	G	13.16	Paris / Jers et coll.	-	-
249	R, A	N	19.81	Paris / Jers et coll.	-	-
267	V	L	29.91	-	5 : 5326	5 : 11688
275	G	T	20.77	-	2 : 65	2 : 3742
276	P	N	17.49	-	2 : 6210	2 : 2
279	K, r	T	25.47	-	-	-
290	T	V	22.71	-	1 : 7254	1 : 10065
297	M, i	K	37.23	-	-	-
313	L	M, i	13.11	-	4 : 1198	4 : 471
321	L	M, v	31.54	-	-	-
328	Y	F	32.82	-	5 : 62	5 : 153
329	N, k	D	27.39	-	-	-
369	A	V, i	23.96	-	-	-
387	K, n	E	46.39	-	-	-
388	W	E	44.96	-	-	5 : 6
390	S	N	47.44	-	-	-
393	S	A	48.09	-	-	-

Tableau 8. Liste des positions divergentes entre les séquences de sialidases (SL) et de trans-sialidases (TS) de *T. rangeli* et *T. cruzi*. Les positions listées ont été identifiées à partir de l'alignement multiple présenté en figure 111. Il s'agit des positions pour lesquelles on observe une divergence d'au moins 90 % entre l'ensemble de séquence de SL de *T. rangeli* et des séquences de TS de *T. cruzi*. 90 % de divergence sur une position signifie que ces deux blocs de séquences ne partagent pas plus que 10 % de résidus en commun (soit 1 résidu maximum) sur une position commune dans l'alignement. La colonne « # PDB » indique l'identifiant numérique de chacune des positions divergente (à plus de 90%) au sein de la structure PDB de TcTS (numérotation de 1S0I). Les colonnes « TcTS MSA » et « TrSA MSA » regroupent respectivement les différents résidus pouvant être trouvés sur chacune de ces positions dans la séquence de TcTS et de TrSA. La quatrième colonne indique la distance relative (en Å) de chaque position divergente au résidu acide-base catalytique. La colonne « références » indique quelles positions ont déjà été expérimentalement testées ou prédites en référant les auteurs des travaux respectifs. Les colonnes « MD TcTS » et « MD TrSA » quantifient le nombre de contacts (en bleu) entre la chaîne

latérale du résidu localisé sur une position donné et l'eau d'un canal donné (l'identifiant du canal, de 1 à 5, est indiqué en rouge) chez TcTS et TrSA complexées au 3'-sialyl-lactose sur 500 ns de simulation de dynamique moléculaire.

La comparaison des trajectoires des chaînes latérales des résidus correspondants chez TcTS et TrSA montre que huit d'entre eux présentent des différences significatives d'exposition aux molécules d'eau internes aux canaux identifiés (Tableau 8). Il est donc possible que ces résidus puissent moduler la circulation des molécules d'eau au sein de ces canaux et être ainsi à l'origine des différences de dynamique observées entre les deux enzymes. Dans le cadre de ce projet, trois de ces positions (S115, S118, L313) ont été testées par mutagenèse.

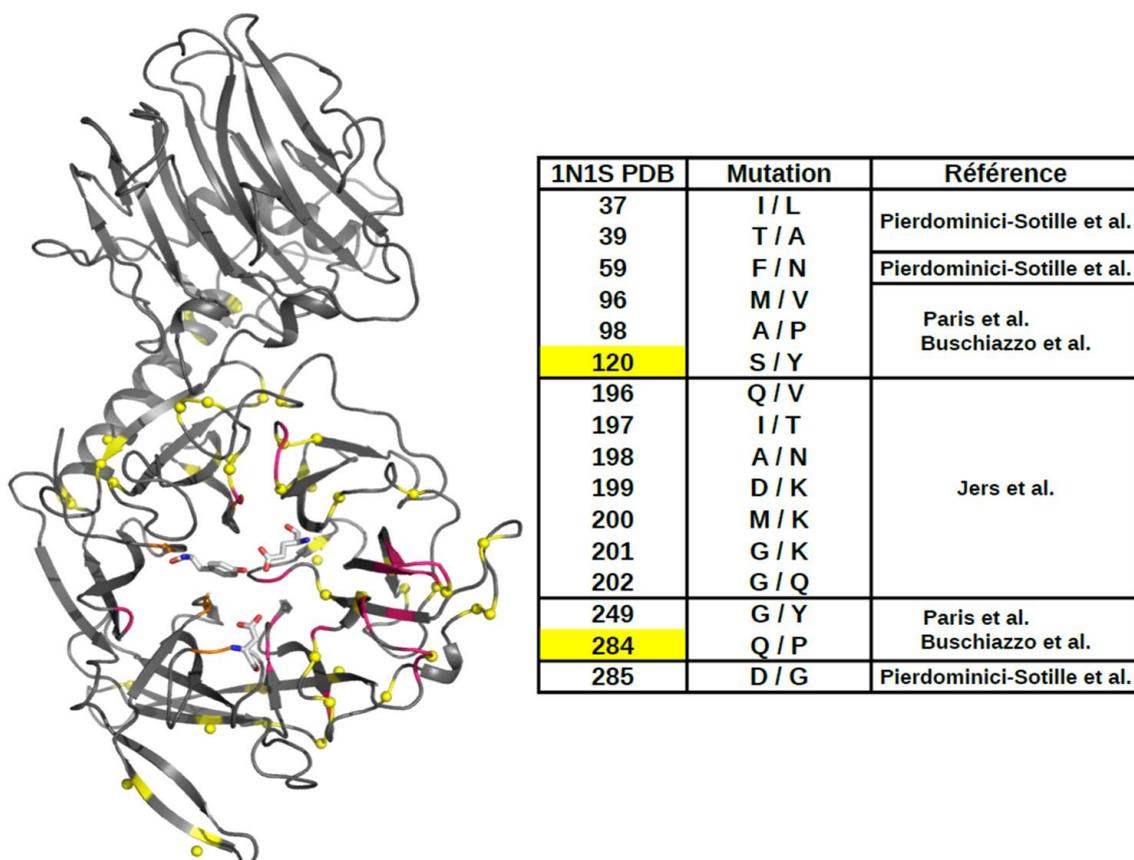


Figure 112. Cartographie des positions mutées et variables entre TcTS et TrSA.

L'ensemble des mutations réalisées sur la séquence de TrSA et répertoriées dans la littérature sont listées dans le tableau ci-contre. La colonne « 1N1S PDB » indiquent les numéros des positions correspondantes suivant la numérotation du fichier PDB 1N1S (structure de référence de TrSA en complexe avec le sialyl-lactose). Les positions surlignées en jaune mettent en valeur les deux résidus les plus impliqués dans le contrôle de la balance H/T. Les chaînes latérales des résidus de la triade

catalytique sont représentées en bâtonnets. Les sphères jaunes indiquent les carbones α des positions variables à plus de 90 % entre TcTS et TrSA mis en évidence à partir de l'alignement multiple présenté en figure 111. Le squelette peptidique des résidus mutés par Paris et coll. et Jers et coll. est coloré en magenta, tandis que celui des résidus prédits par Pierdominici-Sotille et coll. est coloré en orange.

8.8 Mutagenèse des résidus S115, S118, et L313

Afin de vérifier si les résidus S115, S118, et L313 pourraient être impliqués dans le contrôle entre les activités d'hydrolyse et de transfert chez TcTS, leur substitutions par les résidus correspondants chez TrSA a été réalisée. Bordant les canaux 3 et 4 (Figure 113), ces trois résidus ont été retenus en priorité parmi les 37 autres proposés du fait de leur différence de nature physicochimique en comparaison à leurs homologues respectifs chez TrSA (K115, N118, et L313). A l'inverse de quasiment tous les travaux publiés à ce jour qui se focalisaient sur l'ingénierie d'une activité trans-sialidase chez TrSA, une tentative de conversion de TcTS en sialidase a ici été entreprise. Le choix de travailler sur TcTS à la place de TrSA était motivé par le fait que seul le gène synthétique codant la forme sauvage de TcTS était disponible au laboratoire.

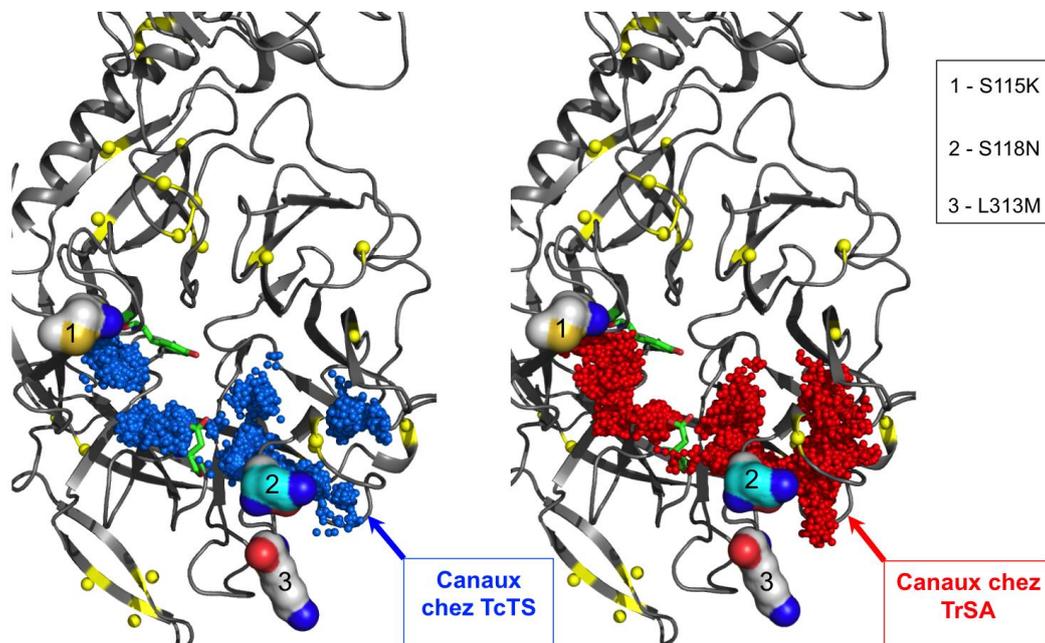


Figure 113. Localisation des résidus S115, S118, et L313 sur le domaine catalytique.

Les chaînes latérales des résidus mutés sont représentées en surface. Les résidus catalytiques acide-base et nucléophile sont illustrés en bâtonnets. Les sphères jaunes indiquent les carbones α des

positions variables à plus de 90 % mis en évidence à partir de l'alignement multiple présenté en figure 111. Les sphères bleues et rouges indiquent respectivement les atomes d'oxygènes des molécules d'eau des canaux d'eau en contact avec les résidus ciblés.

8.8.1 Caractérisation des produits de réaction par chromatographie

La caractérisation des mutants consiste donc à vérifier si l'introduction d'au moins une des trois mutations testées permettait d'augmenter l'activité d'hydrolyse chez TcTS au détriment de son activité de transfert. Afin d'estimer rapidement les conséquences des mutations introduites sur le rapport d'activité hydrolyse/transglycosylation, les mutants ainsi que l'enzyme sauvage ont été testés en présence de 3'-sialyl-lactose (3'SL) et de lactose en tant que sucre donneur et accepteur de groupement sialyl, respectivement. La séparation du milieu réactionnel par chromatographie montre chez toutes les enzymes l'apparition d'un composé de poids moléculaire supérieur à celui des substrats de départ (*Figure 114*). Ce composé est un produit de transglycosylation dont la formation s'explique par le transfert du groupement sialyl du 3'SL sur l'extrémité réductrice du lactose accepteur accompagnée de la libération du lactose issu du 3'SL en tant que groupe partant. L'obtention d'un unique produit de transglycosylation chez l'ensemble des mutants testés montre également que les mutations n'altèrent pas la régiospécificité de l'enzyme sauvage.

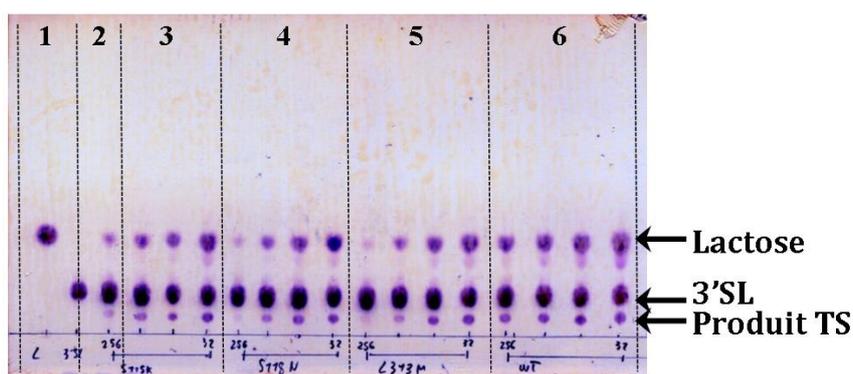


Figure 114. Caractérisation de l'activité des mutants de TcTS.

1 : Lactose (témoin de la réaction d'hydrolyse)

2 : 3'SL (substrat donneur)

3, 4, 5, 6 : produits de réaction des mutants S115K, S118N, L313M, et de l'enzyme sauvage (TcTS)

Pour les pistes de 3 à 6, quatre essais d'activités ont été réalisés avec quatre dilutions décroissantes de l'enzyme (au 1/256, 1/128, 1/64, et 1/32) afin de mimer un suivi cinétique. Produit TS indique le(s) produit(s) de transglycosylation formé(s).

Bien que l'ensemble des mutants soient actifs, leur activité globale respective semble significativement réduite par rapport à l'enzyme sauvage. La nature des chaînes latérales de ces résidus semble donc être importante pour l'efficacité de la catalyse. En plus de diminuer l'activité globale, l'introduction de ces mutations n'a cependant pas permis d'augmenter l'activité d'hydrolyse. Afin de confirmer ces résultats et de quantifier la perte d'activité globale chez ces mutants, un suivi de la réaction enzymatique par spectrophotométrie d'absorption a été effectué.

8.8.2 Détermination des constantes cinétiques de la réaction

En vue de pouvoir suivre la réaction par cette technique, l'utilisation d'un substrat présentant un groupement absorbant dans le visible a été requise. Les substrats donneur et accepteur utilisés précédemment ont donc été respectivement remplacés par le sialyl-PNP et de lacto-N-tétraose (LNT). Le suivi de l'absorbance du milieu réactionnel au cours du temps confirme bien que les mutations réduisent significativement les performances cinétiques de l'enzyme (*Figure 115*). La diminution de l'efficacité catalytique (K_{cat}/K_m) des mutants suggère que la cinétique de la première étape du mécanisme réactionnel a été affectée (*Tableau 9*).

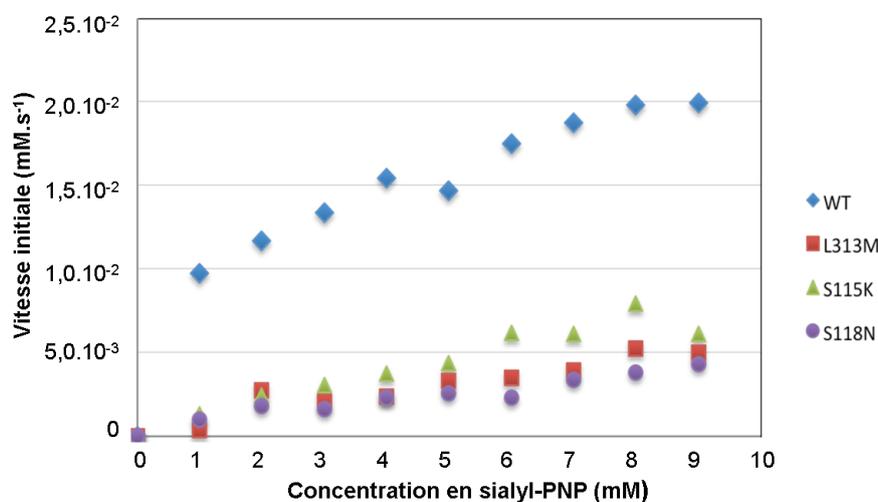


Figure 115. Profil cinétique de TcTS et de ses mutants.

Les tests ont été effectués en présence de concentrations croissantes en sialyl-PNP comme seul substrat de la réaction. Ce substrat joue donc à la fois le rôle de donneur et d'accepteur de groupement sialyl pour la réaction de transglycosylation.

Les résultats obtenus ont montré que les résidus S115, S118, et M313 ne sont pas impliqués dans le contrôle de la balance H/T. Ils semblent en revanche nécessaires au maintien des performances catalytiques de TcTS. Au vu de leur distance importante du sous-site -1, il est possible qu'ils affectent indirectement la catalyse en jouant un rôle dans la stabilisation du domaine catalytique. Une simulation de dynamique moléculaire de ces mutants pourrait permettre de confirmer cela. De part leur nature polaire, les résidus S115 et S118 pourraient participer à la formation d'un réseau de liaisons hydrogènes stabilisant le sous-site -1. Si l'effet recherché n'a pas été obtenu (augmentation de l'activité d'hydrolyse), ces expériences ont néanmoins permis de mettre en évidence de nouveaux résidus influençant la catalyse chez TcTS. Il n'est donc pas exclu qu'au moins quelques unes des 34 autres positions mises en évidence *in-silico* puissent également affecter la catalyse de cette enzyme.

	k_{cat}/K_m ($s^{-1}.M^{-1}$)	k_{cat} (s^{-1})	K_m (mM)	V_m (mM. s^{-1})
Enzyme sauvage	0.77	2.13	2.77	0.021
S115K	0.13			
S118N	0.10			
L313M	0.04			

Tableau 9. Constantes cinétiques de TcTS et de ses mutants.

Ces paramètres ont été déterminés à partir du suivi cinétique présenté en figure 115. Les constantes cinétiques K_m , K_{cat} , et V_m n'ont pas pu être déterminées pour les mutants car leurs cinétiques ne suivent pas la loi de Michaëlis-Menten.

8.9 Vers la reconstruction de l'ancêtre commun à TcTS et TrSA

En vue d'obtenir une meilleure compréhension des bases moléculaires régissant l'équilibre entre activité d'hydrolyse et de transglycosylation au sein des glycosides hydrolases trypanosomales (GH33), une analyse bioinformatique de la totalité des séquences de domaines catalytiques de (trans)sialidases de trypanosomes répertoriées

dans les bases de données UniProt/Interpro a été implémentée en vue d'identifier de nouveaux résidus susceptibles de jouer un rôle dans la régulation de la balance hydrolyse/trans-sialidation chez ces enzymes. La consolidation d'un ensemble non-redondant de séquences obtenues sur différentes bases de données a ainsi été réalisée. En complément, ce jeu de données a également été supplémenté par l'ensemble des séquences de domaines catalytiques des trans-sialidases des trypanosomes *T. brucei*^{195, 196}, *T. congolense*^{197, 198, 199, 200, 201}, et *T. vivax*²⁰² caractérisées biochimiquement à ce jour et répertoriées dans la littérature.

Une distribution statistique de la taille des séquences sélectionnées affiche un aspect bimodal montrant que les tailles de séquences les plus fréquemment retrouvées oscillent entre 400 et 500 résidus (*Figure 116*). Considérant que toutes les séquences de domaines catalytiques de trans-sialidases trypanosomales caractérisées présentent souvent une taille comprise entre 300 et 400 résidus, l'ensemble des séquences du jeu de données de départ présentant une taille inférieure à 300 résidus ont été éliminées. Après élimination des séquences (pseudogènes potentiels) ne possédant pas de machinerie catalytique (absente ou incomplète), le jeu de séquences résultant a été ensuite servi de support pour la construction d'un alignement multiple à partir duquel un arbre phylogénétique a été réalisé dans le but de caractériser les relations évolutives entre ces séquences (*Figure 117*).

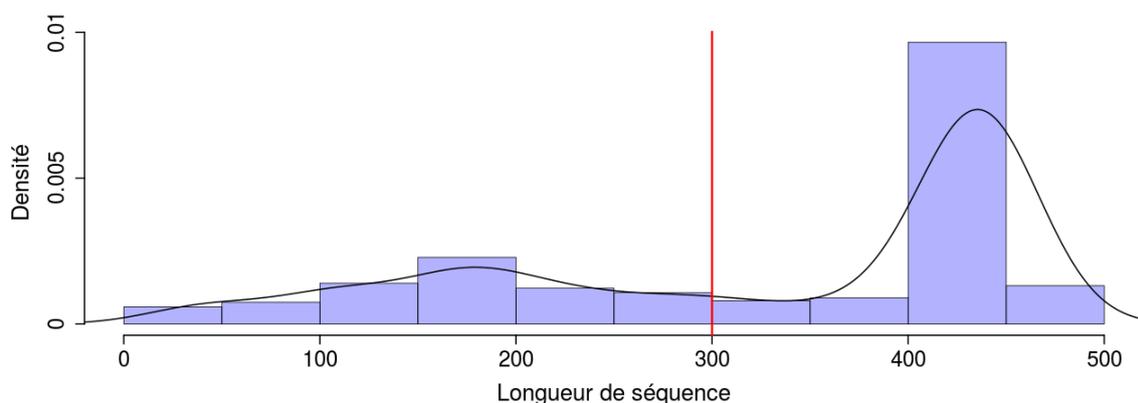


Figure 116. Distribution de longueur des séquences de domaines catalytiques de (trans)sialidases de trypanosomes. L'histogramme est représenté sous la forme d'une densité de probabilité normalisée à 1. La limite de taille (300 résidus) imposée en dessous de laquelle toute séquence de taille inférieure a été rejetée est indiquée en rouge.

Elucider les bases moléculaires régissant l'équilibre entre les activités d'hydrolyse et de trans-sialidation entre TcTS et TrSA revient à identifier les mutations qui ont conduit l'activité de ces deux enzymes à diverger au cours de l'évolution. La structure de TcTS présente par ailleurs de très fortes similitudes avec les sialidases bactériennes qui sont pourtant connues pour être exclusivement hydrolytiques et qui appartiennent à un clade distinct de celui des (trans)sialidases trypanosomales. Partageant en outre une architecture commune, on y retrouve également la même triade catalytique, une triade d'arginines chargées de stabiliser le groupement sialyl, ainsi que la présence de résidus hydrophobes conservés (V95, L176, W120) impliqués dans la formation de la poche de liaison de ce même groupement. Ces similarités laissent alors penser que l'activité trans-sialidase de TcTS pourrait être un caractère dérivé acquis chez une sialidase bactérienne ancestrale qui aurait ensuite été transmis aux trypanosomes par transfert horizontal de gène.

Une analyse phylogénétique invalide cependant cette hypothèse et suggère que l'ancêtre commun de TcTS et TrSA serait une trans-sialidase trypanosomale ²⁵⁸. Cette étude propose également qu'un transfert vertical du gène ancestral correspondant ait pu avoir lieu entre deux formes infectieuses du parasite. Il est donc tentant de supposer que TrSA ait pu évoluer à partir à partir de cette forme ancestrale hypothétique et perdre son activité trans-sialidase par l'acquisition d'une activité sialidase. Ces études laissent donc penser qu'une association de mutations clés apparues successivement au cours de l'évolution sur la séquence de TrSA aurait donc permis le passage d'une activité trans-sialidase à une activité sialidase. L'identification de ces mutations ainsi que de leur historique, c'est-à-dire de leur ordre d'apparition au cours de l'évolution, permettrait d'élucider le mécanisme évolutif permettant l'interconversion entre ces deux activités chez ces deux enzymes. Ce mécanisme pourrait alors potentiellement être généralisé à d'autres membres de la famille des GH33.

Afin de mettre en évidence les mutations responsables du passage de l'activité sialidase à l'activité de trans-sialidase et vice versa, il convient de déterminer la séquence de l'ancêtre commun aux sialidases et trans-sialidases de *T. rangeli* et *T. cruzi*, respectivement. La reconstruction du gène des ancêtres communs sur plusieurs nœuds

successifs de l'arbre phylogénétique est en effet une méthode de choix pour identifier la chronologie des mutations accumulées au cours de l'évolution dans la séquence d'une des deux enzymes ²⁵⁹. L'arbre phylogénétique construit précédemment a donc servi de support pour la prédiction des séquences protéiques ancestrales hypothétiques aux nœuds ancestraux les plus probables, c'est-à-dire ceux présentant les valeurs de bootstrap les plus élevés (>70 %).

Comme plusieurs possibilités de séquences ancestrales sont en général prédites par nœud, une étape ultérieure consistera à implémenter une méthodologie permettant de réduire le nombre de séquences à considérer en ne sélectionnant que les plus probables d'entre elles. Cette étape reviendra donc à établir un classement de l'ensemble des séquences ancestrales possibles sur la base de leur probabilité d'adopter un repliement stable et donc potentiellement fonctionnel. L'utilisation de logiciels tels que RosettaDesign ²⁰⁷ ou FoldX ²⁶⁰ devrait permettre de réaliser ces prédictions. Dans la continuité de ces travaux, les protéines ancestrales putatives correspondant aux séquences ancestrales potentielles sélectionnées à la suite de l'analyse précédente pourraient être exprimées au laboratoire et caractérisées biochimiquement.

8.10 Conclusions et perspectives

Les simulations effectuées sur les formes sauvages de TcTS et TrSA ont montré que la dynamique des résidus catalytiques E59 et E230 ainsi que celle des molécules d'eau internes aux domaines catalytiques de ces deux enzymes était fortement modifiée. Une perturbation du réseau de liaisons hydrogènes stabilisant les résidus situés en bordure des canaux a également été mis en évidence. Il est donc possible que ces effets puissent avoir des conséquences importantes sur la catalyse. Par analogie aux deux chapitres précédents, une analyse plus poussée de la trajectoire de l'eau interne aux domaines catalytiques pourra être réalisée ultérieurement afin de caractériser plus précisément la nature des modifications affectant sa dynamique.

L'alignement multiple et l'analyse bioinformatique de deux jeux de séquences protéiques de domaines catalytiques de sialidases et trans-sialidases de *T. rangeli* et *T. cruzi* sélectionnées dans diverses bases de données a permis d'isoler 37 positions divergentes entre ces deux sous-ensembles. Figurant parmi celles-ci, les résidus 115, 118, et 313 se sont révélés importants pour le maintien des performances catalytiques de TcTS. Un total de 35 autres positions restent néanmoins à tester afin d'évaluer leur contribution sur le contrôle de la balance hydrolyse/transglycosylation et sur la catalyse en général.

Une étude phylogénétique d'un jeu de séquences protéiques étendu aux domaines catalytiques de trans-sialidases d'autres espèces de trypanosomes a également permis de prédire un ensemble de séquences possibles de l'ancêtre commun aux sialidases et trans-sialidases de *T. rangeli* et *T. cruzi*. La sélection des séquences les plus probables suivie de la synthèse et de la caractérisation biochimique ultérieure des protéines correspondantes permettra de fournir de nouvelles informations sur l'historique des mutations ayant conduit à la divergence fonctionnelle entre ces deux enzymes.

Conclusion générale
et
Perspectives

9 Conclusion générale et perspectives

Ce travail de thèse a permis de réaliser une avancée dans la compréhension des bases moléculaires régissant la balance entre les activités d'hydrolyse et de transglycosylation chez quatre glycoside hydrolases issues respectivement des familles 1 (Tt β gly), 16 (AgaD), et 33 (la trans-sialidase TcTS et la sialidase TrSA). A travers une combinaison d'approches expérimentales et théoriques, il a été montré que l'eau interne à certaines régions protéiques éloignées du site catalytique pouvait jouer un rôle fonctionnel dans la catalyse des GH.

Des simulations de dynamique moléculaire de 500 ns effectuées sur ces quatre enzymes à différentes étapes de leur cycle catalytique ont mis en évidence l'existence de canaux d'eau connectant le milieu externe au site actif et ainsi potentiellement impliqués dans l'apport de molécules d'eau au voisinage du résidu acide-base catalytique. Chez Tt β gly et AgaD, deux canaux d'eau internes principaux ont été identifiés le long de deux chaînes d'eau cristallographiques débouchant dans le site actif et organisées suivant une géométrie analogue. L'analyse des simulations de TcTS et TrSA complexées au sialyl-lactose a révélé la présence de quatre canaux d'eau internes à leurs domaines catalytiques respectifs et débouchant sur le sous-site -1. Chez TcTS, la trajectoire fragmentée de trois de ces canaux d'eau suggère une circulation plus contrainte et inégale des molécules d'eau au sein de son domaine catalytique en comparaison à TrSA. Le déficit important de molécules d'eau au sein de ce domaine semble appuyer les hypothèses formulées par Demir et Roitberg²⁴⁶ et suggère ainsi qu'une accessibilité restreinte de l'eau du milieu externe au domaine catalytique de TcTS pourrait en partie expliquer la prédominance de l'activité de transglycosylation au sein de cette enzyme par opposition à TrSA qui ne catalyse que l'activité d'hydrolyse.

Les expériences de mutagenèse menées sur Tt β gly et AgaD ont révélé que trois résidus, respectivement W120, H178, C167 chez Tt β gly et D341, S351, Q342 chez AgaD, identifiés en bordure du canal d'eau principal étaient impliqués dans le contrôle de la balance entre les activités d'hydrolyse et de transglycosylation chez ces deux enzymes. Bordant l'extrémité du canal à l'interface avec le site actif, les résidus W120 et H178 chez Tt β gly et Q342 chez AgaD pourraient notamment contrôler l'accès des molécules

d'eau du canal vers l'acide-base catalytique et ainsi influencer la réaction d'hydrolyse. La mutation de ces résidus a également permis de générer des mutants transglycosidases très efficaces. La mutation W120I chez Tt β gly a notamment provoqué une augmentation du rendement de transglycosylation de 60 % par rapport à l'enzyme sauvage. De manière plus spectaculaire, l'introduction de la triple mutation D341L/S351F/Q342H au sein de la séquence d'AgaD a conduit à une inactivation presque totale de l'activité d'hydrolyse au profit de la transglycosylation. Ce triple-mutant s'est avéré être à ce jour la première endo-transagarase capable de catalyser la synthèse de polymères d'agarose à partir de substrats naturels. En plus de posséder une activité transglycosidase remarquable, les mutants Tt β gly-W120I et AgaD-D341L/S351F/Q342H sont également intéressants puisqu'ils préservent une efficacité catalytique peu réduite par rapport à l'enzyme sauvage. Le mutant Tt β gly-W120I garde une activité d'hydrolyse secondaire non-négligeable. Par contre, la stabilité des produits de transglycosylation générés par AgaD-D341L/S351F/Q342H et les performances catalytiques élevées de ce mutant rendent son utilisation particulièrement attractive pour des applications de synthèses enzymatiques *in-vitro*.

En vue d'expliquer les résultats expérimentaux, plusieurs analyses des simulations de dynamique moléculaire des formes sauvages et mutées de Tt β gly et d'AgaD ont été réalisées. Celles-ci ont montré que l'augmentation de l'activité de transglycosylation était systématiquement liée à une accélération de la dynamique des molécules d'eau internes au sein du canal d'eau principal. Cet effet se manifeste dans la plupart des cas par un relargage important des molécules d'eau du canal principal dans le sous-site -1. L'ensemble de ces résultats suggère ainsi que l'ingénierie de la dynamique de l'eau interne aux GH pourrait s'imposer comme une nouvelle approche rationnelle pour l'introduction d'une activité transglycosidase au sein de ces enzymes. Malgré les corrélations observées entre les résultats théoriques et expérimentaux, il n'est cependant pas aisé d'établir une relation claire entre l'augmentation de l'activité de transglycosylation et la dynamique de l'eau interne.

Si une modification importante de la dynamique de l'eau interne à l'ensemble des canaux étudiés a pu être observée chez les mutants transglycosidases de Tt β gly et d'AgaD, ainsi que chez TcTS, une altération de la dynamique conformationnelle des

chaînes latérales des résidus du sous-site -1 impliqués dans la stabilisation du substrat ou dans l'acte catalytique a également pu être mise en évidence. Compte tenu de l'importance de la géométrie de ces chaînes latérales dans la stabilisation de l'état de transition, l'ensemble de ces observations suggère donc que la balance entre les activités d'hydrolyse et de transglycosylation n'est probablement pas exclusivement contrôlée par la dynamique de l'eau interne aux canaux étudiés.

Lors de la catalyse de la réaction d'hydrolyse, le positionnement correct de la molécule d'eau catalytique au voisinage de l'acide-base catalytique pourrait nécessiter la formation d'un réseau de liaisons hydrogène servant à maintenir cette dernière dans une géométrie optimale pour initier son attaque nucléophile sur le carbone anomérique du substrat. En modifiant par mutagenèse la dynamique des molécules d'eau participant à la formation de ce réseau de liaison hydrogène hypothétique, on pourrait supposer dans un premier temps qu'une perturbation très locale de ce réseau puisse avoir lieu au voisinage des mutations mentionnées précédemment. Par effet de connectivité, cette perturbation locale du réseau de liaisons hydrogène pourrait au cours de la dynamique de la protéine se propager par la suite sur l'ensemble du réseau jusqu'au voisinage du point d'attaque nucléophile. Cet effet déstabiliserait *in-fine* la géométrie des chaînes latérales des résidus de la machinerie catalytique impliqués dans ce même réseau et compromettrait à fortiori le bon positionnement de la molécule d'eau catalytique et la formation de l'état de transition correspondant. Bien qu'elle nécessite d'être prouvée, cette hypothèse permettrait ainsi de faire le lien entre la modification de la dynamique de l'eau des canaux et de celle des chaînes latérales des résidus fonctionnels du site catalytique.

Bien qu'ayant fourni des résultats prometteurs, les analyses réalisées dans le cadre de ce projet sont incomplètes et nécessitent d'être poursuivies afin de confirmer l'ensemble des hypothèses soulevées à l'issue de ces travaux. L'étude de la dynamique de l'eau interne au site catalytique au moment de la formation de l'intermédiaire covalent (ou glycosyl-enzyme) reste essentielle car il s'agit de l'étape du cycle catalytique conduisant à la formation des produits d'hydrolyse et/ou de transglycosylation. La simulation des intermédiaires covalents d'AgaD ainsi que de TcTS et de TrSA est donc nécessaire afin de vérifier si les effets caractérisés dans les

simulations réalisées précédemment sont également bien observés à l'étape clé de la catalyse. La détermination expérimentale de la structure 3D de la forme glycosyl-enzyme d'AgaD et la paramétrisation *ab-initio* des résidus anhydro-galactose de la partie glycosyle reste cependant les principaux obstacles à la réalisation d'analyses théoriques de l'étape catalytique de cette enzyme. La connaissance de la structure 3D d'AgaD ainsi que des paramètres mentionnés précédemment serait en effet essentielle afin de réaliser une simulation des intermédiaires covalents de la forme sauvage et de ses mutants afin de confirmer les effets observés dans ces travaux et d'en valider les hypothèses sous-jacentes. Ces nouvelles simulations nous permettraient notamment de quantifier précisément l'impact des mutations sur l'accessibilité des molécules d'eau catalytiques au point d'attaque nucléophile. La réalisation de simulations QM/MM couplées à des calculs d'énergie libre sur les formes sauvages de Tt β gly et d'AgaD ainsi que sur leurs meilleurs mutants respectifs pourrait également être envisagée dans le but d'obtenir des informations détaillées sur la stabilisation des états de transition propres aux réactions d'hydrolyse et de transglycosylation.

Une description plus fine de la trajectoire des molécules d'eau internes pourrait être obtenue par un échantillonnage plus fréquent des clichés de trajectoire sur un intervalle de l'ordre de la dizaine de picosecondes (au lieu d'un intervalle de 100 ps choisi dans ces travaux). La qualité des prédictions obtenues par simulations la dynamique moléculaire pourrait également être améliorée par le choix d'un modèle d'eau plus réaliste (TIP4P¹⁷⁸, AMOEBA²⁶³...) que celui utilisé dans cette étude (TIP3P) ou/et en utilisant des modèles bien plus détaillés.

Par analogie avec la sialidase et la trans-sialidase des parasites *T. rangeli* et *T. cruzi*, les approches développées au cours de cette thèse pourrait également être appliquées à l'étude de nouvelles paires d'homologues glycoside-hydrolase/transglycosidase^{159,230}. Il a précédemment été mentionné que la trajectoire du canal d'eau principal mis en évidence chez Tt β gly et AgaD a été identifiée le long d'une chaîne de trois molécules d'eau cristallographiques adoptant une géométrie commune en sortie du canal à l'interface avec le site catalytique. Ces des, mises en évidence au sein de GH présentant des repliements structuraux différents, laissent donc supposer que ces découvertes pourraient être communes à d'autres familles de GH.

Annexe

Modification du champ de force AMBER99SB-ildn pour la simulation d'intermédiaires covalents glucosyl-enzymes

I. Modification du fichier ffbonded.itp

La topologie de la partie glucosyle a été dérivée à partir de celle de la glucosyl-thréonine paramétrée dans le champ de force Glycam06. Afin de pouvoir être exploitée, la topologie du glucose généré par Glycam06 a été préalablement convertie au format utilisé par le logiciel GROMACS *via* le script `acpype.py`²¹⁵ avant d'être intégrée au fichier `ffbonded.itp`. Le fichier `ffbonded.itp` précise l'ensemble des interactions interatomiques liées pouvant être décrites par le champ de force. Les ajouts suivants ont donc été introduits au sein des différentes sections spécifiant chaque type d'interaction:

[bondtypes]

- Paramètres d'élongation des liaisons de la partie Glucosyle :

```
Cg H92  1  1.0900e-01  2.8451e+05
Cg Os   1  1.4600e-01  2.3849e+05
Cg Cg   1  1.5200e-01  2.5941e+05
Cg H91  1  1.0900e-01  2.8451e+05
Cg O9h  1  1.4300e-01  2.6778e+05
O9h H9o 1  9.6000e-02  4.6275e+05
```

- Paramètres d'élongation de la liaison Glucosyl-glutamate :

```
Cg O2   1  1.4600e-01  2.3849e+05
```

Les 2 premières colonnes spécifient les types d'atomes en interaction, la colonne 3 le type de liaison, et les deux dernières colonnes la valeur d'élongation de la liaison à l'équilibre et sa constante de force associée.

[angletypes]

- Paramètres de déformation des angles de la partie Glucosyle :

```
Cg Os Cg   1  1.1160e+02  4.1840e+02
Cg Cg Cg   1  1.1350e+02  3.7656e+02
Cg Cg H91  1  1.1100e+02  3.7656e+02
Cg Cg H92  1  1.1100e+02  3.7656e+02
Cg Cg O9h  1  1.0750e+02  5.8576e+02
H92 Cg Os  1  1.1000e+02  5.0208e+02
O9h Cg Os  1  1.1200e+02  8.3680e+02
H92 Cg Cg  1  1.1100e+02  3.7656e+02
Os Cg Cg   1  1.0850e+02  5.8576e+02
Os Cg H91  1  1.1000e+02  5.0208e+02
```

H91 Cg Cg	1	1.1100e+02	3.7656e+02
Cg O9h H9o	1	1.0950e+02	4.6024e+02
H91 Cg H91	1	1.0950e+02	3.7656e+02
H91 Cg O9h	1	1.1000e+02	5.0208e+02
H92 Cg O9h	1	1.1000e+02	5.0208e+02
O9h Cg Cg	1	1.0750e+02	5.8576e+02

- Paramètres de déformation des angles de la liaison Glucosyl-glutamate :

C O2 Cg	1	1.1160e+02	4.1840e+02
O2 Cg H92	1	1.1000e+02	5.0208e+02
O2 Cg Os	1	1.1200e+02	8.3680e+02
O2 Cg Cg	1	1.0850e+02	5.8576e+02

Les 3 premières colonnes spécifient les types d'atomes en interaction, la colonne 4 le type de l'angle, et les deux dernières colonnes la valeur de l'angle de valence à l'équilibre et sa constante de force associée en (kJ/mol/nm²).

[dihedraltypes]

- Paramètres de torsion des dièdres de la liaison Glucosyl-glutamate

CT C O2 Cg	3	0.66944	2.00832	0.00000	-2.67776	0.00000	0.00000
C O2 Cg H92	3	0.41840	1.25520	5.02080	-1.67360	0.00000	0.00000
C O2 Cg Os	3	2.80328	3.38904	10.62736	-6.19232	0.00000	0.00000
C O2 Cg Cg	3	0.66944	2.00832	0.00000	-2.67776	0.00000	0.00000
O2 Cg Os Cg	3	8.53536	7.53120	11.54784	-16.06656	0.00000	0.00000
O2 Cg Cg Cg	3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
O2 Cg Cg H92	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
O2 Cg Cg H91	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
O2 Cg Cg O9h	3	0.00000	0.00000	2.09200	0.00000	0.00000	0.00000

- Paramètres de torsion des dièdres de la partie Glucosyle :

O9h Cg Cg Cg	3	0.41840	1.25520	0.00000	-1.67360	0.00000	0.00000
O9h Cg Cg H92	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
O9h Cg Cg O9h	3	2.30120	6.90360	7.94960	-9.20480	0.00000	0.00000
Cg Os Cg H91	3	1.12968	3.38904	0.00000	-4.51872	0.00000	0.00000
Cg Os Cg Cg	3	0.66944	2.00832	0.00000	-2.67776	0.00000	0.00000
Cg Cg O9h H9o	3	0.75312	2.25936	0.00000	-3.01248	0.00000	0.00000
H92 Cg Os Cg	3	0.41840	1.25520	5.02080	-1.67360	0.00000	0.00000
H92 Cg Cg Cg	3	0.62760	1.88280	0.00000	-2.51040	0.00000	0.00000
H92 Cg Cg H92	3	0.71128	2.13384	0.00000	-2.84512	0.00000	0.00000
H92 Cg Cg O9h	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
Os Cg Cg Cg	3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
Os Cg Cg H92	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
Os Cg Cg O9h	3	0.00000	0.00000	2.09200	0.00000	0.00000	0.00000
Os Cg Cg H91	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
Cg Cg Cg H92	3	0.62760	1.88280	0.00000	-2.51040	0.00000	0.00000
Cg Cg Cg O9h	3	0.41840	1.25520	0.00000	-1.67360	0.00000	0.00000
Cg Cg Cg Cg	3	1.88280	-1.88280	0.00000	0.00000	0.00000	0.00000
H91 Cg Cg H91	3	0.71128	2.13384	0.00000	-2.84512	0.00000	0.00000
H91 Cg Cg O9h	3	0.20920	0.62760	0.00000	-0.83680	0.00000	0.00000
H91 Cg Cg Cg	3	0.62760	1.88280	0.00000	-2.51040	0.00000	0.00000
Cg Cg Cg H91	3	0.62760	1.88280	0.00000	-2.51040	0.00000	0.00000
H91 Cg O9h H9o	3	0.75312	2.25936	0.00000	-3.01248	0.00000	0.00000
H91 Cg Cg H92	3	0.71128	2.13384	0.00000	-2.84512	0.00000	0.00000
H92 Cg O9h H9o	3	0.75312	2.25936	0.00000	-3.01248	0.00000	0.00000
Cg Cg Os Cg	3	0.66944	2.00832	0.00000	-2.67776	0.00000	0.00000

Les 4 premières colonnes spécifient les types d'atomes en interaction, la colonne 5 le type de dièdre (Ryckaert-Bellemans), et les autres colonnes les coefficients de Ryckaert-Bellemans.

Le paramétrage du glucose *via* Glycam06 génère des angles dièdres de type Ryckaert-Bellemans (angle dièdre de type 3) alors que le champ de force utilisé pour la simulation est paramétré pour reconnaître par défaut les angles dièdres propres de type 9 (à l'exception des angles dièdres impropres, de type 4). Le programme attribue donc par défaut le type 9 à tout dièdre rajouté au champ de force. Le type 9 attribué aux angles dièdres nouvellement introduits doit alors être remplacé par 3 à l'exception des atomes OE1 CD OE2 C1 (fonction carboxylate) qui forment un angle dièdre impropre qui se voient attribuer le type 4. Les types atomiques dérivés par Glycam06 étant redondants avec ceux d'autres atomes dans la version originale du champ de force, de nouveaux types atomiques doivent être introduits afin de décrire les atomes de la partie glucosyle. Les types atomiques existants C, Oh, Ho, H1, et H2, sont alors respectivement remplacés par : Cg, O9h, H9o, H91, H92.

II. Modification du fichier ffnonbonded.itp

Le fichier ffnonbonded.itp précise l'ensemble des interactions interatomiques non-liées pouvant être décrites par le champ de force. Les ajouts suivants ont donc été introduits au sein du fichier:

```
[ atomtypes ]
; name  at.num  mass  charge  ptype  sigma  epsilon
H9o    1    1.008  0.0000  A    3.56359e-02  1.25520e-01
O9h    8    16.00  0.0000  A    3.06647e-01  8.80314e-01
Cg     6    12.01  0.0000  A    3.39967e-01  4.57730e-01
H91    1    1.008  0.0000  A    2.29317e-01  6.56888e-02
H92    1    1.008  0.0000  A    2.29317e-01  6.56888e-02
Os     8    16.00  0.0000  A    3.00001e-01  7.11280e-01
```

Les colonnes 1, 2, 3, 4, et 5 spécifient respectivement les types d'atomes de la partie glucosyle, leur identifiants numérique, leur masse, leur charge et leur type de particule. Les deux dernières colonnes spécifient les paramètres de Van der Waals de l'atome concerné.

III. Modification du fichier aminoacids.rtp

Le fichier aminoacids.rtp liste l'ensemble des atomes de chaque résidu décrit dans le champ de force, la charge partielle associée à chaque atome, ainsi que la nature de leurs interactions au sein du résidu. Le nouveau résidu glucosyl-glutamate (ou GLG) a donc été décrit de la manière suivante :

[GLG]

[atoms]

N	N	-0.531117	1
H	H	0.333334	2
CA	CT	0.354358	3
HA	H1	0.009158	4
CB	CT	-0.180667	5
HB1	HC	0.035399	6
HB2	HC	0.106395	7
CG	CT	-0.223820	8
HG1	HC	0.081917	9
HG2	HC	0.094616	10
CD	C	0.812173	11
OE1	O2	-0.563207	12
OE2	O2	-0.428627	13
C	C	0.292589	14
O	O	-0.400910	15
C1	Cg	0.110019	16
H1	H92	0.134630	17
O5	Os	-0.317211	18
C5	Cg	0.074512	19
H5	H91	0.059466	20
C6	Cg	0.198507	21
H62	H91	0.005454	22
H61	H91	0.066987	23
O6	O9h	-0.586567	24
H60	H9o	0.385227	25
C4	Cg	0.221358	26
H4	H91	0.039914	27
O4	O9h	-0.615440	28
H40	H9o	0.419247	29
C3	Cg	0.157640	30
H3	H91	0.055769	31
O3	O9h	-0.640479	32
H30	H9o	0.446491	33
C2	Cg	0.132506	34
H2	H91	0.058206	35
O2	O9h	-0.620405	36
H20	H9o	0.422581	37

[bonds]

N H
N CA
CA HA
CA CB
CA C
CB HB1
CB HB2
CB CG
CG HG1
CG HG2
CG CD
CD OE1
CD OE2
C O
-C N
C1 OE2
C1 H1
C1 O5
C1 C2
O5 C5
C5 H5
C5 C6
C5 C4
C6 H61
C6 H62
C6 O6
O6 H60
C4 H4
C4 O4
C4 C3
O4 H40
C3 H3
C3 O3
C3 C2
O3 H30
C2 H2
C2 O2
O2 H20

[impropers]

-C CA N H
CA +N C O
CG OE1 CD OE2

Les charges partielles du résidu (4^{ème} colonne, section [atoms]) ont été déterminées de manière *ab-initio* à partir du calcul de son potentiel électrostatique (méthode RESP ²¹⁷). Préalablement au calcul des charges, la géométrie du résidu a été optimisée *in-vacuo* en utilisant la méthode de DFT et le niveau de théorie B3LYP/6-31G*. Le

logiciel Gaussian (version 02) a été utilisé pour ces calculs, qui ont été effectués par Philippe Arnaud.

IV. Modification du fichier atomtypes.atp

Le fichier atomtypes.atp précise la liste des types d'atomes décrits par le champ de force. A chaque type atomique est associée sa masse correspondante. La section suivante a donc été rajoutée au fichier :

Cg	12.01000
H92	1.00800
Os	16.00000
H91	1.00800
O9h	16.00000
H9o	1.00800

V. Modifications complémentaires

Les fichiers residuetypes.dat et xlateat.dat ont également été modifiés. L'identifiant du nouveau résidu, nommé ici GLG (pour Glucosyl GLutamate) doit être explicitement renseigné dans le fichier residuetypes.dat. La ligne « GLG Protein » y a donc été rajoutée. L'identifiant O2 étant spécifiquement réservé au second atome d'oxygène du cycle osidique du glucosyl glutamate, la ligne : « protein O2 OXT » a été supprimée du fichier xlateat.dat.

Bibliographie

1. Springer, S. A. & Gagneux, P. Glycomics: revealing the dynamic ecology and evolution of sugar molecules. *J. Proteomics* **135**, 90–100 (2016).
2. Boyle, J. Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochem. Mol. Biol. Educ.* **33**, 74–75 (2005).
3. Tavares, E. Q. P. & Buckeridge, M. S. Do plant cell walls have a code? *Plant Sci.* **241**, 286–294 (2015).
4. Mrázek, H. *et al.* Carbohydrate synthesis and biosynthesis technologies for cracking of the glycan code: Recent advances. *Biotechnol. Adv.* **31**, 17–37 (2013).
5. Oliveira, D. L., Wilbey, R. A., Grandison, A. S. & Roseiro, L. B. Milk oligosaccharides: A review. *Int. J. Dairy Technol.* **68**, 305–321 (2015).
6. Quirós, L. M., Carbajo, R. J., Braña, A. F. & Salas, J. A. Glycosylation of Macrolide Antibiotics PURIFICATION AND KINETIC STUDIES OF A MACROLIDE GLYCOSYLTRANSFERASE FROM STREPTOMYCES ANTIBIOTICUS. *J. Biol. Chem.* **275**, 11713–11720 (2000).
7. Xiao, J., Chen, T. & Cao, H. Flavonoid glycosylation and biological benefits. *Biotechnol. Adv.* (2014). doi:10.1016/j.biotechadv.2014.05.004
8. Dade, J. M. E. *et al.* Pyrrolidine alkaloids and their glycosylated derivatives from the root bark of *Dichrostachys cinerea* (L) Wight & Arn. (Fabaceae). *Phytochem. Lett.* **16**, 268–276 (2016).
9. Treutter, D. Significance of flavonoids in plant resistance and enhancement of their biosynthesis. *Plant Biol. Stuttg. Ger.* **7**, 581–591 (2005).
10. Mithöfer, A. & Boland, W. Plant defense against herbivores: chemical aspects. *Annu. Rev. Plant Biol.* **63**, 431–450 (2012).
11. Hart, G. W. & Copeland, R. J. Glycomics Hits the Big Time. *Cell* **143**, 672–676 (2010).
12. Lafite, P. & Daniellou, R. Rare and unusual glycosylation of peptides and proteins. *Nat. Prod. Rep.* **29**, 729 (2012).
13. Amon, R., Reuven, E. M., Leviatan Ben-Arye, S. & Padler-Karavani, V. Glycans in immune recognition and response. *Carbohydr. Res.* **389**, 115–122 (2014).
14. Hennet, T. & Cabalzar, J. Congenital disorders of glycosylation: a concise chart of glycolyx dysfunction. *Trends Biochem. Sci.* **40**, 377–384 (2015).
15. Teclé, E. & Gagneux, P. Sugar-coated sperm: Unraveling the functions of the mammalian sperm glycolyx: FUNCTIONS OF THE MAMMALIAN SPERM GLYCOLYX. *Mol. Reprod. Dev.* **82**, 635–650 (2015).
16. Tan, F. Y. Y., Tang, C. M. & Exley, R. M. Sugar coating: bacterial protein glycosylation and host–microbe interactions. *Trends Biochem. Sci.* **40**, 342–350 (2015).
17. Cooper, G. M. Cell Walls and the Extracellular Matrix. (2000).
18. Calabro, V. & Ronin, C. Glycoengineering of protein-based. *Carbohydr. Chem.* **38**, 94–123 (2012).
19. Shylaja, M. & Seshadri, H. S. Glycoproteins: an overview. *Biochem. Educ.* **17**, 170–178 (1989).
20. Donadio-Andréi, S., El Maï, N., Nicollo, M. & Ronin, C. in *Carbohydrate Chemistry* (eds. Pilar Rauter, A., Lindhorst, T. & Queneau, Y.) **40**, 461–490 (Royal Society of Chemistry, 2014).

21. Chia, J., Goh, G. & Bard, F. Short O-GalNAc glycans: regulation and role in tumor development and clinical perspectives. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1860**, 1623–1639 (2016).
22. Wells, L. The O-Mannosylation Pathway: Glycosyltransferases and Proteins Implicated in Congenital Muscular Dystrophy. *J. Biol. Chem.* **288**, 6930–6935 (2013).
23. Fukuda, T., Roberts, A., Plotz, P. H. & Raben, N. Acid alpha-glucosidase deficiency (Pompe disease). *Curr. Neurol. Neurosci. Rep.* **7**, 71–77 (2007).
24. Horowitz, M., Elstein, D., Zimran, A. & Goker-Alpan, O. New Directions in Gaucher Disease. *Hum. Mutat.* (2016). doi:10.1002/humu.23056
25. Zulueta, M. M. L., Janreddy, D. & Hung, S.-C. One-Pot Methods for the Protection and Assembly of Sugars. *Isr. J. Chem.* **55**, 347–359 (2015).
26. Koenigs, W. & Knorr, E. Ueber einige Derivate des Traubenzuckers und der Galactose. *Berichte Dtsch. Chem. Ges.* **34**, 957–981 (1901).
27. Seeberger, P. H. Automated oligosaccharide synthesis. *Chem. Soc. Rev.* **37**, 19–28 (2007).
28. Davies, G. J. & Williams, S. J. Carbohydrate-active enzymes: sequences, shapes, contortions and cells. *Biochem. Soc. Trans.* **44**, 79–87 (2016).
29. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
30. Gloster, T. M. Advances in understanding glycosyltransferases from a structural perspective. *Curr. Opin. Struct. Biol.* **28**, 131–141 (2014).
31. Ardèvol, A. & Rovira, C. Reaction Mechanisms in Carbohydrate-Active Enzymes: Glycoside Hydrolases and Glycosyltransferases. Insights from *ab Initio* Quantum Mechanics/Molecular Mechanics Dynamic Simulations. *J. Am. Chem. Soc.* **137**, 7528–7547 (2015).
32. Martin, A. *et al.* Catching elusive glycosyl cations in a condensed phase with HF/SbF₅ superacid. *Nat. Chem.* (2015). doi:10.1038/nchem.2399
33. Schuman, B., Evans, S. V. & Fyles, T. M. Geometric Attributes of Retaining Glycosyltransferase Enzymes Favor an Orthogonal Mechanism. *PLoS ONE* **8**, e71077 (2013).
34. Rivas, F., Parra, A., Martinez, A. & Garcia-Granados, A. Enzymatic glycosylation of terpenoids. *Phytochem. Rev.* **12**, 327–339 (2013).
35. De Bruyn, F., Maertens, J., Beauprez, J., Soetaert, W. & De Mey, M. Biotechnological advances in UDP-sugar based glycosylation of small molecules. *Biotechnol. Adv.* **33**, 288–302 (2015).
36. Strasser, R., Altmann, F. & Steinkellner, H. Controlled glycosylation of plant-produced recombinant proteins. *Curr. Opin. Biotechnol.* **30**, 95–100 (2014).
37. Loos, A. & Steinkellner, H. Plant glyco-biotechnology on the way to synthetic biology. *Front. Plant Sci.* **5**, (2014).
38. Mathieu-Rivet, E. *et al.* Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. *Front. Plant Sci.* **5**, (2014).
39. Hofer, B. Recent developments in the enzymatic O-glycosylation of flavonoids. *Appl. Microbiol. Biotechnol.* **100**, 4269–4281 (2016).

40. Merritt, J. H., Ollis, A. A., Fisher, A. C. & DeLisa, M. P. Glycans-by-design: Engineering bacteria for the biosynthesis of complex glycans and glycoconjugates. *Biotechnol. Bioeng.* **110**, 1550–1564 (2013).
41. Cuccui, J. & Wren, B. Hijacking bacterial glycosylation for the production of glycoconjugates, from vaccines to humanised glycoproteins: Novel glycoconjugate assembly. *J. Pharm. Pharmacol.* **67**, 338–350 (2015).
42. De Pourcq, K., De Schutter, K. & Callewaert, N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Appl. Microbiol. Biotechnol.* **87**, 1617–1631 (2010).
43. Krasnova, L. & Wong, C.-H. Exploring human glycosylation for better therapies. *Mol. Aspects Med.* (2016). doi:10.1016/j.mam.2016.05.003
44. Tagliabracci, V. S. & Roach, P. J. Insights into the mechanism of polysaccharide dephosphorylation by a glucan phosphatase. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15312–15313 (2010).
45. Sutherland, I. W. Polysaccharide lyases. *FEMS Microbiol. Rev.* **16**, 323–347 (1995).
46. Hemsworth, G. R., Henrissat, B., Davies, G. J. & Walton, P. H. Discovery and characterization of a new family of lytic polysaccharide monoxygenases. *Nat. Chem. Biol.* **10**, 122–126 (2014).
47. Vocadlo, D. J. & Davies, G. J. Mechanistic insights into glycosidase chemistry. *Curr. Opin. Chem. Biol.* **12**, 539–555 (2008).
48. Zechel, D. L. & Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33**, 11–18 (2000).
49. Vuong, T. V. & Wilson, D. B. Glycoside hydrolases: Catalytic base/nucleophile diversity. *Biotechnol. Bioeng.* **107**, 195–205 (2010).
50. Armstrong, Z. & Withers, S. G. Synthesis of Glycans and Glycopolymers Through Engineered Enzymes: Synthesis of Glycans and Glycopolymers. *Biopolymers* **99**, 666–674 (2013).
51. Botti, M. G., Taylor, M. G. & Botting, N. P. Studies on the Mechanism of Myrosinase INVESTIGATION OF THE EFFECT OF GLYCOSYL ACCEPTORS ON ENZYME ACTIVITY. *J. Biol. Chem.* **270**, 20530–20535 (1995).
52. Bojarová, P. & Křen, V. Glycosidases: a key to tailored carbohydrates. *Trends Biotechnol.* **27**, 199–209 (2009).
53. Rajan, S. S. *et al.* Novel Catalytic Mechanism of Glycoside Hydrolysis Based on the Structure of an NAD⁺/Mn²⁺-Dependent Phospho- α -Glucosidase from *Bacillus subtilis*. *Structure* **12**, 1619–1629 (2004).
54. Liu, Q. P. *et al.* Bacterial glycosidases for the production of universal red blood cells. *Nat. Biotechnol.* **25**, 454–464 (2007).
55. Wang, L.-X. & Huang, W. Enzymatic transglycosylation for glycoconjugate synthesis. *Curr. Opin. Chem. Biol.* **13**, 592–600 (2009).
56. Bissaro, B., Monsan, P., Fauré, R. & O'Donohue, M. J. Glycosynthesis in a waterworld: new insight into the molecular basis of transglycosylation in retaining glycoside hydrolases. *Biochem. J.* **467**, 17–35 (2015).
57. Skov, L. K. Amylosucrase, a Glucan-synthesizing Enzyme from the α -Amylase Family. *J. Biol. Chem.* **276**, 25273–25278 (2001).

58. Peterbauer, T., Mach, L., Mucha, J. & Richter, A. Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. *Planta* **215**, 839–846 (2002).
59. Scheurwater, E., Reid, C. W. & Clarke, A. J. Lytic transglycosylases: Bacterial space-making autolysins. *Int. J. Biochem. Cell Biol.* **40**, 586–591 (2008).
60. Larsbrink, J., Izumi, A., Hemsworth, G. R., Davies, G. J. & Brumer, H. Structural Enzymology of Cellvibrio japonicus Agd31B Protein Reveals -Transglucosylase Activity in Glycoside Hydrolase Family 31. *J. Biol. Chem.* **287**, 43288–43299 (2012).
61. Hurtado-Guerrero, R. *et al.* Molecular Mechanisms of Yeast Cell Wall Glucan Remodeling. *J. Biol. Chem.* **284**, 8461–8469 (2009).
62. Blanco, N. *et al.* Structural and functional analysis of yeast Crh1 and Crh2 transglycosylases. *FEBS J.* **282**, 715–731 (2015).
63. Qin, Z. *et al.* The first crystal structure of a glycoside hydrolase family 17 β -1,3-glucanosyltransferase displays a unique catalytic cleft. *Acta Crystallogr. D Biol. Crystallogr.* **71**, 1714–1724 (2015).
64. Lammens, W. *et al.* Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. *J. Exp. Bot.* **60**, 727–740 (2009).
65. Moulis, C., André, I. & Remaud-Simeon, M. GH13 amylosucrases and GH70 branching sucrases, atypical enzymes in their respective families. *Cell. Mol. Life Sci.* **73**, 2661–2679 (2016).
66. Moellering, E. R., Muthan, B. & Benning, C. Freezing Tolerance in Plants Requires Lipid Remodeling at the Outer Chloroplast Membrane. *Science* **330**, 226–228 (2010).
67. Matsuba, Y. *et al.* A Novel Glucosylation Reaction on Anthocyanins Catalyzed by Acyl-Glucose-Dependent Glucosyltransferase in the Petals of Carnation and Delphinium. *Plant Cell* **22**, 3374–3389 (2010).
68. Luang, S. *et al.* Rice Os9BGlu31 Is a Transglucosidase with the Capacity to Equilibrate Phenylpropanoid, Flavonoid, and Phytohormone Glycoconjugates. *J. Biol. Chem.* **288**, 10111–10123 (2013).
69. Cabib, E. *et al.* Assembly of the Yeast Cell Wall: Crh1p AND Crh2p ACT AS TRANSGLYCOSYLASES IN VIVO AND IN VITRO. *J. Biol. Chem.* **283**, 29859–29872 (2008).
70. Franková, L. & Fry, S. C. Trans- α -xylosidase and trans- β -galactosidase activities, widespread in plants, modify and stabilize xyloglucan structures: Trans- α -xylosidase and trans- β -galactosidase. *Plant J.* **71**, 45–60 (2012).
71. Welzel, P. Syntheses around the Transglycosylation Step in Peptidoglycan Biosynthesis. *Chem. Rev.* **105**, 4610–4660 (2005).
72. Garcia, G. A. & Kittendorf, J. D. Transglycosylation: A mechanism for RNA modification (and editing?). *Bioorganic Chem.* **33**, 229–251 (2005).
73. Monsan, P., Remaud-Siméon, M. & André, I. Transglucosidases as efficient tools for oligosaccharide and glucoconjugate synthesis. *Curr. Opin. Microbiol.* **13**, 293–300 (2010).
74. Overwin, H., Wray, V. & Hofer, B. Flavonoid glucosylation by non-Leloir glycosyltransferases: formation of multiple derivatives of 3,5,7,3',4'-pentahydroxyflavane stereoisomers. *Appl. Microbiol. Biotechnol.* **99**, 9565–9576 (2015).

75. Ko, J.-A. *et al.* Glucosyl Rubusosides by Dextranucrases Improve the Quality of Taste and Sweetness. *J. Microbiol. Biotechnol.* **26**, 493–497 (2016).
76. Barea-Alvarez, M., Benito, M. T., Olano, A., Jimeno, M. L. & Moreno, F. J. Synthesis and Characterization of Isomaltulose-Derived Oligosaccharides Produced by Transglucosylation Reaction of *Leuconostoc mesenteroides* Dextranucrase. *J. Agric. Food Chem.* **62**, 9137–9144 (2014).
77. Chlubnova, I. *et al.* Specific and non-specific enzymes for furanosyl-containing conjugates: biosynthesis, metabolism, and chemo-enzymatic synthesis. *Carbohydr. Res.* **356**, 44–61 (2012).
78. Stick, R. V. & Williams, S. J. *Carbohydrates: The Essential Molecules of Life*. (Elsevier, 2009).
79. Teze, D. *et al.* Alkoxyamino glycoside acceptors for the regioselective synthesis of oligosaccharides using glycosynthases and transglycosidases. *Bioorg. Med. Chem. Lett.* **23**, 448–451 (2013).
80. Sakaguchi, K., Katoh, T. & Yamamoto, K. Transglycosidase-like activity of *Mucor hiemalis* endoglycosidase mutants enabling the synthesis of glycoconjugates using a natural glycan donor: Endo-M N175H Is a Superior Transglycosidase. *Biotechnol. Appl. Biochem.* n/a-n/a (2015). doi:10.1002/bab.1433
81. Umekawa, M. *et al.* Mutants of *Mucor hiemalis* Endo- β -N-acetylglucosaminidase Show Enhanced Transglycosylation and Glycosynthase-like Activities. *J. Biol. Chem.* **283**, 4469–4479 (2008).
82. Honda, Y. & Kitaoka, M. The First Glycosynthase Derived from an Inverting Glycoside Hydrolase. *J. Biol. Chem.* **281**, 1426–1431 (2006).
83. Aragunde, H., Castilla, E., Biarnés, X., Faijes, M. & Planas, A. A transitional hydrolase to glycosynthase mutant by Glu to Asp substitution at the catalytic nucleophile in a retaining glycosidase. *Carbohydr. Res.* **389**, 85–92 (2014).
84. Honda, Y. *et al.* Alternative strategy for converting an inverting glycoside hydrolase into a glycosynthase. *Glycobiology* **18**, 325–330 (2008).
85. Hidaka, M. *et al.* Structural explanation for the acquisition of glycosynthase activity. *J. Biochem. (Tokyo)* **147**, 237–244 (2010).
86. Pengthaisong, S., Withers, S. G., Kuaprasert, B., Svasti, J. & Ketudat Cairns, J. R. The role of the oligosaccharide binding cleft of rice BGl1 in hydrolysis of cellooligosaccharides and in their synthesis by rice BGl1 glycosynthase. *Protein Sci.* **21**, 362–372 (2012).
87. Ohnuma, T. *et al.* A glycosynthase derived from an inverting GH19 chitinase from the moss *Bryum coronatum*. *Biochem. J.* **444**, 437–443 (2012).
88. Spadiut, O. *et al.* Building Custom Polysaccharides in Vitro with an Efficient, Broad-Specificity Xyloglucan Glycosynthase and a Fucosyltransferase. *J. Am. Chem. Soc.* **133**, 10892–10900 (2011).
89. Ben-David, A. *et al.* Glycosynthase Activity of *Geobacillus stearothermophilus* GH52 β -Xylosidase: Efficient Synthesis of Xylooligosaccharides from α -D-Xylopyranosyl Fluoride through a Conjugated Reaction. *ChemBioChem* **8**, 2145–2151 (2007).
90. Vaughan, M. D. *et al.* Glycosynthase-mediated synthesis of glycosphingolipids. *J. Am. Chem. Soc.* **128**, 6300–6301 (2006).
91. Smith, E. L. *et al.* Chemoenzymatic Fc Glycosylation via Engineered Aldehyde Tags. *Bioconjug. Chem.* **25**, 788–795 (2014).

92. Yang, M., Davies, G. J. & Davis, B. G. A Glycosynthase Catalyst for the Synthesis of Flavonoid Glycosides. *Angew. Chem. Int. Ed.* **46**, 3885–3888 (2007).
93. Kim, Y.-W., Lee, S. S., Warren, R. A. J. & Withers, S. G. Directed Evolution of a Glycosynthase from *Agrobacterium* sp. Increases Its Catalytic Activity Dramatically and Expands Its Substrate Repertoire. *J. Biol. Chem.* **279**, 42787–42793 (2004).
94. Hancock, S. M., Rich, J. R., Caines, M. E. C., Strynadka, N. C. J. & Withers, S. G. Designer enzymes for glycosphingolipid synthesis by directed evolution. *Nat. Chem. Biol.* **5**, 508–514 (2009).
95. Jahn, M., Marles, J., Warren, R. A. J. & Withers, S. G. Thioglycoligases: mutant glycosidases for thioglycoside synthesis. *Angew. Chem. Int. Ed Engl.* **42**, 352–354 (2003).
96. Jahn, M. *et al.* Thioglycosynthases: double mutant glycosidases that serve as scaffolds for thioglycoside synthesis. *Chem Commun* 274–275 (2004). doi:10.1039/B313155F
97. Li, C., Kim, J.-H. & Kim, Y.-W. α -Thioglycoligase-based synthesis of O-aryl α -glycosides as chromogenic substrates for α -glycosidases. *J. Mol. Catal. B Enzym.* **87**, 24–29 (2013).
98. Kim, Y.-W. *et al.* Thioglycoligase-Based Assembly of Thiodisaccharides: Screening as β -Galactosidase Inhibitors. *ChemBioChem* **8**, 1495–1499 (2007).
99. Müllegger, J., Chen, H. M., Warren, R. A. J. & Withers, S. G. Glycosylation of a Neoglycoprotein by Using Glycosynthase and Thioglycoligase Approaches: The Generation of a Thioglycoprotein. *Angew. Chem. Int. Ed.* **45**, 2585–2588 (2006).
100. Drone, J. *et al.* *Thermus thermophilus* Glycosynthases for the Efficient Synthesis of Galactosyl and Glucosyl β -(1 \rightarrow 3)-Glycosides. *Eur. J. Org. Chem.* **2005**, 1977–1983 (2005).
101. Sridhar, P. R., Prabhu, K. R. & Chandrasekaran, S. Synthesis of Thioglycosides by Tetrathiomolybdate-Mediated Michael Additions of Masked Thiolates. *Eur. J. Org. Chem.* **2004**, 4809–4815 (2004).
102. Feng, H.-Y. *et al.* Converting a β -Glycosidase into a β -Transglycosidase by Directed Evolution. *J. Biol. Chem.* **280**, 37088–37097 (2005).
103. Kone, F. M. T., Le Behec, M., Sine, J.-P., Dion, M. & Tellier, C. Digital screening methodology for the directed evolution of transglycosidases. *Protein Eng. Des. Sel.* **22**, 37–44 (2008).
104. Teze, D. *et al.* Semi-rational approach for converting a GH1 β -glycosidase into a β -transglycosidase. *Protein Eng. Des. Sel.* **27**, 13–19 (2014).
105. Teze, D., Daligault, F., Ferrieres, V., Sanejouand, Y.-H. & Tellier, C. Semi-rational approach for converting a GH36 β -glycosidase into an α -transglycosidase. *Glycobiology* **25**, 420–427 (2015).
106. Arab-Jaziri, F. *et al.* Engineering transglycosidase activity into a GH51 α -L-arabinofuranosidase. *New Biotechnol.* **30**, 536–544 (2013).
107. Bissaro, B. *et al.* Molecular Design of Non-Leloir Furanose-Transferring Enzymes from an α -L-Arabinofuranosidase: A Rationale for the Engineering of Evolved Transglycosylases. *ACS Catal.* **5**, 4598–4611 (2015).
108. Sirimontree, P. *et al.* Mutation strategies for obtaining chitooligosaccharides with longer chains by transglycosylation reaction of family GH18 chitinase. *Biosci. Biotechnol. Biochem.* **78**, 2014–2021 (2014).

109. Placier, G., Watzlawick, H., Rabiller, C. & Mattes, R. Evolved β -Galactosidases from *Geobacillus stearothermophilus* with Improved Transgalactosylation Yield for Galacto-Oligosaccharide Production. *Appl. Environ. Microbiol.* **75**, 6312–6321 (2009).
110. Zakariassen, H., Hansen, M. C., Jøranli, M., Eijsink, V. G. H. & Sørli, M. Mutational Effects on Transglycosylating Activity of Family 18 Chitinases and Construction of a Hypertransglycosylating Mutant. *Biochemistry* **50**, 5693–5703 (2011).
111. Lundemo, P., Adlercreutz, P. & Karlsson, E. N. Improved transferase/hydrolase ratio through rational design of a family 1 β -glucosidase from *Thermotoga neapolitana*. *Appl. Environ. Microbiol.* **79**, 3400–3405 (2013).
112. Paris, G. *et al.* Probing molecular function of trypanosomal sialidases: single point mutations can change substrate specificity and increase hydrolytic activity. *Glycobiology* **11**, 305–311 (2001).
113. Oliveira, I. A., Goncalves, A. S., Neves, J. L., von Itzstein, M. & Todeschini, A. R. Evidence of Ternary Complex Formation in *Trypanosoma cruzi* trans-Sialidase Catalysis. *J. Biol. Chem.* **289**, 423–436 (2014).
114. Ochs, M. *et al.* Role of hydrophobic residues in the aglycone binding subsite of a GH39 β -xylosidase in alkyl xylosides synthesis. *J. Mol. Catal. B Enzym.* **96**, 21–26 (2013).
115. García, A. E. & Hummer, G. Water penetration and escape in proteins. *Proteins Struct. Funct. Bioinforma.* **38**, 261–272 (2000).
116. Persson, F. & Halle, B. Transient Access to the Protein Interior: Simulation versus NMR. *J. Am. Chem. Soc.* **135**, 8735–8748 (2013).
117. Teze, D. *et al.* Conserved Water Molecules in Family 1 Glycosidases: A DXMS and Molecular Dynamics Study. *Biochemistry* **52**, 5900–5910 (2013).
118. Loris, R. *et al.* Conserved water molecules in a large family of microbial ribonucleases. *Proteins Struct. Funct. Bioinforma.* **36**, 117–134 (1999).
119. Knight, J. D. R., Hamelberg, D., McCammon, J. A. & Kothary, R. The role of conserved water molecules in the catalytic domain of protein kinases. *Proteins Struct. Funct. Bioinforma.* **76**, 527–535 (2009).
120. Sreenivasan, U. & Axelsen, P. H. Buried water in homologous serine proteases. *Biochemistry* **31**, 12785–12791 (1992).
121. Prasad, B. V. L. S. & Suguna, K. Role of water molecules in the structure and function of aspartic proteinases. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 250–259 (2002).
122. Busch, S., Bruce, C. D., Redfield, C., Lorenz, C. D. & McLain, S. E. Water Mediation Is Essential to Nucleation of β -Turn Formation in Peptide Folding Motifs. *Angew. Chem. Int. Ed.* **52**, 13091–13095 (2013).
123. Rhee, Y. M., Sorin, E. J., Jayachandran, G., Lindahl, E. & Pande, V. S. Simulations of the role of water in the protein-folding mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6456–6461 (2004).
124. Helms, V. Protein Dynamics Tightly Connected to the Dynamics of Surrounding and Internal Water Molecules. *ChemPhysChem* **8**, 23–33 (2007).
125. Ball, P. Water as an Active Constituent in Cell Biology. *Chem. Rev.* **108**, 74–108 (2008).

126. Fischer, S. & Verma, C. S. Binding of buried structural water increases the flexibility of proteins. *Proc. Natl. Acad. Sci.* **96**, 9613–9615 (1999).
127. Yasuda, S., Oshima, H. & Kinoshita, M. Structural stability of proteins in aqueous and nonpolar environments. *J. Chem. Phys.* **137**, 135103 (2012).
128. Takano, K., Yamagata, Y. & Yutani, K. Buried water molecules contribute to the conformational stability of a protein. *Protein Eng. Des. Sel.* **16**, 5–9 (2003).
129. Rye, C. S. & Withers, S. G. Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580 (2000).
130. Byers, J. A. & Jamison, T. F. Entropic factors provide unusual reactivity and selectivity in epoxide-opening reactions promoted by water. *Proc. Natl. Acad. Sci.* **110**, 16724–16729 (2013).
131. Pocker, Y. Water in enzyme reactions: biophysical aspects of hydration-dehydration processes. *Cell. Mol. Life Sci. CMLS* **57**, 1008–1017
132. Fitch, C. A. *et al.* Experimental pK_a values of buried residues: analysis with continuum methods and role of water penetration. *Biophys. J.* **82**, 3289–3304 (2002).
133. Dwyer, J. J. *et al.* High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys. J.* **79**, 1610–1620 (2000).
134. Park, S. & Saven, J. G. Statistical and molecular dynamics studies of buried waters in globular proteins. *Proteins Struct. Funct. Bioinforma.* **60**, 450–463 (2005).
135. Kanaujia, S. P. & Sekar, K. Structural and functional role of water molecules in bovine pancreatic phospholipase A₂: a data-mining approach. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 74–84 (2009).
136. Wang, L. *et al.* Structural Analyses of Covalent Enzyme–Substrate Analog Complexes Reveal Strengths and Limitations of De Novo Enzyme Design. *J. Mol. Biol.* **415**, 615–625 (2012).
137. Hendil-Forsell, P., Martinelle, M. & Syrén, P.-O. Exploring water as building bricks in enzyme engineering. *Chem Commun* **51**, 17221–17224 (2015).
138. Robinson, D. *et al.* Differential Water Thermodynamics Determine PI3K-Beta/Delta Selectivity for Solvent-Exposed Ligand Modifications. *J. Chem. Inf. Model.* **56**, 886–894 (2016).
139. Breiten, B. *et al.* Water Networks Contribute to Enthalpy/Entropy Compensation in Protein–Ligand Binding. *J. Am. Chem. Soc.* **135**, 15579–15584 (2013).
140. Michel, J., Tirado-Rives, J. & Jorgensen, W. L. Energetics of displacing water molecules from protein binding sites: consequences for ligand optimization. *J. Am. Chem. Soc.* **131**, 15403–15411 (2009).
141. Vidossich, P. *et al.* On the role of water in peroxidase catalysis: a theoretical investigation of HRP compound I formation. *J. Phys. Chem. B* **114**, 5161–5169 (2010).
142. Chelikani, P., Carpena, X., Fita, I. & Loewen, P. C. An Electrical Potential in the Access Channel of Catalases Enhances Catalysis. *J. Biol. Chem.* **278**, 31290–31296 (2003).
143. Carugo, O. Statistical survey of the buried waters in the Protein Data Bank. *Amino Acids* **48**, 193–202 (2016).
144. Carugo, O. Structure and function of water molecules buried in the protein core. *Curr. Protein Pept. Sci.* **16**, 259–265 (2015).

145. Preston, G. M., Carroll, T. P., Guggino, W. B. & Agre, P. Appearance of water channels in *Xenopus* oocytes expressing red cell CHIP28 protein. *Science* **256**, 385–387 (1992).
146. Maurel, C. Plant aquaporins: Novel functions and regulation properties. *FEBS Lett.* **581**, 2227–2236 (2007).
147. Soveral, G., Prista, C., Moura, T. F. & Loureiro-Dias, M. C. Yeast water channels: an overview of orthodox aquaporins. *Biol. Cell* **103**, 35–54 (2011).
148. Day, R. E. *et al.* Human aquaporins: Regulators of transcellular water flow. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1840**, 1492–1506 (2014).
149. Agmon, N. The Grotthuss mechanism. *Chem. Phys. Lett.* **244**, 456–462 (1995).
150. Friedrich, T., Steinmüller, K. & Weiss, H. The proton-pumping respiratory complex I of bacteria and mitochondria and its homologue in chloroplasts. *FEBS Lett.* **367**, 107–111 (1995).
151. Sode, O. & Voth, G. A. Electron transfer activation of a second water channel for proton transport in [FeFe]-hydrogenase. *J. Chem. Phys.* **141**, 22D527 (2014).
152. Klimacek, M., Brunsteiner, M. & Nidetzky, B. Dynamic Mechanism of Proton Transfer in Mannitol 2-Dehydrogenase from *Pseudomonas fluorescens*: MOBILE GLU292 CONTROLS PROTON RELAY THROUGH A WATER CHANNEL THAT CONNECTS THE ACTIVE SITE WITH BULK SOLVENT. *J. Biol. Chem.* **287**, 6655–6667 (2012).
153. Jha, V., Chelikani, P., Carpena, X., Fita, I. & Loewen, P. C. Influence of main channel structure on H₂O₂ access to the heme cavity of catalase KatE of *Escherichia coli*. *Arch. Biochem. Biophys.* **526**, 54–59 (2012).
154. Garczarek, F. & Gerwert, K. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature* **439**, 109–112 (2006).
155. Rich, P. R. & Marechal, A. Functions of the hydrophilic channels in protonmotive cytochrome c oxidase. *J. R. Soc. Interface* **10**, 20130183–20130183 (2013).
156. Oprea, T. I., Hummer, G. & García, A. E. Identification of a functional water channel in cytochrome P450 enzymes. *Proc. Natl. Acad. Sci.* **94**, 2133–2138 (1997).
157. Aghajari, N., Roth, M. & Haser, R. Crystallographic Evidence of a Transglycosylation Reaction: Ternary Complexes of a Psychrophilic α -Amylase. *Biochemistry* **41**, 4273–4280 (2002).
158. Rebuffet, E. *et al.* Discovery and structural characterization of a novel glycosidase family of marine origin: Discovery of a family of glycosidases of marine origin. *Environ. Microbiol.* **13**, 1253–1270 (2011).
159. Tonozuka, T. *et al.* Crystal structure of a lactosucrose-producing enzyme, *Arthrobacter* sp. K-1 β -fructofuranosidase. *Enzyme Microb. Technol.* **51**, 359–365 (2012).
160. Bjerregaard-Andersen, K. *et al.* A Proton Wire and Water Channel Revealed in the Crystal Structure of Isatin Hydrolase. *J. Biol. Chem.* **289**, 21351–21359 (2014).
161. Sainz, G. *et al.* Interruption of the Internal Water Chain of Cytochrome *f* Impairs Photosynthetic Function. *Biochemistry* **39**, 9164–9173 (2000).
162. Syrén, P.-O., Henche, S., Eichler, A., Nestl, B. M. & Hauer, B. Squalene-hopene cyclases—evolution, dynamics and catalytic scope. *Curr. Opin. Struct. Biol.* **41**, 73–82 (2016).

163. Wittrup Larsen, M. *et al.* Suppression of Water as a Nucleophile in *Candida antarctica* Lipase B Catalysis. *ChemBioChem* **11**, 796–801 (2010).
164. Sykora, J. *et al.* Dynamics and hydration explain failed functional transformation in dehalogenase design. *Nat. Chem. Biol.* **10**, 428–430 (2014).
165. Erhardt, S., Jaime, E. & Weston, J. A Water Sluice Is Generated in the Active Site of Bovine Lens Leucine Aminopeptidase. *J. Am. Chem. Soc.* **127**, 3654–3655 (2005).
166. Pavlova, M. *et al.* Redesigning dehalogenase access tunnels as a strategy for degrading an anthropogenic substrate. *Nat. Chem. Biol.* **5**, 727–733 (2009).
167. Fishelovitch, D., Shaik, S., Wolfson, H. J. & Nussinov, R. How Does the Reductase Help To Regulate the Catalytic Cycle of Cytochrome P450 3A4 Using the Conserved Water Channel? *J. Phys. Chem. B* **114**, 5964–5970 (2010).
168. Gora, A., Brezovsky, J. & Damborsky, J. Gates of Enzymes. *Chem. Rev.* **113**, 5871–5923 (2013).
169. Frutuoso, M. A. & Marana, S. R. A single amino acid residue determines the ratio of hydrolysis to transglycosylation catalyzed by β -glucosidases. *Protein Pept. Lett.* **20**, 102–106 (2013).
170. Yan, X. & Maier, C. S. Hydrogen/deuterium exchange mass spectrometry. *Methods Mol. Biol. Clifton NJ* **492**, 255–271 (2009).
171. Teze, D. *Recherche des bases moléculaires de l'équilibre transglycosylation / hydrolyse d'une glycoside hydrolase de la famille 1.* (Université de Nantes. Faculté des sciences et des techniques, 2012).
172. Gupta, S., D'Mello, R. & Chance, M. R. Structure and dynamics of protein waters revealed by radiolysis and mass spectrometry. *Proc. Natl. Acad. Sci.* **109**, 14882–14887 (2012).
173. Maleknia, S. D. & Downard, K. M. Advances in radical probe mass spectrometry for protein footprinting in chemical biology applications. *Chem. Soc. Rev.* **43**, 3244 (2014).
174. Kim, Y. S., Liu, L., Axelsen, P. H. & Hochstrasser, R. M. 2D IR provides evidence for mobile water molecules in β -amyloid fibrils. *Proc. Natl. Acad. Sci.* **106**, 17751–17756 (2009).
175. Nucci, N. V., Pometun, M. S. & Wand, A. J. Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **18**, 245–249 (2011).
176. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
177. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).
178. Jorgensen, W. L. & Madura, J. D. Temperature and size dependence for Monte Carlo simulations of TIP4P water. *Mol. Phys.* **56**, 1381–1392 (1985).
179. Mahoney, M. W. & Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* **112**, 8910–8922 (2000).
180. Leach, A. R. *Molecular Modelling: Principles and Applications.* (Pearson Education, 2001).
181. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

182. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
183. Chovancova, E. *et al.* CAVER 3.0: A Tool for the Analysis of Transport Pathways in Dynamic Protein Structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).
184. Masood, T. B., Sandhya, S., Chandra, N. & Natarajan, V. CHEXVIS: a tool for molecular channel extraction and visualization. *BMC Bioinformatics* **16**, (2015).
185. Raunest, M. & Kandt, C. dxTuber: Detecting protein cavities, tunnels and clefts based on protein and solvent dynamics. *J. Mol. Graph. Model.* **29**, 895–905 (2011).
186. Hehemann, J.-H. *et al.* Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
187. Hehemann, J.-H. *et al.* Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J. Biol. Chem.* **287**, 30571–30584 (2012).
188. Smith, T. F. & Waterman, M. S. Identification of common molecular subsequences. *J. Mol. Biol.* **147**, 195–197 (1981).
189. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
190. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
191. Chuenchor, W. *et al.* Structural insights into rice BGlu1 beta-glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol.* **377**, 1200–1215 (2008).
192. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
193. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
194. Goujon, M. *et al.* A new bioinformatics analysis tools framework at EMBL–EBI. *Nucleic Acids Res.* **38**, W695–W699 (2010).
195. Montagna, G. *et al.* The trans-sialidase from the african trypanosome *Trypanosoma brucei*. *Eur. J. Biochem.* **269**, 2941–2950 (2002).
196. Montagna, G. N., Donelson, J. E. & Frasch, A. C. C. Procyclic *Trypanosoma brucei* expresses separate sialidase and trans-sialidase enzymes on its surface membrane. *J. Biol. Chem.* **281**, 33949–33958 (2006).
197. Tiralongo, E. *et al.* Two Trans-sialidase Forms with Different Sialic Acid Transfer and Sialidase Activities from *Trypanosoma congolense*. *J. Biol. Chem.* **278**, 23301–23310 (2003).
198. Tiralongo, E., Martensen, I., Grötzinger, J., Tiralongo, J. & Schauer, R. Trans-sialidase-like sequences from *Trypanosoma congolense* conserve most of the critical active site residues found in other trans-sialidases. *Biol. Chem.* **384**, 1203–1213 (2003).
199. Joachim Schmitt, Hosbach (DE); *et al.* Trans-sialidases obtained from *Trypanosoma congolense*. **22** (2010).

200. Koliwer-Brandl, H. *et al.* Biochemical characterization of trans-sialidase TS1 variants from *Trypanosoma congolense*. *BMC Biochem.* **12**, 39 (2011).
201. Coustou, V., Plazolles, N., Guegan, F. & Baltz, T. Sialidases play a key role in infection and anaemia in *Trypanosoma congolense* animal trypanosomiasis. *Cell. Microbiol.* **14**, 431–445 (2012).
202. Guegan, F., Plazolles, N., Baltz, T. & Coustou, V. Erythrophagocytosis of desialylated red blood cells is responsible for anaemia during *Trypanosoma vivax* infection. *Cell. Microbiol.* **15**, 1285–1303 (2013).
203. Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
204. Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718 (2005).
205. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655 (2008).
206. Lindorff-Larsen, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **78**, 1950–1958 (2010).
207. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
208. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
209. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
210. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463–1472 (1997).
211. Amaya, M. F. *et al.* Structural insights into the catalytic mechanism of *Trypanosoma cruzi* trans-sialidase. *Struct. Lond. Engl.* **1993** **12**, 775–784 (2004).
212. Amaya, M. F., Buschiazzo, A., Nguyen, T. & Alzari, P. M. The high resolution structures of free and inhibitor-bound *Trypanosoma rangeli* sialidase and its comparison with *T. cruzi* trans-sialidase. *J. Mol. Biol.* **325**, 773–784 (2003).
213. Sali, A. Comparative protein modeling by satisfaction of spatial restraints. *Mol. Med. Today* **1**, 270–277 (1995).
214. Fiser, A., Do, R. K. G. & Šali, A. Modeling of loops in protein structures. *Protein Sci.* **9**, 1753–1773 (2000).
215. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5**, 367 (2012).
216. Zechel, D. L. *et al.* Iminosugar glycosidase inhibitors: structural and thermodynamic dissection of the binding of isofagomine and 1-deoxynojirimycin to beta-glucosidases. *J. Am. Chem. Soc.* **125**, 14313–14323 (2003).
217. Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *ResearchGate* **97**, 10269–10280 (1992).

218. Mark, P. & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **105**, 9954–9960 (2001).
219. Dion, M., Fourage, L., Hallet, J.-N. & Colas, B. Cloning and expression of a β -glycosidase gene from *Thermus thermophilus*. Sequence and biochemical characterization of the encoded enzyme. *Glycoconj. J.* **16**, 27–37 (1999).
220. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
221. Zoi, I. *et al.* Modulating Enzyme Catalysis through Mutations Designed to Alter Rapid Protein Dynamics. *J. Am. Chem. Soc.* **138**, 3403–3409 (2016).
222. Luk, L. Y. P., Loveridge, E. J. & Allemann, R. K. Protein motions and dynamic effects in enzyme catalysis. *Phys Chem Chem Phys* **17**, 30817–30827 (2015).
223. Luo, M., Li, L. & Schramm, V. L. Remote Mutations Alter Transition-State Structure of Human Purine Nucleoside Phosphorylase †. *Biochemistry* **47**, 2565–2576 (2008).
224. Krynicki, K., Green, C. D. & Sawyer, D. W. Pressure and temperature dependence of self-diffusion in water. *Faraday Discuss. Chem. Soc.* **66**, 199 (1978).
225. Brás, N. F., Fernandes, P. A. & Ramos, M. J. QM/MM Studies on the β -Galactosidase Catalytic Mechanism: Hydrolysis and Transglycosylation Reactions. *J. Chem. Theory Comput.* **6**, 421–433 (2010).
226. Wang, J. *et al.* Theoretical study on the deglycosylation mechanism of rice BGLu1 β -glucosidase. *Int. J. Quantum Chem.* **113**, 1071–1075 (2013).
227. Jam, M. *et al.* The endo- β -agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours. *Biochem. J.* **385**, 703–713 (2005).
228. Baumann, M. J. *et al.* Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: biological implications for cell wall metabolism. *Plant Cell* **19**, 1947–1963 (2007).
229. Eklof, J. M. & Brumer, H. The XTH Gene Family: An Update on Enzyme Structure, Function, and Phylogeny in Xyloglucan Remodeling. *PLANT Physiol.* **153**, 456–466 (2010).
230. Qin, Z. *et al.* Catalytic Mechanism of a Novel Glycoside Hydrolase Family 16 ‘Elongating’ β -Transglycosylase. *J. Biol. Chem.* **292**, 1666–1678 (2017).
231. Fu, X. T. & Kim, S. M. Agarase: Review of Major Sources, Categories, Purification Method, Enzyme Characteristics and Applications. *Mar. Drugs* **8**, 200–218 (2010).
232. Chi, W.-J., Chang, Y.-K. & Hong, S.-K. Agar degradation by microorganisms and agar-degrading enzymes. *Appl. Microbiol. Biotechnol.* **94**, 917–930 (2012).
233. Leaver-Fay, A. *et al.* ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
234. Rod, T. H., Radkiewicz, J. L. & Brooks, C. L. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proc. Natl. Acad. Sci.* **100**, 6980–6985 (2003).

235. Edwards, S. J., Soudackov, A. V. & Hammes-Schiffer, S. Impact of Distal Mutation on Hydrogen Transfer Interface and Substrate Conformation in Soybean Lipoxygenase. *J. Phys. Chem. B* **114**, 6653–6660 (2010).
236. Saen-Oon, S., Ghanem, M., Schramm, V. L. & Schwartz, S. D. Remote Mutations and Active Site Dynamics Correlate with Catalytic Properties of Purine Nucleoside Phosphorylase. *Biophys. J.* **94**, 4078–4088 (2008).
237. Warshel, A. *et al.* Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
238. Merceron, R. *et al.* The molecular mechanism of thermostable α -galactosidases AgaA and AgaB explained by x-ray crystallography and mutational studies. *J. Biol. Chem.* **287**, 39642–39652 (2012).
239. Buschiazzo, A. & Alzari, P. M. Structural insights into sialic acid enzymology. *Curr. Opin. Chem. Biol.* **12**, 565–572 (2008).
240. Freire-de-Lima, L. *et al.* Sialic acid: a sweet swing between mammalian host and *Trypanosoma cruzi*. *Front. Immunol.* **3**, 356 (2012).
241. Oliveira, I. A., Freire-de-Lima, L., Penha, L. L., Dias, W. B. & Todeschini, A. R. *Trypanosoma cruzi* Trans-sialidase: structural features and biological implications. *Subcell. Biochem.* **74**, 181–201 (2014).
242. Buschiazzo, A., Amaya, M. F., Cremona, M. L., Frasch, A. C. & Alzari, P. M. The crystal structure and mode of action of trans-sialidase, a key enzyme in *Trypanosoma cruzi* pathogenesis. *Mol. Cell* **10**, 757–768 (2002).
243. Dc-Rubin, S. S. C. & Schenkman, S. T. *Trypanosoma cruzi* trans-sialidase as a multifunctional enzyme in Chagas' disease. *Cell. Microbiol.* **14**, 1522–1530 (2012).
244. Holck, J. *et al.* Enzyme catalysed production of sialylated human milk oligosaccharides and galactooligosaccharides by *Trypanosoma cruzi* trans-sialidase. *New Biotechnol.* **31**, 156–165 (2014).
245. Watts, A. G., Opezzo, P., Withers, S. G., Alzari, P. M. & Buschiazzo, A. Structural and kinetic analysis of two covalent sialosyl-enzyme intermediates on *Trypanosoma rangeli* sialidase. *J. Biol. Chem.* **281**, 4149–4155 (2006).
246. Demir, O. & Roitberg, A. E. Modulation of catalytic function by differential plasticity of the active site: case study of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase. *Biochemistry* **48**, 3398–3406 (2009).
247. Mitchell, F. L., Miles, S. M., Neres, J., Bichenkova, E. V. & Bryce, R. A. Tryptophan as a molecular shovel in the glycosyl transfer activity of *Trypanosoma cruzi* trans-sialidase. *Biophys. J.* **98**, L38–40 (2010).
248. Nordvang, R. T. *et al.* It All Starts with a Sandwich: Identification of Sialidases with Trans-Glycosylation Activity. *PLoS One* **11**, e0158434 (2016).
249. Buschiazzo, A. *et al.* Structural basis of sialyltransferase activity in trypanosomal sialidases. *EMBO J.* **19**, 16–24 (2000).
250. Damager, I. *et al.* Kinetic and mechanistic analysis of *Trypanosoma cruzi* trans-sialidase reveals a classical ping-pong mechanism with acid/base catalysis. *Biochemistry* **47**, 3507–3512 (2008).
251. Watts, A. G. *et al.* *Trypanosoma cruzi* trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile. *J. Am. Chem. Soc.* **125**, 7532–7533 (2003).

252. Paris, G. *et al.* A sialidase mutant displaying trans-sialidase activity. *J. Mol. Biol.* **345**, 923–934 (2005).
253. Pierdominici-Sottile, G., Palma, J. & Roitberg, A. E. Free-energy computations identify the mutations required to confer trans-sialidase activity into *Trypanosoma rangeli* sialidase. *Proteins* **82**, 424–435 (2014).
254. Jers, C. *et al.* Rational design of a new *Trypanosoma rangeli* trans-sialidase for efficient sialylation of glycans. *PLoS One* **9**, e83902 (2014).
255. Bueren-Calabuig, J. A., Pierdominici-Sottile, G. & Roitberg, A. E. Unraveling the differences of the hydrolytic activity of *Trypanosoma cruzi* trans-sialidase and *Trypanosoma rangeli* sialidase: a quantum mechanics-molecular mechanics modeling study. *J. Phys. Chem. B* **118**, 5807–5816 (2014).
256. Shinobu, A. & Agmon, N. The Hole in the Barrel: Water Exchange at the GFP Chromophore. *J. Phys. Chem. B* **119**, 3464–3478 (2015).
257. Nyffenegger, C., Nordvang, R. T., Jers, C., Meyer, A. S. & Mikkelsen, J. D. Design of *Trypanosoma rangeli* sialidase mutants with improved trans-sialidase activity. *PLOS ONE* **12**, e0171585 (2017).
258. Briones, M. R. S., Egima, C. M., Eichinger, D. & Schenkman, S. Trans-sialidase genes expressed in mammalian forms of *Trypanosoma cruzi* evolved from ancestor genes expressed in insect forms of the parasite. *J. Mol. Evol.* **41**, 120–131 (1995).
259. Harms, M. J. & Thornton, J. W. Analyzing protein structure and function using ancestral gene reconstruction. *Curr. Opin. Struct. Biol.* **20**, 360–366 (2010).
260. Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
261. Hall, B. G. Building Phylogenetic Trees from Molecular Data with MEGA. *Mol. Biol. Evol.* **30**, 1229–1235 (2013).
262. Tamura, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
263. Ren, P., Ponder, J. W. & others. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **107**, 5933–5947 (2003).

Liste des illustrations

◆ Liste des figures

FIGURE 1. EXEMPLES D'OLIGOSACCHARIDES RENCONTRES DANS LE MONDE VIVANT	22
FIGURE 2. QUELQUES ROLES MAJEURS EXERCES PAR LES GLYCOPROTEINES	23
FIGURE 3. MECANISME A RETENTION DE CONFIGURATION DES GT	25
FIGURE 4. MECANISME A INVERSION DE CONFIGURATION DES GT	26
FIGURE 5. MECANISME CATALYTIQUE A RETENTION DE CONFIGURATION DES GH.....	27
FIGURE 6. MECANISME CATALYTIQUE A INVERSION DE CONFIGURATION DES GH	28
FIGURE 7. MECANISME CATALYTIQUE DES N-ACETYL-B-HEXOSAMINIDASES AVEC ASSISTANCE PAR LE SUBSTRAT	29
FIGURE 8. MECANISME REACTIONNEL DE L'HYDROLYSE ET DE LA TRANSGLYCOSYLATION	29
FIGURE 9. MECANISME CATALYTIQUE D'UNE GLYCOSYNTHESE.....	32
FIGURE 10. MECANISME CATALYTIQUE D'UNE GLYCOSYNTHESE AVEC ASSISTANCE DU SUBSTRAT.....	33
FIGURE 11. TRANSGLYCOSYLATION D'UNE PROTEINE CATALYSEE PAR UNE ENGASES.....	34
FIGURE 12. MECANISME CATALYTIQUE D'UNE THIOGLYCOLIGASE	35
FIGURE 13. MECANISME CATALYTIQUE D'UNE THIOGLYCOSYNTHESE.....	35
FIGURE 14. NOMENCLATURE DES SITES DE LIAISON DES GH AUX SUCRES DONNEURS ET ACCEPTEURS	37
FIGURE 15. QUELQUES ROLES BIOLOGIQUES DE L'EAU	40
FIGURE 16. QUELQUES CANAUX D'EAU CRISTALLOGRAPHIQUES RENCONTRES CHEZ DIVERSES ENZYMES	42
FIGURE 17. MISE EN EVIDENCE DE PEPTIDES DE LA TTBGly ACCESSIBLES AU SOLVANT PAR DXMS	45
FIGURE 18. MISE EN EVIDENCE DE PEPTIDES SOLVATES DE LA CYTOCHROME-C-OXYDASE PAR RPMS	46
FIGURE 19. MISE EN EVIDENCE D'UN TRANSFERT DE PROTON ENTRE LE RESIDU D85 DE LA BACTERORHODOPSINE ET UNE MOLECULE D'EAU PAR SPECTROSCOPIE FTIR	47
FIGURE 20. CARACTERISATION DES SITES D'HYDRATATION INTERNES A L'UBIQUITINE PAR RMN 3D EN MICELLE INVERSE	48

FIGURE 21. DEFINITION D'UN SYSTEME TOUT-ATOME EN SIMULATION DE DYNAMIQUE MOLECULAIRE	49
FIGURE 22. STRUCTURES DE MOLECULE D'EAU UTILISEES EN SIMULATION DE DYNAMIQUE MOLECULAIRE	50
FIGURE 23. ENSEMBLE DES TYPES D'INTERACTIONS LIEES	51
FIGURE 24. CARTE DU PLASMIDE PBTAC2 UTILISE POUR L'EXPRESSION DE LA TTBGLY A TITRE D'EXEMPLE	64
FIGURE 25. SCHEMA RECAPITULATIF DE L'ENSEMBLE DES ETAPES EXPERIMENTALES APPLIQUEES POUR LA PRODUCTION ET LA CARACTERISATION DE TTBGLY A TITRE D'EXEMPLE	65
FIGURE 26. EXEMPLE DE GEL SDS-PAGE OBTENU POUR LE CONTROLE DE LA PURIFICATION DE LA TTBGLY	71
FIGURE 27. EXEMPLE DE CHROMATOGRAMME OBTENU PAR CHROMATOGRAPHIE SUR COUCHE MINCE	73
FIGURE 28. PROFIL TYPE D'ELECTROPHORESE CAPILLAIRE DANS LE CAS D'UN SUIVI CINETIQUE DE LA TTBGLY SAUVAGE A TITRE D'EXEMPLE	78
FIGURE 29. REPRESENTATION SCHEMATIQUE D'UNE LIAISON HYDROGENE	90
FIGURE 30. ANGLES DE TORSION (X1, X2, X3) DE LA CHAINE LATERALE D'UN RESIDU ASPARTATE	91
FIGURE 31. IDENTIFICATION DES SOUS-RESEAUX D'EAU INTERNE SUR UN CLICHE DE TRAJECTOIRE	92
FIGURE 32. RECONSTITUTION DES RESEAUX D'EAU INTERNES	93
FIGURE 33. EXEMPLE DE TRACE DE FONCTIONS DE SURVIE DE L'EAU INTERNE A UN CANAL ENFOUI AU SEIN DE LA CONNEXINE HUMAINE CX26	95
FIGURE 34. EXEMPLE DE DETERMINATION DE COEFFICIENT DE DIFFUSION LATERAL DE MOLECULES D'EAU ENTRE DEUX SURFACES CHARGEES	96
FIGURE 35. TOPOLOGIE DE SURFACE DE TTBGLY ET LOCALISATION DU SITE CATALYTIQUE	99
FIGURE 36. STRUCTURE DE TTBGLY ET RESIDUS CONSERVES DU SOUS-SITE -1	100
FIGURE 37. ARBRE PHYLOGENETIQUE CONSTRUIT A PARTIR DE L'ALIGNEMENT DES SEQUENCES DE GH1	102
FIGURE 38. CINETIQUE DE TRANSGLYCOSYLATION DE LA TTBGLY SAUVAGE ET D'UNE SELECTION DE MUTANTS TRANSGLYCOSIDASES	105
FIGURE 39. RENDEMENTS MAXIMAUX DE TRANSGLYCOSYLATION D'UNE SELECTION DE MUTANTS	106
FIGURE 40. PROFIL CINETIQUE DE LA TTBGLY SAUVAGE ET D'UNE SELECTION DE MUTANTS TRANSGLYCOSIDASES EN PRESENCE DE DIVERSES CONCENTRATIONS EN SUBSTRAT	107
FIGURE 41. EFFICACITES CATALYTIQUES D'UNE SELECTION DE MUTANTS TRANSGLYCOSIDASES	108
FIGURE 42. MODELE CINETIQUE DES REACTIONS CATALYSEES PAR LA TTBGLY	109

FIGURE 43. TEST DE MUTANTS DE LA POSITION 167 PAR CHROMATOGRAPHIE SUR COUCHE MINCE.....	111
FIGURE 44. RMSD CALCULE SUR LE SQUELETTE PEPTIDIQUE	113
FIGURE 45. PROJECTION STRUCTURALE DU RMSD DU SQUELETTE PEPTIDIQUE DES REGIONS DE LA PROTEINE MONTRANT DES DIFFERENCES DE CONTACT AVEC L'EAU ENFOUIE ENTRE LA TTBGLY SAUVAGE ET CHAQUE MUTANT	114
FIGURE 46. RMSD DU SQUELETTE PEPTIDIQUE DES REGIONS ENFOUIES SUBISSANT LES PLUS FORTS CHANGEMENTS CONFORMATIONNELS.....	115
FIGURE 47. RMSF MOYENNE SUR LES ATOMES DU SQUELETTE PEPTIDIQUE.....	116
FIGURE 48. PROJECTION STRUCTURALE DES FACTEURS B MOYENS DE CHAQUE RESIDU SUR LE SQUELETTE PEPTIDIQUE DES STRUCTURES DYNAMIQUES MOYENNES	117
FIGURE 49. ANGLES DE TORSION (X1, X3) DE LA CHAINE LATERALE DU RESIDU ACIDE-BASE CATALYTIQUE (E164)	119
FIGURE 50. ANGLES DE TORSION (X1, X3) DE LA CHAINE LATERALE DU RESIDU NUCLEOPHILE CATALYTIQUE (E338)	120
FIGURE 51. INTERACTIONS DE RESIDUS CONSERVES DU SOUS-SITE -1 AVEC L'ACIDE-BASE CATALYTIQUE	121
FIGURE 52. ANGLES DE TORSION (X1, X2) DE LA CHAINE LATERALE DU RESIDU Y284.....	122
FIGURE 53. ANGLES DE TORSION (X1, X2) DE LA CHAINE LATERALE DU RESIDU H119	123
FIGURE 54. ANGLES DE TORSION (X1, X2) DE LA CHAINE LATERALE DES RESIDUS H178 ET W120	124
FIGURE 55. SUIVI DE LA PLUS PETITE DISTANCE INTERATOMIQUE ENTRE LES CHAINES LATERALES DES RESIDUS W120 ET H178.....	125
FIGURE 56. DEPLACEMENT QUADRATIQUE MOYEN DES MOLECULES D'EAU INTERNES.....	127
FIGURE 57. DISTRIBUTION DU TEMPS MIS PAR LES MOLECULES D'EAU INTERNES A TTBGLY POUR PARCOURIR UNE DISTANCE D'AU MINIMUM 3 Å AU SEIN DE LA PROTEINE	128
FIGURE 58. LOCALISATION DES CLUSTERS DE MOLECULES D'EAU CRISTALLOGRAPHIQUE CHEZ LA TTBGLY SAUVAGE	130
FIGURE 59. CANAUX D'EAU ISOLES SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE CHEZ TTBGLY.....	131
FIGURE 60. CANAL D'EAU PRINCIPAL (OU CANAL 1) ISOLE SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE CHEZ LA TTBGLY SAUVAGE.....	132
FIGURE 61. TRAJECTOIRE DU CANAL PRINCIPAL ISOLE SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE CHEZ TTBGLY	133
FIGURE 62. LOCALISATION DES MOLECULES D'EAU POTENTIELLEMENT CATALYTIQUES.....	136

FIGURE 63. PROPORTION DE MOLECULES D'EAU ISSUES DU CANAL 4 PARMIS L'ENSEMBLE DES MOLECULES D'EAU INTERNES ECHANTILLONNEES AU POINT D'ATTAQUE NUCLEOPHILE ET TEMPS DE RESIDENCE MAXIMAL DE CES MOLECULES AU SEIN DU CANAL 4	137
FIGURE 64. NOMBRE DE VISITE(S) DES MOLECULES D'EAU ISSUES DU CANAL 4 OU DU MILIEU AU POINT D'ATTAQUE NUCLEOPHILE.....	137
FIGURE 65. TRAJECTOIRE DE DEUX MOLECULES D'EAU EMPRUNTANT LE CANAL 4 POUR ACCEDER AU POINT D'ATTAQUE NUCLEOPHILE CHEZ LA TTBGLY SAUVAGE	138
FIGURE 66. TRAJECTOIRE DES CANAUX 3 ET 4 ISOLEES SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE	140
FIGURE 67. TEMPS DE PURGE DES MOLECULES D'EAU AU SEIN DES CANAUX DESSERVANT LE SOUS-SITE -1	141
FIGURE 68. COURBES DE SURVIE DE L'EAU INTERNE AUX CANAUX DESSERVANT LE SOUS-SITE -1	142
FIGURE 69. STRUCTURE D'AGAD ET RESIDUS CONSERVES DU SOUS-SITE -1.....	147
FIGURE 70. CANAUX D'EAU CRISTALLOGRAPHIQUES POTENTIELS.....	149
FIGURE 71. CANAL D'EAU POTENTIEL FOURNISSANT LES MOLECULES D'EAU CATALYTIQUES AU SOUS-SITE -1	150
FIGURE 72. DISPOSITION CONSERVEE DES MOLECULES D'EAU CRISTALLOGRAPHIQUES EN BORDURE DU SOUS-SITE -1 CHEZ TTBGLY ET AGAD	151
FIGURE 73. MECANISME REACTIONNEL DE L'AGAD SAUVAGE.....	152
FIGURE 74. TEST D'ACTIVITE D'AGAD EN PRESENCE D'AGARODECAOSE ET D'AGAROOCTAOSE.....	154
FIGURE 75. SEPARATION DES PRODUITS DE REACTION PAR HPAEC-PAD	156
FIGURE 76. BALANCE DES VITESSES INITIALES D'HYDROLYSE ET DE TRANSGLYCOSYLATION	157
FIGURE 77. RMSD CALCULE SUR LE SQUELETTE PEPTIDIQUE	160
FIGURE 78. RMSD CALCULE SUR LE SQUELETTE PEPTIDIQUE DE LA REGION MUTEE	161
FIGURE 79. PROJECTION STRUCTURALE DES ECARTS MOYENS DE RMSD CALCULES SUR LE SQUELETTE PEPTIDIQUE DES REGIONS DE LA PROTEINE MONTRANT DES DIFFERENCES DE CONTACT AVEC L'EAU ENFOUIE ENTRE L'AGAD SAUVAGE ET CHAQUE MUTANT	162
FIGURE 80. RMSF CALCULE SUR LE SQUELETTE PEPTIDIQUE ET PROJECTION STRUCTURALE DES VALEURS CALCULEES SUR LES ATOMES DU SQUELETTE PEPTIDIQUE	163
FIGURE 81. ANGLES DE TORSION (X1, X3) DE LA CHAINE LATERALE DU RESIDU ACIDE-BASE CATALYTIQUE (E179)	164
FIGURE 82. LOCALISATION DES RESIDUS ETUDIES SUR LA STRUCTURE CRISTALLOGRAPHIQUE.....	166
FIGURE 83. ANGLES DE TORSION (X1, X2) DE LA CHAINE LATERALE DU RESIDU Y181.....	167
FIGURE 84. ANGLES DE TORSION (X1, X2) DE LA CHAINE LATERALE DU RESIDU D176	168

FIGURE 85. ANGLES DE TORSION (X1, X3) DE LA CHAÎNE LATÉRALE DU RÉSIDU Q214	169
FIGURE 86. ANGLES DE TORSION (X1, X3) DE LA CHAÎNE LATÉRALE DU RÉSIDU Q342	169
FIGURE 87. LIAISONS HYDROGÈNE STABILISANT LE RÉSIDU ACIDE-BASE CATALYTIQUE	171
FIGURE 88. DÉPLACEMENT QUADRATIQUE MOYEN DES MOLECULES D'EAU INTERNES A LA STRUCTURE PROTÉIQUE D'AGAD	174
FIGURE 89. DISTRIBUTION DU TEMPS MIS PAR LES MOLECULES D'EAU INTERNES A AGAD POUR PARCOURIR UNE DISTANCE D'AU MINIMUM 3 Å AU SEIN DE LA PROTÉINE	175
FIGURE 90. ENSEMBLE DES POCHEs ET CANAUX D'EAU POTENTIELS ISOLÉS SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE	176
FIGURE 91. ENSEMBLE DES POSITIONS SPÉCIFIQUEMENT EXPLORÉES PAR LES MOLECULES D'EAU INTERNES CHEZ LES MUTANTS	177
FIGURE 92. TRAJECTOIRE DU CANAL D'EAU POTENTIEL MIS EN ÉVIDENCE PAR SIMULATION DE DYNAMIQUE MOLECULAIRE	178
FIGURE 93. TEMPS DE PURGE DE L'EAU AU SEIN DU CANAL ÉTUDIÉ	179
FIGURE 94. COURBES DE SURVIE DE L'EAU AU SEIN DU CANAL ÉTUDIÉ	180
FIGURE 95. STRUCTURE 3D DE TCTS ET TRSA COMPLEXÉES AU SIALYL-LACTOSE	186
FIGURE 96. MÉCANISME RÉACTIONNEL DE TCTS	188
FIGURE 97. RMSD CALCULÉ SUR LE SQUELETTE PEPTIDIQUE DE TCTS ET TRSA	190
FIGURE 98. RMSF CALCULÉ SUR LE SQUELETTE PEPTIDIQUE ET L'ENSEMBLE DE LA PROTÉINE	191
FIGURE 99. ANGLES DE TORSION DE LA CHAÎNE LATÉRALE DES RÉSIDUS CATALYTIQUES CHEZ TCTS ET TRSA	193
FIGURE 100. ENSEMBLE DES CANAUX D'EAU POTENTIELS ISOLÉS SUR 500 NS DE TRAJECTOIRE DE DYNAMIQUE MOLECULAIRE AU SEIN DU DOMAINE CATALYTIQUE DE TCTS ET TRSA	194
FIGURE 101. COMPARAISON DE TROIS CANAUX DU DOMAINE CATALYTIQUE DE TCTS ET TRSA	195
FIGURE 102. RÉSIDUS FONCTIONNELS BORDANT LES CANAUX D'EAU INTERNES A TRSA ET TCTS	196
FIGURE 103. APÉRÇU DU CANAL 1 ET LOCALISATION DU RÉSIDU 96	197
FIGURE 104. APÉRÇU DU CANAL 2 ET DU RÉSEAU DE LIAISONS HYDROGÈNE IMPLIQUANT H304/305	199
FIGURE 105. IMPACT DU RÉSIDU 276/277 CHEZ TCTS/TRSA SUR LA STABILISATION DE H304/305	200
FIGURE 106. CONFORMATION DU RÉSIDU R314/315 EN SORTIE DU CANAL 2 CHEZ TCTS/TRSA	201
FIGURE 107. LOCALISATION DU CANAL 3 ET DES RÉSIDUS VICINAUX RÉPERTORIÉS DANS LA LITTÉRATURE	202

FIGURE 108. APERÇU DU CANAL 3 ET DU RESEAU DE LIAISONS HYDROGENE IMPLIQUANT H (123/124)	203
FIGURE 109. DEPLACEMENT QUADRATIQUE MOYEN DES MOLECULES D'EAU INTERNES AU DOMAINE CATALYTIQUE DE TRSA ET TCTS.....	204
FIGURE 110. DISTRIBUTION DU TEMPS MIS PAR LES MOLECULES D'EAU INTERNES A TCTS ET TRSA POUR PARCOURIR UNE DISTANCE D'AU MINIMUM 3 Å AU SEIN DE LA PROTEINE.....	205
FIGURE 111. ALIGNEMENT MULTIPLE DE SEQUENCES DE SIALIDASES ET TRANS-SIALIDASES DE T. RANGELI ET T. CRUZI.....	207
FIGURE 112. CARTOGRAPHIE DES POSITIONS MUTEES ET VARIABLES ENTRE TCTS ET TRSA.....	210
FIGURE 113. LOCALISATION DES RESIDUS S115, S118, ET L313 SUR LE DOMAINE CATALYTIQUE.....	211
FIGURE 114. CARACTERISATION DE L'ACTIVITE DES MUTANTS DE TCTS.....	212
FIGURE 115. PROFIL CINETIQUE DE TCTS ET DE SES MUTANTS.....	213
FIGURE 116. DISTRIBUTION DE LONGUEUR DES SEQUENCES DE DOMAINES CATALYTIQUES DE (TRANS)SIALIDASES DE TRYPANOSOMES.....	215
FIGURE 117. ARBRE PHYLOGENETIQUE ENRACINE DES DOMAINES CATALYTIQUES DE (TRANS)SIALIDASES DE TRYPANOSOMES.....	218

◆ Liste des tableaux

TABLEAU 1. EXEMPLES DE TRANSGLYCOSIDASES REPERTORIEES DANS CAZY.....	30
TABLEAU 2. SEQUENCES DE GH33 DE TRYPANOSOMES COLLECTEES DANS LA BASE DE DONNEES CAZY ET ISSUES DE LA LITTERATURE.....	81
TABLEAU 3. POSITIONS CIBLEES EN MUTAGENESE LE LONG DU CANAL D'EAU PRINCIPAL.....	103
TABLEAU 4. PROPORTION DES ACIDES AMINES BORDANT LES CANAUX D'EAU.....	131
TABLEAU 5. CINETIQUE A L'EQUILIBRE D'AGAD EN PRESENCE DE DECA-AGAROSE.....	153
TABLEAU 6. EFFET DES MUTATIONS SUR LA CINETIQUE A L'EQUILIBRE D'AGAD EN PRESENCE DE DECA-AGAROSE.....	155
TABLEAU 7. EFFETS DES MUTATIONS SUR LA DYNAMIQUE CONFORMATIONNELLE DES RESIDUS ETUDIES PRECEDEMMENT.....	170
TABLEAU 8. LISTE DES POSITIONS DIVERGENTES ENTRE LES SEQUENCES DE SIALIDASES ET DE TRANS-SIALIDASES DE T. RANGELI ET T. CRUZI.....	209
TABLEAU 9. CONSTANTES CINETIQUES DE TCTS ET DE SES MUTANTS.....	214

Thèse de Doctorat

Résumé

Catalyseurs de la dégradation de polysaccharides dans le cadre de diverses applications industrielles, de nombreuses glycoside hydrolases (GH) possèdent également une activité de transglycosylation qui peut être exploitée pour la synthèse d'oligosaccharides. Afin d'augmenter cette activité, minoritaire par rapport à l'hydrolyse, des expériences de mutagenèse rationnelle peuvent être employées. Toutefois, l'ensemble des bases moléculaires régissant l'équilibre entre ces deux activités reste en revanche difficile à élucider. L'étude de quatre GH (Tt β gly, AgaD, TcTS, TrSA) par simulation de dynamique moléculaire a permis la découverte de canaux d'eau internes à leurs structures et connectant le site actif au milieu. Cette observation suggère que les canaux d'eau internes aux GH pourraient être impliqués dans leur activité d'hydrolyse. Plusieurs paires de résidus bordant deux de ces canaux ont été mis en évidence chez Tt β gly et AgaD et semblent contrôler le passage de l'eau du canal vers le site actif. La mutagenèse de ces résidus a été entreprise afin de tenter d'augmenter l'activité de transglycosylation chez ces deux enzymes. Une réduction de l'hydrolyse d'un facteur 7 et 50 au profit de l'activité de transglycosylation a été caractérisée chez les deux meilleurs mutants de Tt β gly et AgaD, respectivement. L'analyse des simulations a révélé que ces résultats étaient corrélés à une augmentation de la dynamique des molécules d'eau internes aux deux canaux étudiés. Cette étude souligne ainsi l'importance fonctionnelle de l'eau interne aux hydrolases et suggère que l'ingénierie de sa dynamique peut constituer une approche originale pour convertir les GH en transglycosidases.

Mots clés

balance hydrolyse/transglycosylation, glycosidase, transglycosidase, eau interne, dynamique moléculaire, mutagenèse rationnelle

Abstract

Known for their ability to hydrolyse glycosidic linkages, numerous glycoside hydrolases (GH) are also able to catalyse transglycosylation reaction which can be harnessed for the synthesis of oligosaccharides. Although in the vast majority of cases hydrolysis prevails over transglycosylation reaction, the latter has already been increased through mutagenesis and directed evolution experiments. However, little is known about the regulation of the balance between both activities. We discover, via molecular dynamics (MD) simulations, several potential water channels connecting the bulk to the active site in four GH: Tt β gly, AgaD and the pair of homologs TcTS/TrSA. This observation supports the hypothesis that channels could be involved in the hydrolytic activity of GH. Amino acid residues forming bottlenecks at the interface between two water channels and the active site in Tt β gly and AgaD were suspected to control water access from the bulk to the channel interior and the active site. Mutagenesis of key amino acids in the vicinity of selected channels was performed in order to attempt to increase the transglycosylation/hydrolysis ratio balance. Characterization of the best mutants showed a 7 and 50 fold decrease of hydrolysis compared to the wild type for Tt β gly and AgaD respectively, while the transglycosylase activity was improved. MD simulations showed that these modifications were correlated with greater water dynamics in the corresponding channels. These results highlight the importance of water dynamics in hydrolases catalysis and suggest that modifying the protein internal water dynamics could serve as a generic approach to engineer transglycosylase activity in GH.

Key Words

transglycosylation/hydrolysis balance, glycosidase, transglycosidase, water dynamics, molecular dynamics simulations, rational mutagenesis