

Bio informatique structurale : prédiction de la structure 3D d'une protéine de la famille des MFS : A1A5C7 (solute carrier family 22 member 23)

Nicolas BRUNEAU

Master 2 en Bioinformatique, Unité de Formation et de Recherche en Sciences de la Vie, Université PARIS VII

Résumé

Les protéines de transport transmembranaires sont généralement constitués d'hélices alpha, pour permettre le passage de molécule à travers la membrane. SLC22A23 (code UniProt : A1A5C7) fait partie de cette famille et permet le transport de molécule organique anionique. Après avoir identifié 12 hélices alpha transmembranaires grâce à différents outils de prédiction secondaire nous avons confronté ses données aux informations qui nous étaient disponibles pour cette protéine. À l'issue nous avons généré des modèles de structure de cette protéine par une méthode de 'threading'. Enfin, nous avons analysé nos différents modèles et sommes parvenues à la conclusion que le modèle obtenu avec RaptorX correspondait le plus à ce que nous attendions par rapport aux expérimentations déjà obtenues pour cette protéine. Puis avec ce modèle, nous avons affiné les chaînes latérales, pour par la suite effectuer une analyse des mouvements de la protéine par la technique des 'Modes Normaux' pour comprendre les mécanismes correspondant aux mouvements de la protéine, pour faire passer les molécules organiques anioniques à travers la membrane.

Introduction

SLC22 (solute carrier family 22) est une famille de transporteur de molécule organique cationique, anionique et zwitterionique. Ces protéines contiennent 12 hélices alpha transmembranaires traversant la membrane et une boucle large extracellulaire (1). Ces transporteurs peuvent être de type symport (passage actif de deux molécules différent en même temps), antiport (passage actif de deux molécules en sens inverse) et uniport (transporteur passif, contrairement au symport et antiport, et permettant le passage d'une molécule / ion, à travers la membrane). SLC22A23 (UniProt : A1A5C7) a une expression prédominante dans les organes participant à l'absorption et à l'excrétion et particulièrement dans le foie et les reins. Ces transporteurs interviennent dans les voies d'excrétion des toxines extérieures, des substances produites endogènes, et des drogues (2). SLC22 fait partie de la superfamille des MFS (Major facilitator superfamily). La famille des SLC est divisée en 3 catégories : OCTs (organic cation transporters), OCTNs (organic cation / carnitine transporters), et les OATs (organic anion transporter). C'est de cette dernière subdivision qu'appartient la SLC22A23. Les OATs sont responsables principalement de l'élimination du glutarates, salicylates, et de certaines drogues (3).

Une compréhension des mécanismes d'interactions entre ce récepteur et ces ligands est importante afin de mieux comprendre les principes d'excrétions de certaines drogues et ainsi anticiper sur la toxicité de certaines molécules.

A ce jour, aucune structure issue d'imagerie par cristallographie n'est disponible pour cette protéine. Cependant, des données biologiques expérimentales sont connues. En connaissance de ces données, nous allons donc tenter de créer un modèle qui permettra de décrire au mieux la structure de cette protéine, afin que par la suite nous tentions de comprendre les mouvements de cette protéine et des mouvements permettant le passage de molécule à travers la membrane.

Matériels & Méthodes

Nous avons d'abord cherché à obtenir les séquences les plus proches de notre séquence références et possédant des structures 3D connues. C'est pour cela que nous avons effectué un BLAST (E-threshold:10, Matrice de substitution : Blosum 62, avec gaps) et sur la base de donnée cible '3D structure (PDB)'. Cela dans le but d'aligner nos séquences trouvées pour effectuer un modèle par homologie.

En parallèle, nous avons utilisés des méthodes de détection de partie transmembranaires et de prédiction de la structure secondaire. Ces outils ont été OPM, JPred4, Phyre2, et PsiPred, Lomets et Muster.

N'ayant pas obtenu de résultat assez satisfaisant pour construire un modèle par homologie, dû à des alignements de mauvaise qualité, ou de structure PDB trop petite par rapport à la taille de notre séquence, nous avons décidés de construire des modèles par 'threading' et de sélectionner les modèles les plus vraisemblables selon la littérature biologique existante. Nous avons comparé les modèles issues des modalisations par I-Tasser, Phyre2, SwissModel, et RaptorX.

Puis parmi les différents modèles obtenus, nous avons sélectionné les modèles en adéquation avec les résultats expérimentales. Ces derniers modèles ont subits un raffinements de leurs chaînes latérales grâce aux logiciels SIDEPro.

Nous avons ensuite cherché à évaluer notre modèle grâce à PROQM et Verify3D.

Enfin, pour expliquer les principaux mouvements de la protéine nous avons analyser les modes normaux de cette protéine grâce au paquet 'Bio3D' sous R.

Résultats

Alignements

La séquence protéique de A1A5C7 est connus et 'reviewed'. C'est une séquence de 686 acides aminées. Le 'BLAST' sur UniProt de cette séquence avec la base de donnée 'PDB' nous apporte 105 séquences 'reviewed' avec informations sur la structure. Les principaux résultats ressortant de ce BLAST sont les 'GTRx' qui sont des 'Facilitated glucose transporter' et des 'RNA binding protein' (cf figure 1).

P22732	GTR5_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 1.7e-10 Score: 156 Ident.: 23.1%
Q6XP3	GTR7_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 7.1e-10 Score: 151 Ident.: 26.4%
Q496J9	SV2C_HUMAN - Synaptic vesicle glycoprotein 2C Homo sapiens (Human) - View alignment	E-value: 1.4e-3 Score: 99 Ident.: 22.7%
Q86XN8	MEX3D_HUMAN - RNA-binding protein MEX3D Homo sapiens (Human) - View alignment	E-value: 9.4e-3 Score: 92 Ident.: 36.0%
Q86XN8-2	MEX3D_HUMAN - Isoform 2 of RNA-binding protein ME... Homo sapiens (Human) - View alignment	E-value: 9.5e-3 Score: 92 Ident.: 36.0%
Q86XN8-3	MEX3D_HUMAN - Isoform 3 of RNA-binding protein ME... Homo sapiens (Human) - View alignment	E-value: 9.5e-3 Score: 92 Ident.: 36.0%
P11166	GTR1_HUMAN - Solute carrier family 2, facilitate... Homo sapiens (Human) - View alignment	E-value: 4.5e-2 Score: 86 Ident.: 22.1%

Figure 1 : Illustration des 7 premiers alignements en fonction de leur scores.

On remarque dès à présent que les ‘RNA binding Protein’ (aussi appelé RBP) ont une meilleure identité que les ‘GTRx’, mais inversement, les ‘GTRx’ possèdent des structures plus importantes que les ‘RBP’. Nous savons que la structure est bien plus conservée au cours de l’évolution que les séquences. C’est pour cela que nous avons choisi d’aligner les meilleures séquences en termes de score, plutôt que d’aligner les meilleures structures ayant la plus forte identité de séquence.

Nous avons donc aligné notre séquence de référence avec les 7 séquences ayant le meilleur score selon BLAST, grâce à ClustalW. On peut tout de suite voir que les protéines RBP de la famille CNOT (CCR4 Not Transcription) sont retrouvées sous 3 isoformes pour cette alignement. Cependant, les 3 isoformes sont assez significatives car correspondent à 3 épissages différents pour cette protéine et ne couvrent pas le même alignement. De plus, les structures PDB associées sont légèrement différentes en termes de recouvrement.

Nous considérons que nos séquences sont homologues, si l’identité entre ces deux séquences est supérieure à 30 %. De plus, nous choisirons un modèle par homologie si la structure PDB de l’homologue recouvre assez bien notre séquence de référence.

Les séquences ici, ne nous permettent pas de faire de l’homologie. En effet, leur identité est inférieure à 30 %, et pour les quelques alignements dépassant ce seuil, la structure PDB associée ne recouvre pas assez la séquence de référence.

On peut cependant identifier des zones assez bien conservées à travers les différents alignements. Ils correspondent plutôt à des parties de la structure aux propriétés hydrophobes.

A1A5C7	S22AN_HUMAN	1	-----	0
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	1	MADKRKLQGEIDRCLKKVSEGVQFEDIWQKLHNAANANQKEKYEADLKKEIKKLQRLRD	60
075175-3	CNOT3_HUMAN	1	-----	0
075175-2	CNOT3_HUMAN	1	MADKRKLQGEIDRCLKKVSEGVQFEDIWQKLHNAANANQKEKYEADLKKEIKKLQRLRD	60
A1A5C7	S22AN_HUMAN	1	-----	0
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	61	QIKTWVASNEIKDKRQLIDNRKLIETQMERFKVVERETKTKAYSKEGLGLAQKVDPAQKE	120
075175-3	CNOT3_HUMAN	1	-----	0
075175-2	CNOT3_HUMAN	61	QIKTWVASNEIKDKRQLIDNRKLIETQMERFKVVERETKTKAYSKEGLGLAQKVDPAQKE	120
A1A5C7	S22AN_HUMAN	1	-----	0
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	121	KEEVGQWLTNTIDTLNMQVDQFESEVESLSVQTRKKKGDKDKQDRIEGLKRHIKHYHV	180
075175-3	CNOT3_HUMAN	1	-----	0
075175-2	CNOT3_HUMAN	121	KEEVGQWLTNTIDTLNMQVDQFESEVESLSVQTRKKKGDKDKQDRIEGLKRHIKHYHV	180
A1A5C7	S22AN_HUMAN	1	-----	0
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	181	RMLETILRMLDNDLSILVDAIRKIKDDVEYYVDSSQDPDFEENEFLYDLDLEDIPQALVA	240
075175-3	CNOT3_HUMAN	1	-MLETILRMLDNDLSILVDAIRKIKDDVEYYVDSSQDPDFEENEFLYDLDLEDIPQALVA	59
075175-2	CNOT3_HUMAN	181	RMLETILRMLDNDLSILVDAIRKIKDDVEYYVDSSQDPDFEENEFLYDLDLEDIPQALVA	240
A1A5C7	S22AN_HUMAN	1	-----MAIDRRREAA-----G	11
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	241	TSPPSHSHMEDEIFNQSSSTPTSTTSSSPIPPSPANCITENS EDDKKRGRSTDSEVSQSP	300
075175-3	CNOT3_HUMAN	60	TSPPSHSHMEDEIFNQSSSTPTSTTSSSPIPPSPANCITENS EDDKKRGRSTDSEVSQSP	119
075175-2	CNOT3_HUMAN	241	TSPPSHSHMEDEIFNQSSSTPTSTTSSSPIPPSPANCITENS EDDKKRGRSTDSEVSQSP	300
A1A5C7	S22AN_HUMAN	12	GGPGRQPAPAE-----ENGSLPPGDAASAPLGGRAGPGG-GAEIQPLPLHPGGGPH	63
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----	0
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	301	AKNGSKPVHSNQHPQSPAVPPTYP SGPPPAASALSTTPGNNGVPAAPPSALGPKASPA	360
075175-3	CNOT3_HUMAN	120	AKNGSKPVHSNQHPQSPAVPPTYP SGPPPAASALSTTPGNNGVPAAPPSALGPKASPA	179
075175-2	CNOT3_HUMAN	301	AKNGSKPVHSNQHPQSPAVPPTYP SGPPPAASALSTTPGNNGVPAAPPSALGPKASPA	360
A1A5C7	S22AN_HUMAN	64	PSCCSAAAAPSLLLLDYDGSVLPFLGGLGGGYQKTLVLLTWIPA-----	107
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	1	-----MEDSYKDRTSLMKGAKDIAREVKKQTVKKVNQA	33
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	361	PSHNSGTPAPYA-----QAVAPPA-----	379
075175-3	CNOT3_HUMAN	180	PSHNSGTPAPYA-----QAVAPPA-----	198
075175-2	CNOT3_HUMAN	361	PSHNSGTPAPYA-----QAVAPPA-----	379
A1A5C7	S22AN_HUMAN	108	-----LFIGFSQFSDSFLLDQPNFWCRGAGK---GTELAGV-T--TTGRGGD----M	149
P22732	GTR5_HUMAN	1	-----	0
Q6PXP3	GTR7_HUMAN	1	-----	0
Q496J9	SV2C_HUMAN	34	VDRAQDEYTORQSYSRFQDE--EDDDDYYPAGETYNGEANDDEGS-SEATEGHDEDEIYE	90
P11166	GTR1_HUMAN	1	-----	0
075175	CNOT3_HUMAN	380	-----PSGPSTTQ----PRPPSVQPSGGGGGSGGGGSSSSSNSSAGGGAGKQNGA	426
075175-3	CNOT3_HUMAN	199	-----PSGPSTTQ----PRPPSVQPSGGGGGSGGGGSSSSSNSSAGGGAGKQNGA	245
075175-2	CNOT3_HUMAN	380	-----PSGPSTTQ----PRPPSVQPSGGGGGSGGGGSSSSSNSSAGGGAGKQNGA	426

A1A5C7	S22AN_HUMAN	150	GNWTS LP TPFAT-----APWEAAGNRSN SS GADGGDTPPLP-SPPDKG-----	192
P22732	GTR5_HUMAN	1	-----MEQ0DQS-MKEGR LT LV LA L	19
Q6PXP3	GTR7_HUMAN	1	-----MENKEAGTPPPIP-SREGR LQ PTLL L	25
Q496J9	SV2C_HUMAN	91	GEYQGI PS MNQAK-----DSIVSVGQPKGDEYKDRRELESERRADEEELAQQYE-	139
P11166	GTR1_HUMAN	1	-----ME-PSSK-KLTGR L -----ML	14
075175	CNOT3_HUMAN	427	TSYSSV VD SPA E VALSSSGGNNASSQALGPPSGPHNP-----PPSTSK EP SAAAPT G AG	481
075175-3	CNOT3_HUMAN	246	TSYSSV VD SPA E VALSSSGGNNASSQALGPPSGPHNP-----PPSTSK EP SAAAPT G AG	300
075175-2	CNOT3_HUMAN	427	TSYSSV VD SPA E VALSSSGGNNASSQALGPPSGPHNP-----PPSTSK EP SAAAPT G AG	481
A1A5C7	S22AN_HUMAN	193	DNASNCDCRAW DY GIR---AGLVQNVVSK W DLV-----CDNAWKVH	230
P22732	GTR5_HUMAN	20	ATLIAAFGSSFOYGYN---VAAVNSPALLMQOFYNETY Y GRTGEF-MEDFPLTLLWSVT	74
Q6PXP3	GTR7_HUMAN	26	ATLSAAGFSAFOYGYN---LSVNTPHKVF K SFYNETYFERHATF-MDGKLM LL WSCT	80
Q496J9	SV2C_HUMAN	140	LIIQECG H GRFQWALFFVLGMALMADGVEVF---VVG F VLPSA ET DLCPNSGSGWLGS	195
P11166	GTR1_HUMAN	15	A-VGGAVLGSLOFYGN---TGVINAPQK VI EEFY NQ TWVHRYGES-ILPTTLTLLWSLS	68
075175	CNOT3_HUMAN	482	GVAP--GSGNNSGGPSLLV-PLPVNPPSSPT-----	509
075175-3	CNOT3_HUMAN	301	GVAP--GSGNNSGGPSLLV-PLPVNPPSSPT-----	328
075175-2	CNOT3_HUMAN	482	GVAP--GSGNNSGGPSLLV-PLPVNPPSSPT-----	509
A1A5C7	S22AN_HUMAN	231	IAKF S LLVGLIFGYLITGCIADW V GRRPVLLFSIIFILIFGLTVALSVNV TM F---STLR	287
P22732	GTR5_HUMAN	75	VSMFP--FGFGISLLVGPLVNK F GRKGALLFNNIFSIVPAILMGCSR V ATS F ELIIS R	132
Q6PXP3	GTR7_HUMAN	81	VSMFP--LGGLLGSLLVGLLV D SCGRK GT LLINNIFAIIPAILMGVSK V AKAFELIV FS R	138
Q496J9	SV2C_HUMAN	196	I---VYLGMVMVGAFFWGG L ADKVG R KQSL L ICMSVNGFFA---FLSS F VQGYGFFL F CR	248
P11166	GTR1_HUMAN	69	VAIFS--VGGMIGSFSVGLFVN R FGR R NSMLM N LLAFVSAVLMGFSK L GK S FEM L ILGR	126
075175	CNOT3_HUMAN	510	-PSFS--DAKAAGALLNGPPQFS--TAPEIK-----APE--PLSSLK S MA-----ER	549
075175-3	CNOT3_HUMAN	329	-PSFS--DAKAAGALLNGPPQFS--TAPEIK-----APE--PLSSLK S MA-----ER	368
075175-2	CNOT3_HUMAN	510	-PSFS--DAKAAGALLNGPPQFS--TAPEIK-----APE--PLSSLK S MA-----ER	549
A1A5C7	S22AN_HUMAN	288	FFEGFCLAGIILTLYALRIELCP P GKRFMITMVASFVA-----MAGQF-----LM--	332
P22732	GTR5_HUMAN	133	LLVGICAGVSSNVVPMYLGELAPKNL R GALGVV P QLFITV G ILVAQIF-----GL--	182
Q6PXP3	GTR7_HUMAN	139	VVLGVCAGISY S ALPMYLGELAPKNL R GMVGTMT E VFVIVGVFLAQIF-----SL--	188
Q496J9	SV2C_HUMAN	249	LLSGFGIGGAIPTVF S YFAEVLAREK R GEHL S WLCMFWMIGGIYASAMAWAI I PHYGWSF	308
P11166	GTR1_HUMAN	127	FIIGVYCGLT T TGFVPMYVGEVSP T ALRGALGTLHQLGIVV G ILIAQVF-----GL--	176
075175	CNOT3_HUMAN	550	AAISSGIEDPVPTLH L TERD-----IILS S TS-----A	577
075175-3	CNOT3_HUMAN	369	AAISSGIEDPVPTLH L TERD-----IILS S TS-----A	396
075175-2	CNOT3_HUMAN	550	AAISSGIEDPVPTLH L TERD-----IILS S TS-----A	577
A1A5C7	S22AN_HUMAN	333	-PGLAALCRDQV L QALII C PFLLMLLYWSIFPESL R WL M ATQOFESAK R LILH F TQ K NR	391
P22732	GTR5_HUMAN	183	-RNLLANVDGW P ILLGLTGVPAALQLLLLPFFPESPRYLLIQ K KDEAAAK L QTLR---	238
Q6PXP3	GTR7_HUMAN	189	-QAILGNPAGWPVLLAL T GV P ALLQLLTLPFFPESPRYSLIQ K GDEATARQAL R RLR---	244
Q496J9	SV2C_HUMAN	309	SMGSAYQFHS R WV F IVICALPCVSSV V ALT F MPESP R FLLEV G KHDE-AWMIL K L I HD T N	367
P11166	GTR1_HUMAN	177	-DSIMGNKDLW P LLLSI F IPALLQ C IVLP F CPESP R FL L INRNEENRAK S VL K K L R---	232
075175	CNOT3_HUMAN	578	PPAS----AQPLQLSEVNIPLSLGVCPLG P VPLTKE-QLYQQAMEEA A WHM P HP S DS E	632
075175-3	CNOT3_HUMAN	397	PPAS----AQPLQLSEVNIPLSLGVCPLG P VPLTKE-QLYQQAMEEA A WHM P HP S DS E	451
075175-2	CNOT3_HUMAN	578	PPAS----AQPLQLSEVNIPLSLGVCPLG P VPLR-----	609
A1A5C7	S22AN_HUMAN	392	MNPEGDIKGV-----PEL--EKELSR R PKKVCIV-K-----VVGTRNL--	427
P22732	GTR5_HUMAN	239	--GWDSVDREV-----AEIROE E AKAGFISVL-K-----LFRMRSL--	274
Q6PXP3	GTR7_HUMAN	245	--GHTDMEAL-----EDMRAE A RAEAGHLSVL-H-----LCAL R SL--	280
Q496J9	SV2C_HUMAN	368	MRARGOPEKVFTVN K IKTPKQIDELIEIESDTGTWY R RCFV R TEL Y GIWLT F MR C FN Y	427
P11166	GTR1_HUMAN	233	--GTADVTHDL-----QEMKEESRQMMREKKVTIL-E-----LFRSPAY--	268
075175	CNOT3_HUMAN	633	--RIRQYLPRNPCTPPYHHQMP P PHSDTVEFYQ R LSTETL-----FFIFYYLE	679
075175-3	CNOT3_HUMAN	452	--RIRQYLPRNPCTPPYHHQMP P PHSDTVEFYQ R LSTETL-----FFIFYYLE	498
075175-2	CNOT3_HUMAN	610	-----	609
A1A5C7	S22AN_HUMAN	428	--WKNIV---VLCVNSLT-----GYGIH H CFAR S MMGHEV K VPLLENFYA-----	467
P22732	GTR5_HUMAN	275	--RWOLLSII V LMGGQQLS-----GVNAIYY A DOIYLS-AGVPEEH--V-----	314
Q6PXP3	GTR7_HUMAN	281	--RWOLLSII V LMAGQQLS-----GINAINYYAD T IYTS-AGVEAAH--S-----	320
Q496J9	SV2C_HUMAN	428	PVRDNTIKLTIW F TLSFG-----YYGLSV W FPDVI-----KPLQ S DEYALLTR N VE	474
P11166	GTR1_HUMAN	269	--RQPILIAVVLQ L SQQLS-----GINAVFY Y STSI F EK-AGVQO-----P-----	306
075175	CNOT3_HUMAN	680	G-----TKAQYL--AAKALK K QSW R FHTKYMMW F ORHEEP-----K T IT D EF E	720
075175-3	CNOT3_HUMAN	499	P-----RDVRTAW V DEAGPLR H TSAPLPAH W LTRG P SP-----ALLQ O P S	541
075175-2	CNOT3_HUMAN	610	-----	609
A1A5C7	S22AN_HUMAN	468	-----DY-----	469
P22732	GTR5_HUMAN	315	-----QY-----	316
Q6PXP3	GTR7_HUMAN	321	-----QY-----	322
Q496J9	SV2C_HUMAN	475	RDKYANFTINF T ME N QIHTGMEYDNG R FIGV K FKSV T FKDSV F KSC T FEDV T SVNTY F KN	534
P11166	GTR1_HUMAN	307	-----VY-----	308
075175	CNOT3_HUMAN	721	QGTYY I FDY E K--WGQ R K-----KEG--FTFEY-----	744
075175-3	CNOT3_HUMAN	542	RPTPLC S SAHS--QASLE--TTGE-LSSPLPT-----	568
075175-2	CNOT3_HUMAN	610	-----	609

A1A5C7	S22AN_HUMAN	470	-----YTTAS-----	-----IAL	477
P22732	GTR5_HUMAN	317	-----VTAGT-----	-----G-AVNV	326
Q6PXP3	GTR7_HUMAN	323	-----VTVGS-----	-----G-VVNI	332
Q496J9	SV2C_HUMAN	535	CTFIDTVFDNTDFEPYKFIDSEFNKCSFFHNKTGCQITFDDDYSAWYIYFVNFLGTLAVL	-----G-IVNT	594
P11166	GTR1_HUMAN	309	-----ATIGS-----	-----G-IVNT	318
075175	CNOT3_HUMAN	745	-----RYLEDRLQ-----	-----	753
075175-3	CNOT3_HUMAN	569	-----PLL-----	-----	571
075175-2	CNOT3_HUMAN	610	-----	-----	609
A1A5C7	S22AN_HUMAN	478	VSCLAMCVVVRFLGRRGGLLLFMILTALASLLQLGLLNLIGKYSQHPDSGMSDSVKDKFS	-----	537
P22732	GTR5_HUMAN	327	VMTFCAVFVVELLGRRLLLLLGFSICLIACCVLTAALALQDTVSWMP-----	-----YI	375
Q6PXP3	GTR7_HUMAN	333	VMTITSAVLVERLGRRLHLLAGYGICGSACLVLTVVLLFQNRVPELS-----	-----YL	381
Q496J9	SV2C_HUMAN	595	PGNIVSALLMDRIGRLTMLGGSMVLSGISCFLL-----	-----WFGT-----SESMML	637
P11166	GTR1_HUMAN	319	AFTVVSLFVVERAGRRTLHLIGLAGMAGCAILMTIALALLEQLPWMS-----	-----YL	367
075175	CNOT3_HUMAN	754	-----	-----	753
075175-3	CNOT3_HUMAN	572	-----	-----	571
075175-2	CNOT3_HUMAN	610	-----	-----	609
A1A5C7	S22AN_HUMAN	538	IAFSIVGMFASHAVGSL--VF-----FCAEITPTVIRCG-----GLGLVLASAGFGMLTA	-----	586
P22732	GTR5_HUMAN	376	SIVCVISYVIGHALGPSPIPALLLTEIFLQSSRPSAFMVG-----GSVHWLSNFTVGLIFP	-----	431
Q6PXP3	GTR7_HUMAN	382	GIICVFAYIAGHSIGSPVPSVVRTEIFLQSSRRAAFMVD-----GAVHWLTFNFIIGFLFP	-----	437
Q496J9	SV2C_HUMAN	638	GMLCLYNGLTISAWNLS--DVVTVELYPTDRRATGFGFLNALCKAAAVLGNLIFGSLVS	-----	694
P11166	GTR1_HUMAN	368	SIVAIFGFVAFFEVGPGPIPWFIVAELFSQGPRAAIAVA-----GFSNWTSNFIVGMCFQ	-----	423
075175	CNOT3_HUMAN	754	-----	-----	753
075175-3	CNOT3_HUMAN	572	-----	-----	571
075175-2	CNOT3_HUMAN	610	-----	-----	609
A1A5C7	S22AN_HUMAN	587	PIIELNHGKGYFLHHIIFACCTLICICILLPLESRDQNLPENISNGEHYTRQPLL-PHK	-----	645
P22732	GTR5_HUMAN	432	FIOEG---LGPYSFIVFAVICLLTTIYIFLIVPETKAKTFIEINQIFTKMNKVSEVYPEK	-----	488
Q6PXP3	GTR7_HUMAN	438	SIOEA---IGAYSFIIFAGICLLTAIYIYVIPETKGTVEINRIFAKNRNRKLP EEKE	-----	494
Q496J9	SV2C_HUMAN	695	ITKSI---PILLASTVLVCGGLVGLCLPDTRTQVLM-----	-----	727
P11166	GTR1_HUMAN	424	YVEQL---CGPYVFIIFTVLLVLFIFTYFKVPETKGRTFDEIASGFRQGGASQSDKTPE	-----	480
075175	CNOT3_HUMAN	754	-----	-----	753
075175-3	CNOT3_HUMAN	572	-----	-----	571
075175-2	CNOT3_HUMAN	610	-----	-----	609
A1A5C7	S22AN_HUMAN	646	KGEQPLL--LTNAELKDYSGLHDAAGDTLPEGATANGMKAM	-----	686
P22732	GTR5_HUMAN	489	EELKELPPVTSEQ-----	-----	501
Q6PXP3	GTR7_HUMAN	495	ETIDAGPPTASPAKETSF-----	-----	512
Q496J9	SV2C_HUMAN	728	-----	-----	727
P11166	GTR1_HUMAN	481	ELFHPL---GADSQV-----	-----	492
075175	CNOT3_HUMAN	754	-----	-----	753
075175-3	CNOT3_HUMAN	572	-----	-----	571
075175-2	CNOT3_HUMAN	610	-----	-----	609

Figure 2 : alignement entre la séquence de référence (A1A5C7), des séquences de GRTx et de CNOT (Transcription complex)

Détection structure secondaire et partie transmembranaire

Nous savons que les protéines de la famille des SLC22, comporte 12 hélices transmembranaires et une grande boucle extracellulaire entre l'hélice transmembranaire 1 et 2 (réf : 2 et 3). Nous avons donc comparer plusieurs logiciels de prédiction de structure secondaire : Phyre2, Jpred4, Hhpred, Lomets. Tous nous indique bien 12 hélices transmembranaires, avec une boucle extracellulaire entre le 1^{er} et 2eme segment transmembranaires. Pour des soucis de concision, nous choisissons de ne représenter que les résultats obtenus par Phyre2 et Hhpred (respectivement figure : 3 a et b).

Les hélices alpha, et les parties transmembranaires sont assez bien prédit dans les mêmes localisations pour l'ensemble des logiciels. Seul quelques petits variations existent concernant, les positions de début et de fin des hélices alpha pour les différentes prédiction, par exemple, pour la première hélice alpha, Phyre2 la prédit entre les positions 94 et 120, tandis que HHPred la prédit entre les positions 93 et 127. Cela en est de même pour les différentes structures secondaires.



Figure 3 a : Prédiction de la structure secondaire par Phyre2. On observe bien 12 hélices transmembranaires avec une grande boucle entre les deux premiers segments transmembranaires.

Protein ID: sp|A1A5C7|S22AN_HUMAN Solute carrier family 22 mem



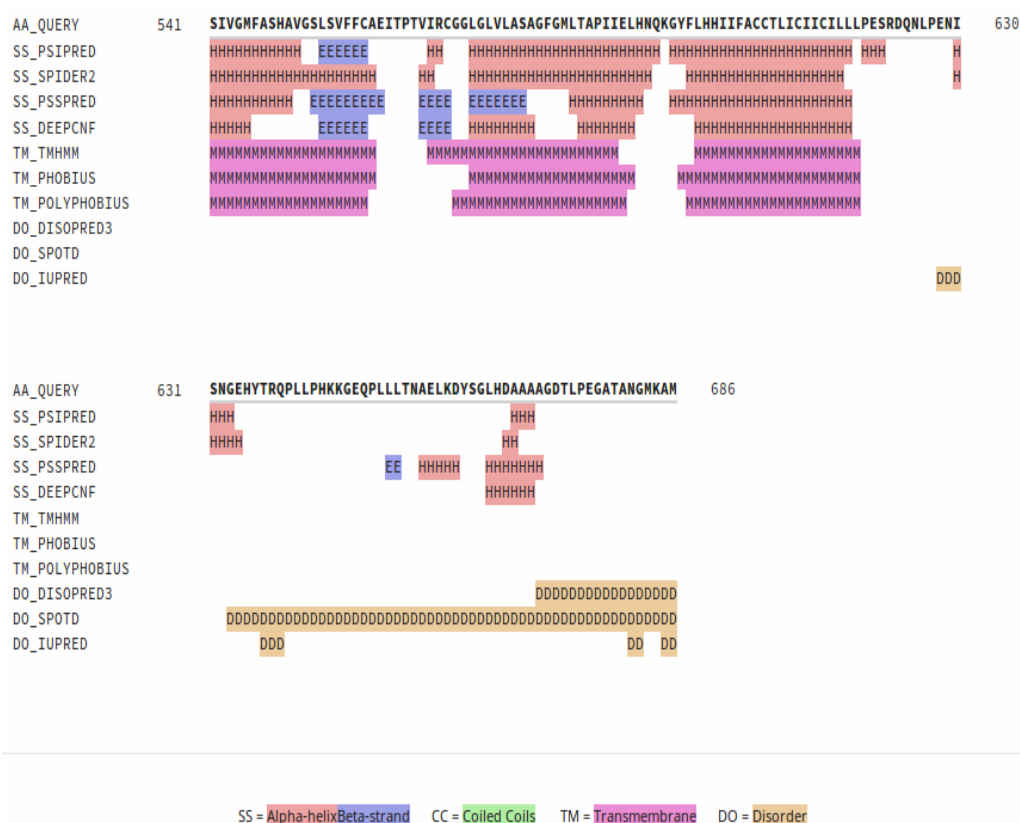


Figure 3 b : prédiction de la structure secondaire par HHPred. On observe bien 12 hélices transmembranaires avec une grande boucle entre les deux premiers segments transmembranaires.

Globalement, on peut affirmer, au vue de la redondance des résultats et des expérimentations biologique, que nos résultats sont conformes à ce qui était attendu. C'est à dire que la protéine contient bel et bien 12 segments transmembranaires en hélice alpha, et une grande boucle coté extracellulaire comprise entre les 2 premiers segments transmembranaires.

De plus, nous observons dans tout les cas, 2 autres hélices alpha non transmembranaire et situé coté intracellulaire de la membrane. La prédiction des structures secondaires par ces différentes méthodes nous a permis d'en savoir un peu plus par rapport à la littérature associé à cette protéine.

Nous avons vu lors de l'alignement, que les propriétés plutôt hydrophobes de la protéine était conservé à travers la séquence. Les hélices alpha transmembranaires, sont elles aussi plutôt hydrophobes pour interagir avec la membrane lipidique.

Nous mettons donc en évidence que les caractéristique structurale en rapport avec les propriétés physico-chimique de passage de membrane sont assez conservés.

Nous pouvons maintenant nous concentrés sur la modélisation d'un modèles par threading.

Modélisation par 'Threading'

La modélisation par threading étant coûteuse en temps de calcul et fournissant des résultats difficilement comparable, nous avons décidé de comparer différents modèles issue de plusieurs logiciels de threading. Nous choisirons le modèle respectant aux mieux ce qui a déjà pu être vérifié expérimentalement, et en accord avec nos prédictions de structure secondaires.

Notre modèle doit donc comporter 12 hélices transmembranaires, une grande boucle coté extracellulaire, et 2 hélices alpha coté intracellulaire. De plus, nous savons que le Nter et Cter de la protéine se trouvent du coté intracellulaire de la membrane. Enfin, nos modèles ne doivent pas être aberrant sur le plan de nos connaissances en biologie.

Nous avons donc utilisés les modélisations suivantes : I-Tasser, Phyre2, SwissModel et RaptorX.

A la vue des résultats préliminaires, nous choisissons de ne vous présenter que les meilleurs modèles obtenues avec SwissModel et RaptorX. En effet, les résultats obtenues avec I-Tasser ne comprenaient que 7 hélices transmembranaires, 3 ou 5 hélices extracellulaires selon les modèles, et soit aucune ou une petite boucle extracellulaire. Nous avons donc juger que les résultats n'étaient pas conforme à ce qui était attendu. Pour les modélisations de Phyre2, les résultats comprenaient 11 à 12 hélices transmembranaires, ce qui est correcte. Cependant, dans sa façon de traiter la modélisation, Phyre2 coupe les boucles de la protéine. Nous nous retrouvions donc avec uniquement une structure composée d'hélice alpha, sans aucune boucle, et donc avec beaucoup trop tronqué. La boucle extracellulaire étant une des caractéristiques de la famille des SLC22, nous avons donc décider de ne pas retenir ces modèles. De plus, l'absence trop nombreuse des boucles, nous aurait poser des problèmes pour analyser la structure par la suite.

Nous allons donc étudier les résultats de SwissModel (figure 4a) et RaptorX (figure 5a).

SwissModel

Le modèle obtenu diffère un peu par rapport à nos espérances. Une boucle est bien présente coté extracellulaire, mais elle n'est pas aussi grande que prévu. De plus, on observe une boucle coté intracellulaire et non deux hélices alpha comme prédit avec les outils de prédictions de structure secondaire. Cependant, on retrouve bien nos hélices alpha transmembranaires, et le Cter et Nter coté intracellulaire. Mais, on ne constate que 10 hélices transmembranaires au lieu de 12 attendues.

Le template retenu pour faire ce modèle est celui de 4zp0.1.A. Ce PDB est celui d'un transporter multidrug d'E.Coli. Cela correspond hypothétiquement à notre protéine. Cependant, la taille de la séquence de cette structure est de 392 résidus. Soit presque la moitié de la taille de notre séquence référence (686 aa). Cela est visible par la présence de nombreux gap lors de l'alignement de ces deux séquences (figure 4b).

Nous décidons de ne pas retenir ce modèle qui nous ne permet d'analyser efficacement la structure de notre séquence initiales.

RaptorX

Le modèle obtenu reflète mieux ce qui était attendu. En effet, nous remarquons la présence de 12 hélices transmembranaires, d'une grande boucle extracellulaire, et la présence coté intracellulaire de la partie Nter et Cter de la protéine.

Nous remarquons tout de même la présence coté intracellulaire de 3 hélices alpha au lieu de 2 prédits.

Le template retenu est 4GBY, qui est celui d'un MFS (Major Facilitator superfamily), transporteur en symport proton:xylose chez E.Coli. Il appartient à la même superfamille que notre protéine.

La séquence du template (491 aa) est plus petite que notre séquence référence (686 aa), mais quand même plus grande que la séquence du modèle SwissModel.

En analysant l'alignement entre les deux séquences (figure 5b), on observe des gap biens moins nombreux que le précédent modèle. De plus, la grande majorité de ces gaps sont situés dans la partie Nter de la protéine : notre protéine plus grande, déborde très largement de ce coté de la séquence. Cela a moins d'impact sur la structure sur une partie plus au 'cœur' de la structure du template.

On peut donc supposer que cet alignement expliquant une grande partie de l'écart de taille sur la partie Nter est plus vraisemblable qu'un alignement démontrant l'écart de taille en plein cœur de la protéine.

Enfin, quand on s'attarde sur la qualité de l'alignement, on observe les meilleurs scores, sur les parties transmembranaires de la structure.

On observe une bonne identité d'alignement même sur le segment 1 de la partie transmembranaire de notre séquence par rapport à la séquence template. Le segment 1 est importante car elle est situé coté Nter. Or on retrouve un bon alignement avec le template.

Chez le transport couplé Proton:Xylose d'E.Coli, on ne retrouve pas de grande boucle entre les deux premiers segments transmembranaire (7).

Or, à la vue de l'alignement, les deux premiers segments sont correctement alignés, et on peut expliquer une partie de l'écart de taille si importante entre les deux séquences, par la présence pour la famille des SLC22, d'une grande boucle coté extracellulaire entre les deux premiers segments transmembranaires.

Pour terminer, le modèle ne contredit pas ce qui déjà été précédemment démontré par expérimentation, et conforte les précédentes prédictions de structure secondaires.

On peut donc, par absence de littérature ou de données nous apportons des précisions contre ce modèle, sélectionné ce modèle par threading de RaptorX, afin d'étudier les mouvements de ce transporteur.

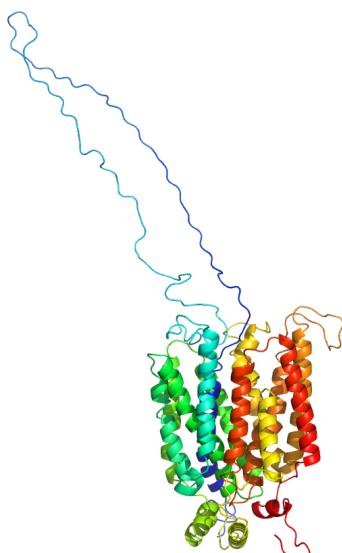


Figure 5a : Représentation 3D du modèle par RaptorX (image obtenue par PyMol)

```

T-COFFEE, Version 11.00.d625267 (2016-01-11 15:25:41 - Revision d625267 - Build 507)
Cedric Notredame
SCORE=64
*
* BAD AVG GOOD
sp|A1A5C7|S22AN : 64
4GBY_A|PDBID|CH : 64
cons : 6
MAIDRRREAAGGGPGRQPAPAEENGSLPPGDAAASAPLGGRAGPGGGAEIQPLPLHPGGGPH
MNTQYNSSYI
cons * : . .
PSCCSAAAPSLLLLDYDGSVLPFLGGLGGGYOKTLVLLTWIPALFIGFSQFSDFSLLDQPNF
FSITLVATLGGLLFGYDTAVIS
cons : . : : . **** * : : :
WCRGAGKGTLAGVTTTGRGDMGNWNTSLPTTPFATAPWEAAGNRSNSGADGGDTPPLPSP
GTVESLNTV FVAPO
cons * . * * : : *
DKGDNASNCDCRANDYGRAGLVQNVVSKWDLVCDNAWKVHIKFSLLVGLIFGYLITGCIAD
NLSESAANS LLGFCVASALIGCIIGGALGGYCSN
cons : . : * : * : : : * : * : * : * :
WVGRRPVLLFSIIFILIFGLTV ALNVNMTFSLRFFEGFC LA
RFGRDLSLKIAAVLFFISGVGSAPPELGFTSINPONTVPVYLAGYYPEFVIYRIIGGIVGLA
cons * * * : : : : * : * : * : * : *
GIITLYALRIELCPPGKRMITMVASFVAMAGFLMPGL AA-L-CRDWDVLOA
SMLSPMYI AELAPAHIRGKLVSNQFAIFGQLLVYCVNYFIARSGDASWLNLDGWRMF
cons : : : * * : * : : * : * : * : * : *
LITCFLLMLLYWSIFPESLRWLMTQOFESAKRLILHFTOKNRNPEGDIKGVIPLEKELS
SECIPALLFLMLLYTVPESPRWLMRSGKQEQAEGLRK IMGNTLATQ AVQEIKHS
cons * * : * : * * * : : * : : : * : : : * : : *
RRPKKVCIVKVVGRNLWKNIVVLCVNSLTGYGIHCFAR SHMGHEVKVPLENFYADY
HGRKTGGRLLMFGVGVIVIGVMSLIFQQFVGINVVLYAPEVFKTLGASTDIALLOTIIVGV
cons : * : : * : : : : * : : * : : * : :
TTASAL VSCLAMCVVRFLLRRGGLLLFMILTALASLLDGLLNL IGKYSOHPDSGMSDS
INLTFTVLAIMTVDKF GRKP LQIIGALG MAIGNFSLGTAFYTQAPG
cons * : * : * : * : * : * : * : * : * : *
VKDKFSIAFSIVGMFASHAVGSLSVFFCAEITPTVIRCGGLGLVLASAGFG MLTAPIE
IVALLSMLFYV AAFA MSWGPVCWVLLSEIFPNAIRGKALAIATAAQLANYFVSWTFPMMD
cons : : * : * : * : * : : : * : * : * : * : *
LHNOKGYFLHHIIFACCTLIC TICILLPESRDONLPENISNGEHYTRPLPH
KNSWLVAHFHNGFS YWYIGCMGLAALFMWKFVPETKGTLEELEALWEPETK
cons : * : * : * : * : : : * : : * : * :
KKCEOPLLLTNAELKDYSGLHDAAGDTLPEGATANGMKAM
KTQTAT
cons * : * :

```

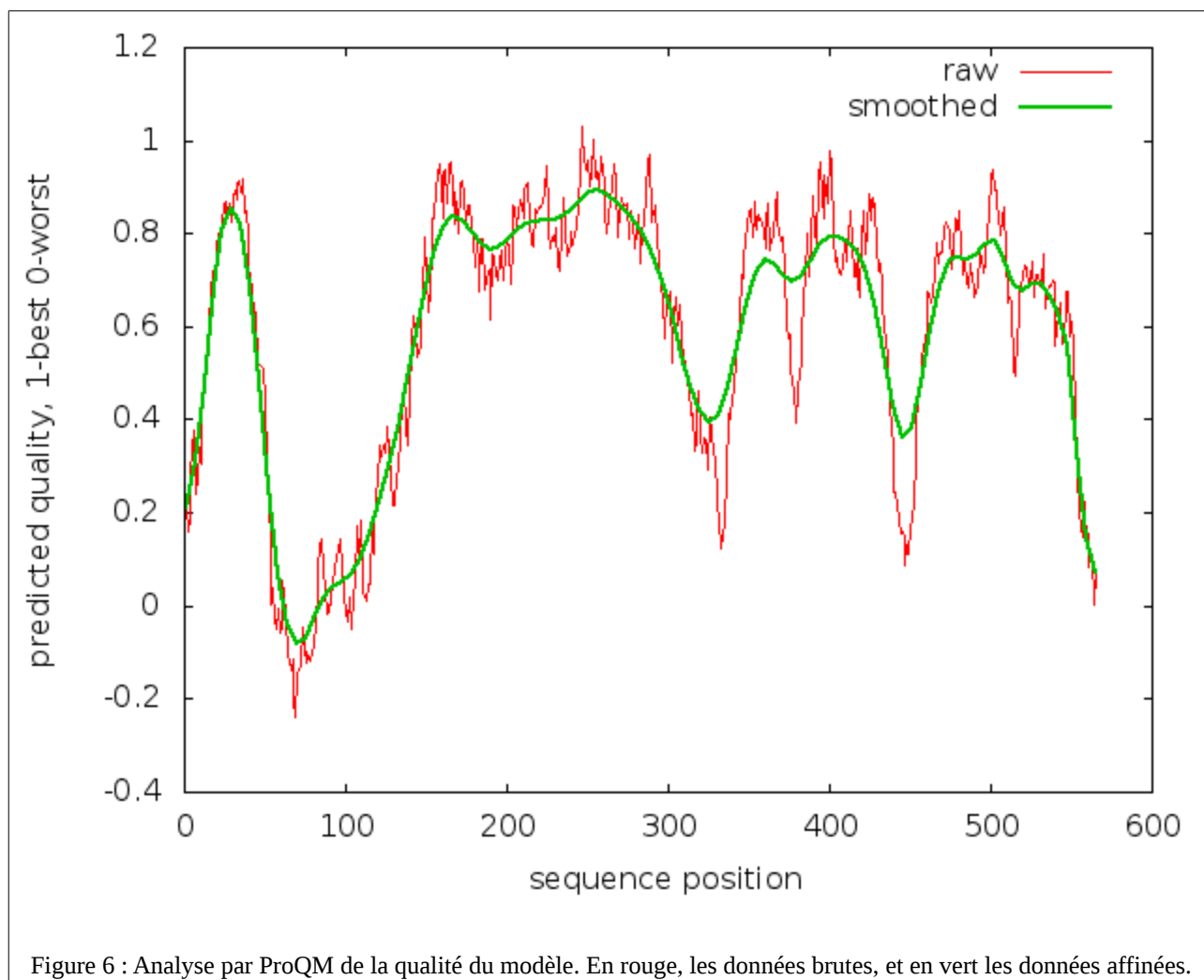
Figure 5b : alignement entre la séquence du template du modèle et notre séquence

Qualité du modèle

Nous avons soumis notre modèle à un outil de raffinement des chaînes latérales et des boucles avec SIDEPro et Modeller.

Malencontreusement, nous n'avons pas pu obtenir les résultats avant le délai imparti... Cela implique que nous ne pouvons non plus faire de dynamique moléculaire de cette protéine. Cependant, dans la suite nous ferons une études des modes normaux sur les carbones alpha de cette protéine.

Nous avons donc estimé la qualité du modèle sur les carbones alpha de notre structure, afin de ne pas tenir compte des chaînes latérales par ProQM (figure 6) et Verify3D (figure7).



ProQM est un outils d'évaluation de structure grâce à de l'apprentissage de modèle. Sa spécificité est de mieux évaluer les protéines membranaires par rapport à d'autres outils d'évaluation. Il nous donne un score compris entre 0 (mauvaise qualité) et 1 (meilleur) globale, et un score par position.

La qualité globale est de 0.589, ce qui le classe en 'moyen-bon'. C'est à l'analyse position par position qui est le plus important. En effet, en figure6, on nous présente les données bruts et lissés.

On observe des scores entre 0.8 et 0.9 (ce qui est bon), pour les carbones alpha correspondant aux parties transmembranaires du transporteur obtenue par RaptorX.

Mais, le score de la grande boucle extracellulaire compris entre les deux premiers segments est très mauvais (proche 0, avec des valeurs négatives). Quand aux autres boucles de la protéine, leurs scores sont compris entre 0.4 et 0.5 ce qui est plutôt moyen.

Les résultats de l'analyse par Verify3D sont obtenues en figure 7, ci-dessous.

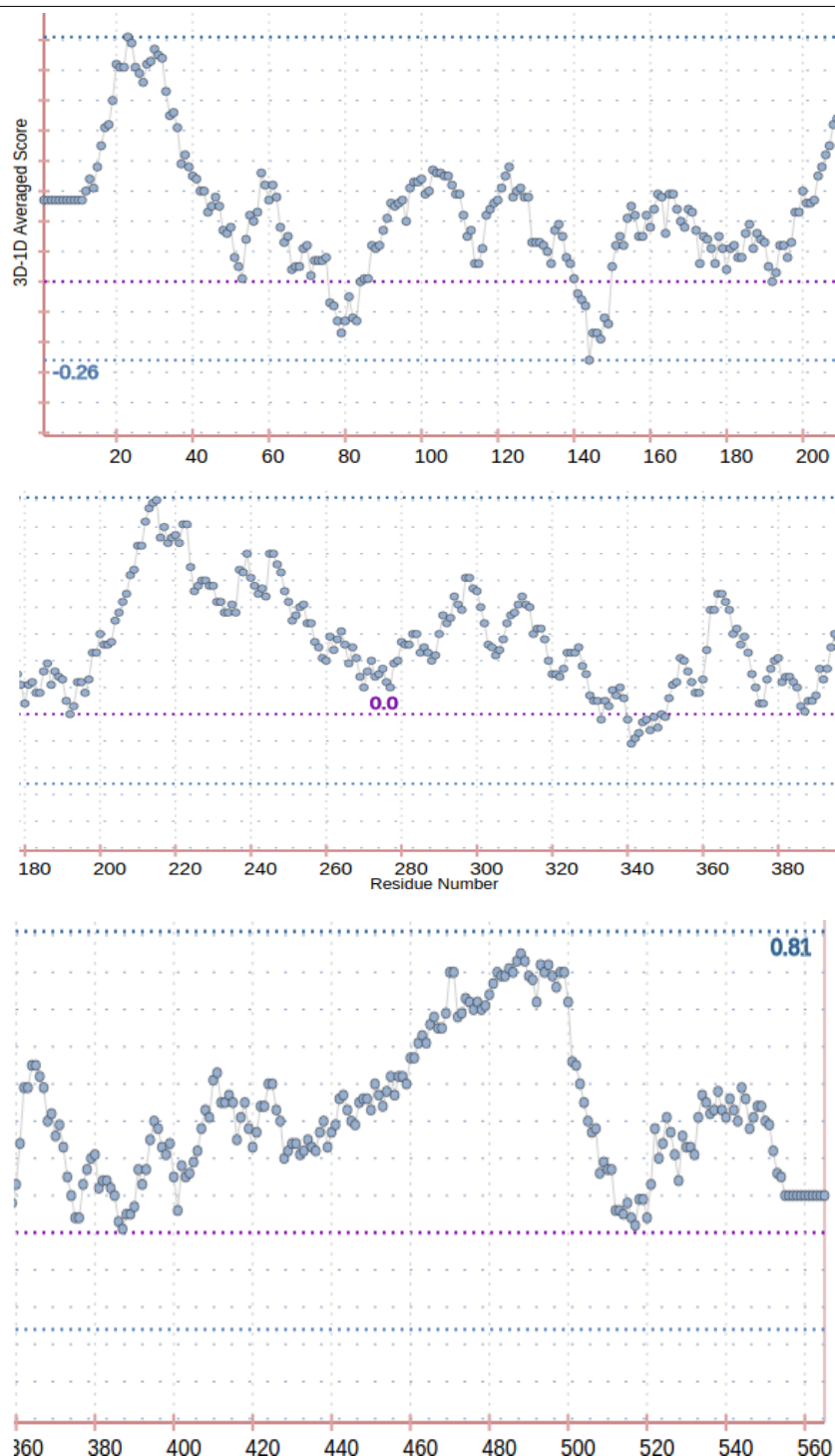


Figure 7 : Evaluation de la structure modèle par Verify3D. Une bonne évaluation est donnée par un score proche de 1

Verify3D va évaluer notre modèle par rapport à un ensemble de protéine générale (contrairement à PROQM qui analyse sur des données plutôt de protéine membranaire). Cependant, comme ProQM, il indique un score 0 étant mauvais et 1 étant comme excellent pour toutes les positions de la structure.

On retrouve sensiblement les mêmes résultats de ProQM : des hélices bien évalué et des boucles ayant des scores proches ou en dessous de 0.

On peut donc déduire que seul les carbones alpha des parties transmembranaires de notre protéine modèle sont assez correctes.

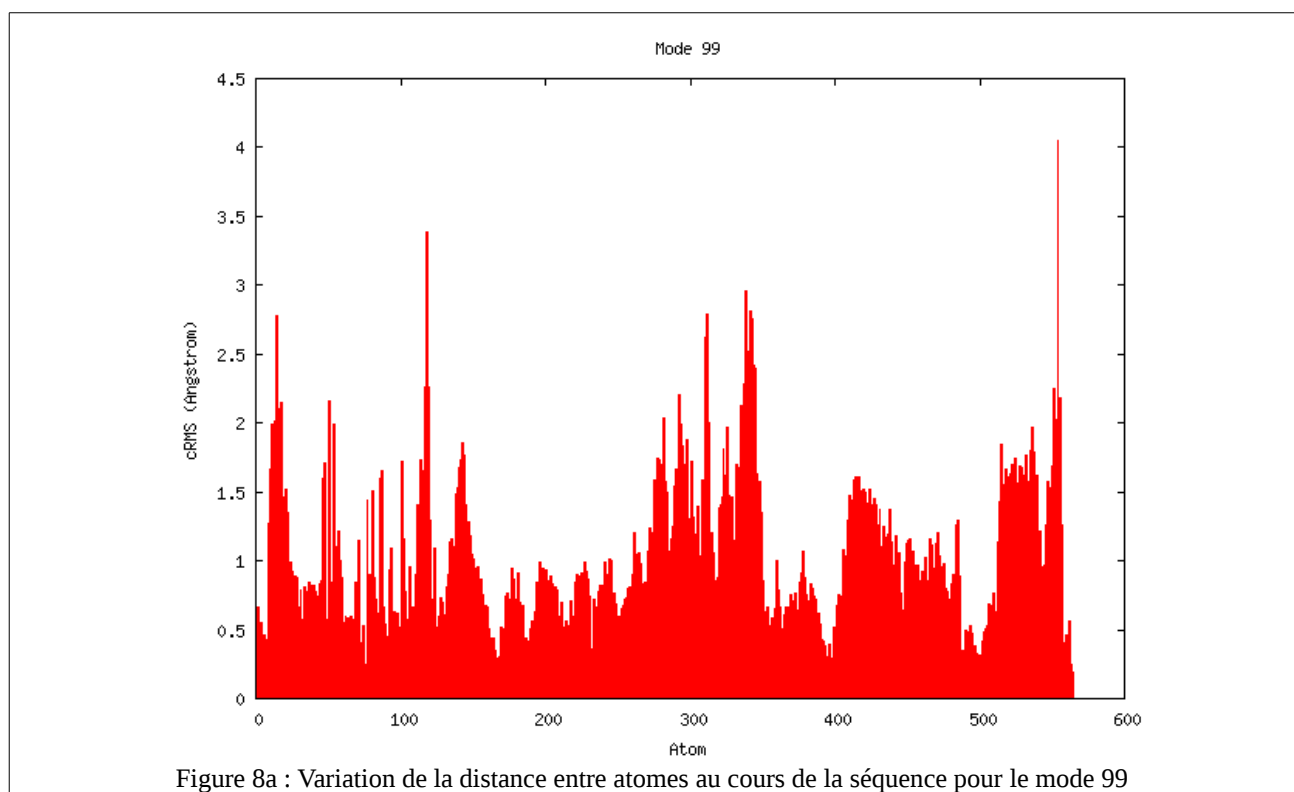
Peut-être qu'avec les résultats issues du raffinement des chaînes latérales et des boucles, nous aurions pu avoir de meilleur score. Cependant, au moment où nous finissons cet article, les résultats ne nous sont toujours pas parvenues.

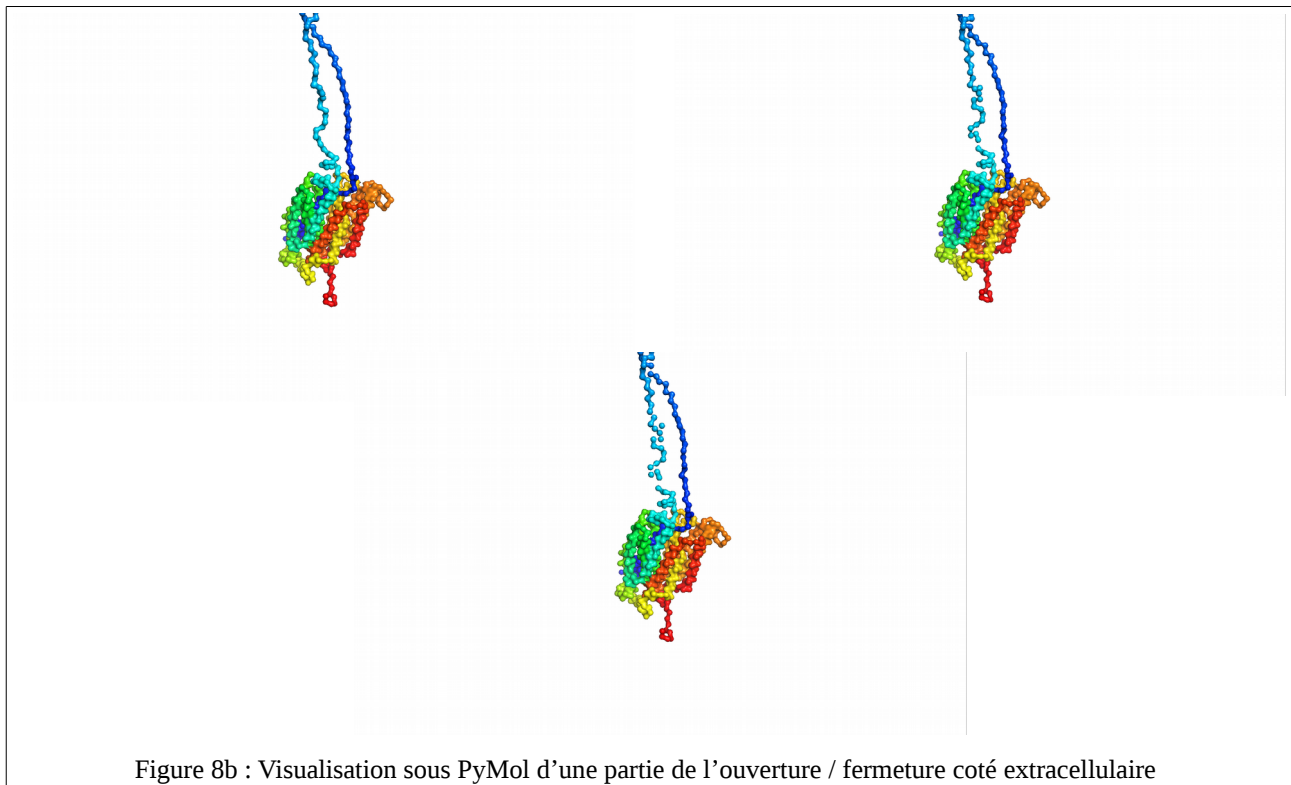
Nous allons à présent, tenter de comprendre les mouvements de cette protéine par l'analyse des modes normaux sur les carbones alphas de ce transporteur.

Modes Normaux

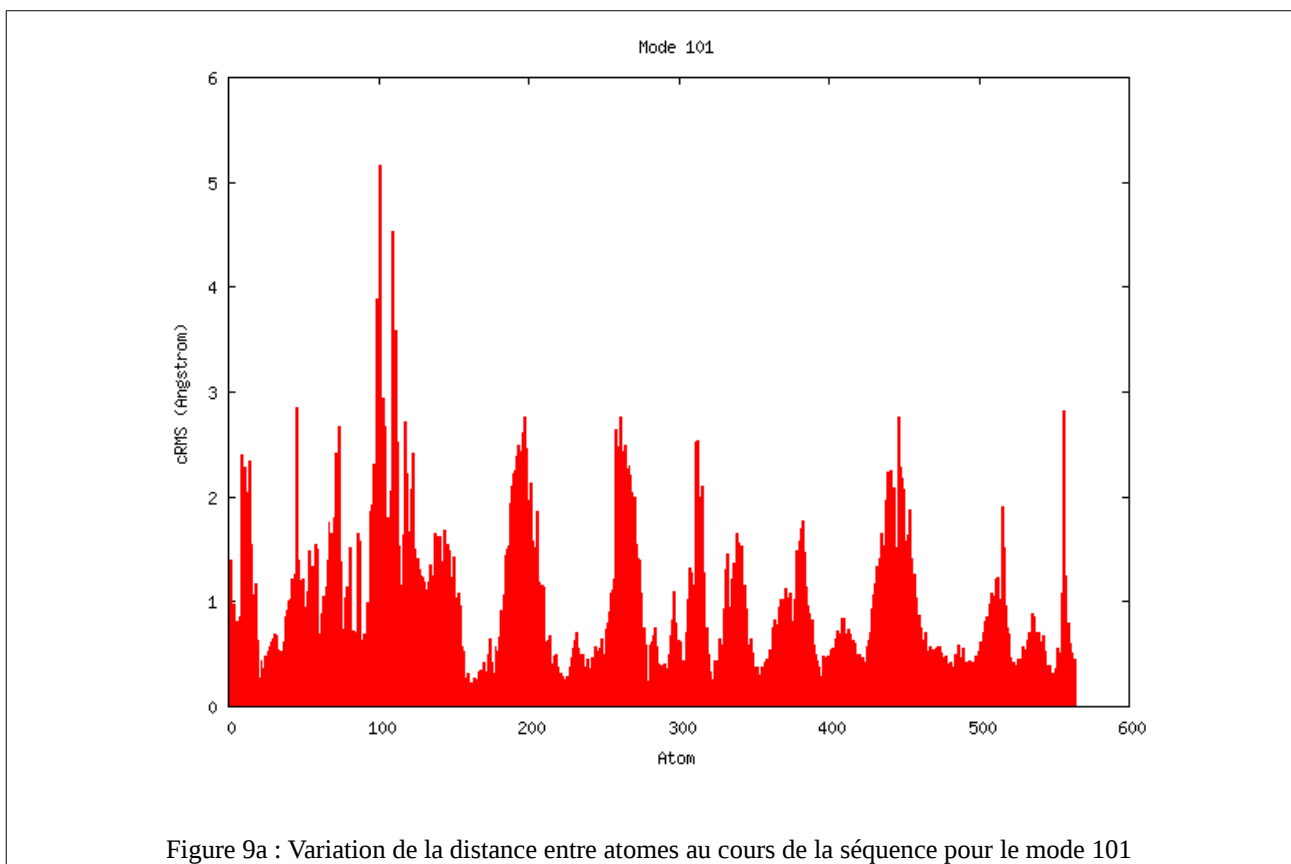
Du fait de la présence d'une grande boucle en extracellulaire, nous avons cherché les modes normaux expliquant au mieux les mouvements des hélices alphas transmembranaires en évitant les mouvements de la boucle.

Nous avons trouvé deux modes qui peuvent en partie expliquer les mouvements de passage de molécule à travers ce transporteur : il s'agit du mode 99 (ouverture / fermeture extracellulaire, cf figure 8) et 101(ouverture / fermeture intracellulaire, cf figure 9).





Sur la figure 8a, les variations entre les positions 0 à 150, expliquent plutôt des mouvements de la boucle. Le reste tendent à expliquer le début d'ouverture des hélices transmembranaires comme figuré sur la figure 8b.



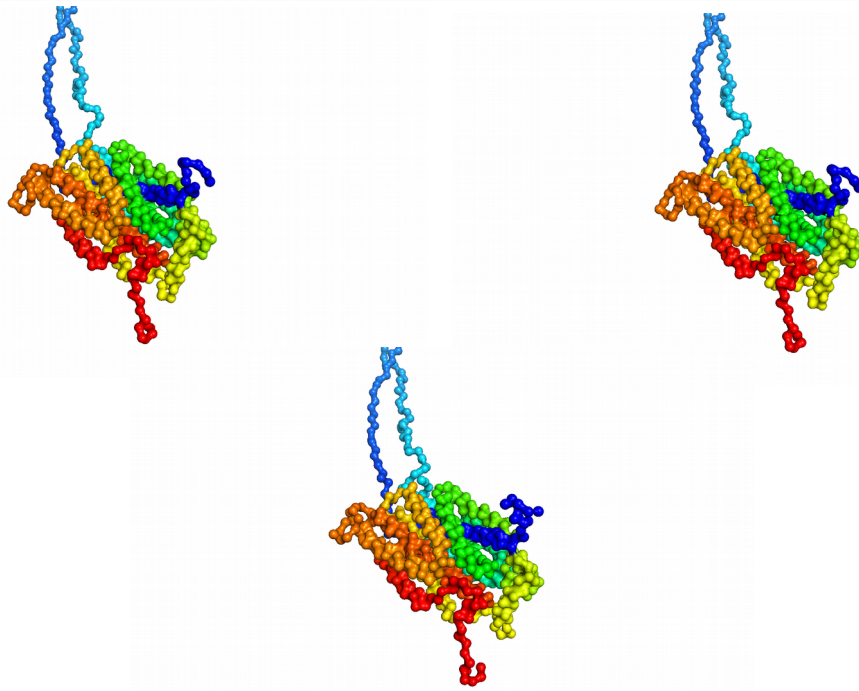


Figure 9b : Visualisation sous PyMol d'une partie de l'ouverture / fermeture coté intracellulaire

Sur la figure 9a, lidem à 8a, les variations entre les positions 0 à 150, expliquent plutôt des mouvements de la boucle. Le reste tendent à expliquer le début d'ouverture des hélices transmembranaires comme figuré sur la figure 9b.

Parmi ces deux modes, on observe plutôt que l'ouverture extracellulaire, impliquent peu de mouvements de l'ensemble de la protéine, tandis que l'ouverture coté intracellulaire impliquent plus de mouvement.

Cependant, les peu de mouvements coté extracellulaire recrutent de grands zones de la protéine, tandis que les mouvements les plus importants de la protéine pour l'ouverture intracellulaire recrutent moins d'atomes pour chaque mouvement.

Bien que nous ayons pu trouver un début de compréhension de l'ouverture coté intracellulaire et extracellulaire du transporteur, nous n'avons pas établie de modes permettant le basculement du ligand à travers la membrane.

Cela peut peut-être s'expliquer par le fait que nous avons fait les modes normaux sur uniquement les carbones alphas, ou peut-être aussi que les mouvements de la boucles 'tire' trop l'information de variation de mouvement vers eux. Nous aurions peut-être dû analyser plus que les 100 premiers modes.

Conclusion

La sous-estimation du temps de calcul de certaines étapes, ne nous permet pas de correctement analyser les variations de mouvement de la protéine concernant le passage des ligands à travers la membrane. En effet, nous sommes toujours en attente des résultats du raffinement des boucles, des chaînes latérales. Par conséquent, nous n'avons pu faire une analyse correcte.

Cependant, nous avons pu avoir un début de réponse concernant la structure de notre transporteur : 12 hélices alphas transmembranaires, 2 hélices intracellulaires, une grande boucle extracellulaire.

De plus, nous avons pu visualiser le début d'ouverture de la protéine coté extracellulaire et coté intracellulaire, mais malheureusement sans pour autant faire le lien entre les deux.

Nous sommes dans l'attente de nos derniers résultats, qui, espérons, vont pouvoir nous aider à améliorer la qualité d'analyse des mouvements de notre protéine. Avec ces derniers résultats, nous pourrions par la suite tenter une modélisation moléculaire afin de comprendre en détails les mouvements de la protéine.

Enfin, une imagerie par cristallographie de notre protéine, apporterait, un grand pas dans la confrontation de nos résultats.

Références

- (1) « The SLC22 drug transporter family », Springer-Verlag, 2003, DOI:10.1007/s00424-003-1089-9
- (2) « The SLC22 family with transporters of organic cations, anions and zwitterions », H.KOEPSSELL, 2012, DOI : 2012.10.010
- (3) « Identification of six putative human transporters with structural similarity to the drug transporter SLC22 family », JACOBSON et al, 2007, DOI : 595 – 609
- (4) « Catalog of 238 variations among six human genes encoding solute carriers (hSLCs) in the Japanese population », SAITO et al, 2002, DOI : 47:576 – 584
- (5) « T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. », Di TOMMASO et al, 2011, DOI : 0.1093/nar/gkr245
- (6) « T-Coffee: A novel method for fast and accurate multiple sequence alignment. », 2000, NOTREDAME et al, DOI : 302(1):205-17
- (7) « Proton-coupled sugar transport in the prototypical major facilitator superfamily protein Xyle », WISEDCHAI SRI et al, 2014, PMCID : PMC4137407
- « The Phyre2 web portal for protein modeling, prediction and analysis », KELLEY et al, 2015, DOI : 10.1038/nprot.2015.053
- « SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information », BIASINI et al, 2014
- « Protein structure homology modelling using SWISS-MODEL Workspace », BORDOLI et al, 2009

- « The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling », ARNOLD et al, 2006
- « Template-based protein structure modeling using the RaptorX web server », KALLBERG et al, 2012, DOI : 7, 1511–1522, 2012.
- « Protein threading using context-specific alignment potential », MA J, 2013, DOI : 10.1093/bioinformatics/btt210