



Reconnaissance moléculaire et régulation en présence d'ARN d'un signal de localisation nucléaire bimodulaire

Allegra Mboukou

► To cite this version:

Allegra Mboukou. Reconnaissance moléculaire et régulation en présence d'ARN d'un signal de localisation nucléaire bimodulaire. Biologie structurale [q-bio.BM]. Université Paris Cité, 2021. Français. NNT : 2021UNIP5007 . tel-04508717

HAL Id: tel-04508717

<https://theses.hal.science/tel-04508717>

Submitted on 18 Mar 2024

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

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

Université de Paris

École doctorale Médicament, Toxicologie, Chimie, Imageries n°563

Laboratoire Expression Génétique Microbienne (UMR 8261)

**Reconnaissance moléculaire et régulation en
présence d'ARN d'un signal de localisation
nucléaire bimodulaire**

Par Allegra MBOUKOU

Thèse de doctorat de **Biologie Structurale**

Dirigée par Pierre BARRAUD
Et par Carine TISNÉ

Présentée et soutenue publiquement le 22 Mars 2021

Devant un jury composé de :

Dr. Thibaut CRÉPIN, Directeur de Recherche, Université de Grenoble – Rapporteur
Dr. Julie MÉNÉTREY, Directrice de Recherche, Université de Paris Saclay – Rapporteur
Pr. Jean-François GUICHOU, Professeur d'université, Université de Montpellier – Examinateur
Dr. Valérie BIOU, Chargée de Recherche – HDR, Université de Paris – Examinateur
Dr. Carine TISNÉ, Directrice de Recherche, Université de Paris – Co-directrice de Thèse
Dr. Pierre BARRAUD, Chargé de Recherche – HDR, Université de Paris – Co-directeur de Thèse



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

Remerciements

Je souhaiterais remercier toutes les personnes ayant contribué de près ou de loin à ma thèse.

Je tiens tout d'abord à remercier Thibaut Crépin ainsi que Julie Ménétrey de me faire l'honneur d'être les rapporteurs de mes travaux de thèse. Merci à Jean-François Guichou et Valérie Biou d'avoir accepté d'évaluer mon travail de thèse et de participer à mon jury de thèse.

Je tiens à exprimer ma gratitude à mes co-directeurs de thèse Pierre Barraud et Carine Tisné. Merci à vous de m'avoir donné l'occasion de faire une thèse et pour tout ce que vous m'avez appris durant ces dernières années. Merci Carine pour ton optimisme et ton soutien, tes précieux conseils et ta disponibilité. Un grand merci pour ton travail de relecture et pour m'avoir aidé à améliorer ce manuscrit. Merci d'avoir créé cette ambiance dans l'équipe qui m'a permis de bien me sentir pendant ces années. Je me suis sentie à ma place et cela m'a permis de m'épanouir dans mon travail. Merci Pierre pour ta confiance, le degré de liberté que tu m'as laissé quant à la conduite de mes travaux ainsi que pour ta disponibilité. Merci pour tes encouragements, tes conseils, ton soutien, d'avoir lu, et relu ce manuscrit et de m'avoir aidé à l'améliorer également. Ta rigueur scientifique sera dorénavant un fil conducteur pour moi.

Je souhaite également remercier les membres de l'équipe « Biogenèse, architecture et interactions des ARNs », particulièrement Marjorie Catala qui m'a encadrée en M2 et qui m'a appris une bonne partie des techniques utilisées pendant la thèse. Merci pour les moments en tête à tête pendant les vacances d'été et de Noël, pour toutes les discussions scientifiques et surtout non scientifiques. J'embarque avec moi avec beaucoup de tes astuces. Un grand merci aux membres de l'équipe (anciens et présents).

Je remercie les membres du laboratoire EGM pour leur accueil et leur convivialité. Merci à l'ensemble de l'IBPC (merci Franck Brachet (mon tueur d'araignées) pour ta disponibilité, un grand merci à Alexandre Pozza pour son aide avec les expériences de SEC-MALS).

Un grand merci à nos collaborateurs avec qui, il a été facile de communiquer et de réfléchir sur ce projet : Michael Jantsch, Renata Kleinova et particulièrement Vinod Rajendra dont les efforts ont permis de mieux orienter mes recherches. Merci également à Tobias Madl et Benjamin Bourgeois pour la réalisation des expériences RMN.

J'aimerais remercier l'organisme qui a financé ma thèse, l'ANR, et le CNRS pour m'avoir permis de prolonger de 3 mois mes travaux.

Enfin, je souhaite exprimer toute ma gratitude à mes amis, olufe mi, ma famille, mes parents, qui ont été d'un soutien sans faille.

MERCI !

Sommaire

Liste des figures	I
Liste des tableaux	V
Liste des structures	VII
Liste des principales abréviations	IX
Introduction	1
1 Le transport nucléo-cytoplasmique	1
1.1 La cellule eucaryote	2
1.1.1 L'organisation structurale de la cellule eucaryote.....	2
1.1.2 Les pores nucléaires : structure et transport	3
1.2 Le cycle de la protéine Ran	9
1.2.1 Les caractéristiques de la protéine Ran	9
1.2.2 La distribution asymétrique des formes de la protéine Ran	10
1.3 La famille des karyophérines β	12
1.3.1 L'organisation structurale des karyophérines β	12
1.3.2 Les fonctions des karyophérines β	14
1.3.3 Les signaux de transport d'import et d'export nucléaire.....	15
1.3.4 Les exportines et l'export nucléaire	17
1.3.5 Les importines et l'import nucléaire.....	18
2 La Transportine 1	19
2.1 La Transportine 1 : une importine des karyophérines β.....	20
2.2 L'organisation structurale de la Transportine 1	22
2.3 La reconnaissance des partenaires de la Transportine 1.....	24
2.3.1 La reconnaissance de RanGTP par la Transportine 1.....	24
2.3.2 La reconnaissance des différents substrats de la Transportine 1.....	26
3 L'import nucléaire d'ADAR1 par la Transportine 1.....	36
3.1 La famille des protéines ADAR.....	36
3.1.1 L'édition des ARN par les enzymes ADAR	37
3.1.2 Les protéines de la famille ADAR.....	38
3.1.3 La sélectivité et la spécificité de l'édition par les enzymes ADAR	40
3.2 La protéine ADAR1	44
3.2.1 L'organisation structurale d'ADAR1.....	44
3.2.2 La localisation des isoformes d'ADAR1.....	49
3.2.3 Le NLS bimodulaire d'ADAR1	51
Objectifs de la thèse et organisation du manuscrit.....	57
Matériels et méthodes.....	61

I.	<i>La reconnaissance d'ADAR1 par Trn1</i>	61
1	<i>Obtention du matériel</i>	62
1.1	Production de la protéine Transportine 1	63
1.1.1	Surexpression	63
1.1.2	Purification	64
1.2	Production de la protéine ADAR1	66
1.2.1	Surexpression	67
1.2.2	Purification	68
1.3	Préparation du peptide ADAR1_pep	71
1.4	Construction et production des protéines chimériques	71
1.4.1	Souches bactériennes, vecteur plasmidique	73
1.4.2	Surexpression et purification	74
2	<i>Caractérisation du complexe Trn1/ADAR1</i>	74
2.1	Caractérisation physico-chimique des protéines	75
2.1.1	GST Pull-Down	75
2.1.2	Differential Scanning Fluorimetry (DSF)	76
2.1.3	Chromatographie d'exclusion stérique couplée à la diffusion de lumière multi-angles et réfractométrie (SEC-MALS)	78
2.1.4	Chromatographie d'exclusion stérique (SEC) analytique	79
2.1.5	Calorimétrie par titration isotherme (ITC)	79
2.2	Caractérisation par des études structurales	80
2.2.1	Études structurales par cristallographie aux rayons X	80
2.2.2	Cristallisation et résolution de structure du complexe Trn1/ADAR1	96
2.2.3	Étude structurale en solution	100
II.	<i>La reconnaissance d'un set de non-PY-NLS par Trn1</i>	104
1	<i>Obtention du matériel</i>	104
1.1	Préparation des peptides synthétiques	104
1.2	Production des peptides recombinants	105
1.2.1	Surexpression	105
1.2.2	Purification	106
1.3	Construction et production des protéines chimériques	108
1.3.1	Souches bactériennes, vecteur plasmidique	108
1.3.2	Surproduction et purification	109
2	<i>Caractérisation des complexes Trn1/non-PY-NLS</i>	109
2.1	Caractérisation physico-chimique des protéines	109
2.1.1	GST Pull-Down	109
2.1.2	Differential Scanning Fluorimetry (DSF)	109
2.1.3	Interférométrie de bio-couche (BLI)	110
2.2	Cristallisation des complexes Trn1/non-PY-NLS	111
III.	<i>La reconnaissance d'un ARN double brin par ADAR1</i>	112

1	<i>Obtention du matériel.....</i>	<i>112</i>
1.1	<i>Production de la protéine ADAR1ΔNLS.....</i>	<i>112</i>
1.1.1	<i>Surexpression.....</i>	<i>112</i>
1.1.2	<i>Purification.....</i>	<i>113</i>
1.2	<i>Préparation de l'ARN double brin.....</i>	<i>115</i>
2	<i>Caractérisation du complexe ADAR1/ARNdb.....</i>	<i>116</i>
	<i>Résultats.....</i>	<i>119</i>
I.	<i>Reconnaissance d'ADAR1 par Trn1.....</i>	<i>119</i>
1	<i>Caractérisation du complexe Trn1ΔH8 / ADAR1court.....</i>	<i>119</i>
1.1	<i>Purification de Trn1ΔH8.....</i>	<i>119</i>
1.2	<i>Purification d'ADAR1court.....</i>	<i>121</i>
1.3	<i>Caractérisation du complexe Trn1ΔH8/ADAR1court.....</i>	<i>123</i>
1.3.1	<i>Cristallisation de Trn1ΔH8.....</i>	<i>123</i>
1.3.2	<i>Résolution de la structure cristallographique de Trn1ΔH8.....</i>	<i>125</i>
1.3.3	<i>Cristallisation du complexe Trn1ΔH8/ADAR1court.....</i>	<i>134</i>
2	<i>Caractérisation du complexe Trn1ΔH8/ADAR1_{pep}.....</i>	<i>135</i>
2.1	<i>Étude de l'interaction Trn1ΔH8/ADAR1_{pep} par DSF et GST Pull-down.....</i>	<i>135</i>
2.2	<i>Utilisation d'une protéine chimère pour la caractérisation de l'interaction</i>	
	<i>Trn1ΔH8/ADAR1_{pep}.....</i>	<i>137</i>
2.2.1	<i>Préparation et purification des protéines chimères.....</i>	<i>137</i>
2.2.2	<i>Cristallisation des protéines chimères.....</i>	<i>139</i>
3	<i>Caractérisation du complexe Trn1ΔH8/ADAR1long.....</i>	<i>141</i>
3.1	<i>Purification d'ADAR1long.....</i>	<i>141</i>
3.2	<i>Étude de l'interaction Trn1ΔH8/ADAR1long par SEC.....</i>	<i>143</i>
3.3	<i>Étude de l'interaction Trn1ΔH8/ADAR1long par SEC-MALS.....</i>	<i>145</i>
4	<i>Caractérisation du complexe Trn1FL/ADAR1long.....</i>	<i>150</i>
4.1	<i>Purification de Trn1FL.....</i>	<i>151</i>
4.2	<i>Caractérisation physico-chimique du complexe Trn1FL/ADAR1long.....</i>	<i>152</i>
4.2.1	<i>Étude de l'interaction Trn1FL/ADAR1long par SEC-MALS.....</i>	<i>152</i>
4.2.2	<i>Étude structurale du complexe Trn1FL/ADAR1long en solution.....</i>	<i>155</i>
4.2.3	<i>Étude structurale du complexe Trn1FL/ADAR1long par cristallographie.....</i>	<i>167</i>
5	<i>Conclusion.....</i>	<i>173</i>
II.	<i>Reconnaissance d'un set de non-PY-NLS par Trn1.....</i>	<i>174</i>
1	<i>Caractérisation des complexes avec l'utilisation de peptides synthétiques.....</i>	<i>175</i>
1.1	<i>Caractérisation physico-chimique des complexes Trn1ΔH8/peptide.....</i>	<i>175</i>

1.1.1	Étude des interactions Trn1ΔH8/peptide par DSF et GST Pull-down	175
1.1.2	Étude des interactions TrnΔH8/peptide par BLI.....	177
1.2	Cristallisation des complexes TrnΔH8/peptide	178
2	<i>Caractérisation des complexes avec l'utilisation de protéines chimères.....</i>	180
2.1	Purification des protéines chimères	181
2.2	Étude de la stabilité des protéines chimères	182
2.3	Cristallisation des protéines chimères	183
2.3.1	Cristallisation de la protéine chimère Trn1ΔH8-HPV16_pep	183
2.3.2	Cristallisation de la protéine chimère Trn1ΔH8-HPV18_pep	184
2.3.3	Cristallisation de la protéine chimère Trn1ΔH8-REV_pep	185
3	<i>Caractérisation des complexes avec l'utilisation de peptides recombinants.....</i>	186
3.1	Purification des protéines SUMO-REV-peptide et SUMO- HPV16-peptide.....	186
3.2	Tests de co-cristallisation de Trn1ΔH8 et des peptides recombinants.	187
4	<i>Conclusion</i>	189
III.	<i>Reconnaissance d'un ARN double brin par ADAR1.....</i>	189
1	<i>Purification d'ADAR1ΔNLS</i>	190
2	<i>Structure cristallographique d'ADAR1ΔNLS</i>	191
2.1	Cristallisation d'ADAR1ΔNLS.....	191
2.2	Collecte et traitement des données de diffraction.....	192
2.3	Phasage par remplacement moléculaire.....	193
2.4	Affinement et validation du modèle.....	194
2.5	Structure générale d'ADAR1ΔNLS.....	195
3	<i>Structure cristallographique du complexe ADAR1ΔNLS/ARNdb</i>	200
3.1	Cristallisation du complexe ADAR1ΔNLS/ARNdb.....	200
3.1.1	Tests de co-cristallisation du complexe ADAR1ΔNLS/ARN-I et ADAR1ΔNLS/ARN-II	200
3.1.2	Cristallisation du complexe ADAR1ΔNLS/ARN-III	203
3.2	Structure générale du complexe ADAR1ΔNLS/ARN-III.....	207
3.3	Détails de l'interaction ADAR1ΔNLS/ARN-III	210
3.3.1	Détails de l'interaction canonique ADAR1ΔNLS/ARN-III	210
3.3.2	Implication de l'hélice αN dans la reconnaissance de l'ARN-III.....	213
4	<i>Conclusion</i>	213
	<i>Conclusion et perspectives</i>	219
I.	<i>La reconnaissance d'ADAR1 par Trn1.....</i>	220
II.	<i>La reconnaissance d'un ARN double brin par ADAR1</i>	222

III. <i>La reconnaissance d'un set de non-PY-NLS par Trn1</i>	226
<i>Bibliographie</i>	231
<i>Article</i>	248

Liste des figures

FIGURE 1 : ORGANISATION STRUCTURALE DE LA CELLULE.	3
FIGURE 2 : CARACTERISTIQUES STRUCTURALES DES PORES NUCLEAIRES.	5
FIGURE 3 : SCHEMATISATION DE L'ORGANISATION DU RESEAU FORME PAR LES DOMAINES FG.	7
FIGURE 4 : DIFFUSION FACILITE DES MACROMOLECULES SELON LE MODELE DE L'HYDROGEL.	8
FIGURE 5 : REPRESENTATION DE LA STRUCTURE DE LA PROTEINE RAN.	9
FIGURE 6 : REPRESENTATION SCHEMATIQUE DU CYCLE DE LA PROTEINE RAN.	11
FIGURE 7 : SCHEMA SIMPLIFIE DE LA STRUCTURE EN SUPER-HELICE DES KARYOPHERINES β .	13
FIGURE 8 : CLASSIFICATION SIMPLIFIEE DES DIFFERENTS SIGNAUX DE TRANSPORT RECONNUS PAR LES KARYOPHERINES β .	16
FIGURE 9 : SCHEMA SIMPLIFIE DU PROCESSUS DE TRANSPORT NUCLEOCYTOPLASMIQUE MEDIE PAR LES KARYOPHERINES β .	17
FIGURE 10 : ORGANISATION STRUCTURALE DE LA TRANSPORTINE 1.	23
FIGURE 11 : RECONNAISSANCE DE RANGTP PAR Trn1.	25
FIGURE 12 : DISSOCIATION DES CARGOS DE Trn1.	26
FIGURE 13 : SURFACES D'INTERACTION DES PY-NLS.	28
FIGURE 14 : STRUCTURES CRISTALLOGRAPHIQUES DE LA RECONNAISSANCE DES PY-NLS PAR Trn1.	31
FIGURE 15 : ALIGNEMENT DE SEQUENCES DE NEUF PY-NLS CONNUS.	32
FIGURE 16 : RECONNAISSANCE D'UN NON-PY-NLS PAR Trn1.	34
FIGURE 17 : REACTION DE DESAMINATION HYDROLYTIQUE DE L'ADENOSINE EN INOSINE.	38
FIGURE 18 : ORGANISATION STRUCTURALE DES MEMBRES DE LA FAMILLE DES PROTEINES ADAR.	39
FIGURE 19 : EDITION D'ADENOSINE (A) EN INOSINE (I) DANS DES SEQUENCES NON CODANTES D'ARNDB.	44
FIGURE 20 : ORGANISATION STRUCTURALE DE L'ENZYME ADAR1.	45
FIGURE 21 : STRUCTURE TOPOLOGIQUE D'UN DOMAINE DE LIAISON A L'ARN DOUBLE BRIN (dsRBD).	48
FIGURE 22 : RECONNAISSANCE CANONIQUE D'UN ARNDB PAR UN dsRBD.	49
FIGURE 23 : STRUCTURE DU dsRBD3 D'ADAR1PORTANT SON NLS BIMODULAIRE.	52
FIGURE 24 : LE MODELE D'INACTIVATION DU NLS D'ADAR1EN PRESENCE D'ARN.	54
FIGURE 25 : LES CONSTRUCTIONS DE Trn1 ET ADAR1.	62
FIGURE 26 : SCHEMA DE LA CONCEPTION DES PROTEINES CHIMERES.	72
FIGURE 27 : CARTE PLASMIDIQUE PAM_NONPY-NLS.	73
FIGURE 28 : PRINCIPE DU GST-PULL DOWN.	75
FIGURE 29 : PRINCIPE DU DSF.	76
FIGURE 30 : DIAGRAMME DE PHASE POUR LA CRISTALLISATION DES PROTEINES.	81
FIGURE 31 : COMPOSITION D'UN CRISTAL.	87
FIGURE 32 : TABLEAU REGROUPANT LES 14 RESEAUX DE BRAVAIS.	87
FIGURE 33 : PRINCIPE DE L'INTERFEROMETRIE DE BIO-COUCHES (BLI).	110
FIGURE 34 : SCHEMA REPRESENTANT LA FORMATION D'UN ARNDB A PARTIR DE LA SEQUENCE "PALINDROMIQUE".	115
FIGURE 35 : PURIFICATION DE Trn1 Δ H8.	121
FIGURE 36 : ANALYSE SDS-PAGE DES DIFFERENTES ETAPES DE PURIFICATION D'ADAR1COURT.	122
FIGURE 37 : CRISTALLISATION DE Trn1 Δ H8.	124
FIGURE 38 : OPTIMISATION DE LA CRISTALLISATION DE Trn1 Δ H8.	125
FIGURE 39 : COMPARAISON DES CARTES DE DENSITE ELECTRONIQUE DES STRUCTURES CRISTALLINES DE Trn1.	129
FIGURE 40 : COMPARAISON DE L'EMPILEMENT CRISTALLIN DES STRUCTURES DE Trn1.	130
FIGURE 41 : COMPARAISON DES STRUCTURES CRISTALLOGRAPHIQUES DE Trn1.	131
FIGURE 42 : REPRESENTATION GRAPHIQUE COMPARATIVE DES FACTEURS B DES STRUCTURES CRISTALLOGRAPHIQUES DE Trn1.	132
FIGURE 43 : COMPARAISON DES HEAT H18-H20 DES STRUCTURES CRISTALLOGRAPHIQUES DE Trn1.	133
FIGURE 44 : COMPARAISON DES HEAT H1-H4 DES STRUCTURES CRISTALLOGRAPHIQUES DE Trn1.	134
FIGURE 45 : ÉTUDE BIOCHIMIQUE DE L'INTERACTION Trn1 Δ H8/ADAR1_PEP.	136

FIGURE 46 : EXPRESSION DES CHIMERES Trn1ΔH8-A1_PEP ET Trn1ΔH8-ADAR1_PEP.	138
FIGURE 47 : PURIFICATION DES CHIMERES Trn1ΔH8-ADAR1_PEP ET Trn1ΔH8-A1_PEP.	139
FIGURE 48 : CRISTALLISATION DE LA CHIMERE Trn1ΔH8-A1_PEP.	140
FIGURE 49 : PURIFICATION DE GST-ADAR1LONG.	142
FIGURE 50 : PURIFICATION D'ADAR1LONG.	143
FIGURE 51 : TESTS D'INTERACTION PAR SEC DES PROTEINES Trn1ΔH8 ET ADAR1LONG.	144
FIGURE 52 : ANALYSE PAR SEC-MALS DE Trn1ΔH8 ET ADAR1LONG SEULES.	146
FIGURE 53 : ANALYSE PAR SEC-MALS DE Trn1ΔH8 EN PRESENCE D'ADAR1LONG / ADAR1COURT.	147
FIGURE 54 : ANALYSE PAR SEC-MALS DE Trn1ΔH8 ET ADAR1LONG SEULES APRES OPTIMISATION.	148
FIGURE 55 : ANALYSE PAR SEC-MALS DE Trn1ΔH8 EN PRESENCE D'ADAR1LONG A UN RATIO 1:10.	149
FIGURE 56 : PURIFICATION DE Trn1FL.	152
FIGURE 57 : ANALYSE PAR SEC-MALS DE Trn1FL SEULE.	153
FIGURE 58 : ANALYSE PAR SEC-MALS DE Trn1FL EN PRESENCE D'ADAR1LONG.	154
FIGURE 59 : COURBES DES INTENSITES DIFFUSEES ET DES RAYONS DE GIRATION DE Trn1FL EN ABSENCE ET PRESENCE D'ADAR1LONG.	156
FIGURE 60 : COURBES EXPERIMENTALES DE Trn1FL EN ABSENCE ET EN PRESENCE D'ADAR1LONG.	157
FIGURE 61 : CALCULS DES RAYONS DE GIRATION DE Trn1FL EN ABSENCE ET EN PRESENCE D'ADAR1LONG.	157
FIGURE 62 : FONCTIONS PAIRES DE Trn1FL EN ABSENCE ET EN PRESENCE D'ADAR1LONG.	158
FIGURE 63 : COMPARAISON DES DONNEES SAXS EXPERIMENTALES DE Trn1FL SEULE AVEC LES COURBES THEORIQUES CALCULEES.	160
FIGURE 64 : MODELE DE Trn1FL GENERE PAR DEFORMATION PAR MODES NORMAUX.	161
FIGURE 65 : REPRESENTATION EN SURFACE DU MODELE STRUCTURAL DU COMPLEXE Trn1/dsRBD3 D'ADAR1.	162
FIGURE 66 : REPRESENTATION DES MODELES DU COMPLEXE Trn1FL/ADAR1LONG GENERES PAR DEFORMATION PAR MODES NORMAUX ET PAR MODELISATION EN CORPS RIGIDE.	163
FIGURE 67 : SUPERPOSITION DES SPECTRES HSQC ¹ H- ¹⁵ N DES EXPERIENCES RMN DE Trn1FL.	165
FIGURE 68 : SUPERPOSITION DE SPECTRES HSQC ¹ H- ¹⁵ N DES EXPERIENCES RMN DE LA BOUCLE H8 DE Trn1.	166
FIGURE 69 : MESURE ITC DES COMPLEXES Trn1FL/Trn1ΔH8 / GST-ADAR1LONG/GST-ADAR1COURT.	168
FIGURE 70 : EXPRESSION DE LA CHIMERE Trn1FL-ADAR1LONG.	169
FIGURE 71 : PURIFICATION DE LA CHIMERE Trn1FL-ADAR1LONG.	170
FIGURE 72 : CRISTALLISATION DE LA CHIMERE Trn1FL-ADAR1LONG.	171
FIGURE 73 : FORMATION DU COMPLEXE Trn1FL/GST-ADAR1LONG.	173
FIGURE 74 : CARACTERISATION DES INTERACTIONS Trn1ΔH8/PEPTIDES.	176
FIGURE 75 : ÉTUDE QUANTITATIVE DE L'INTERACTION Trn1ΔH8/PEPTIDES.	178
FIGURE 76 : TESTS DE CO-CRISTALLISATION DES COMPLEXES Trn1ΔH8/PEPTIDE.	179
FIGURE 77 : EXPRESSION DES CHIMERES Trn1ΔH8-PEPTIDES.	181
FIGURE 78 : PURIFICATION DES CHIMERES Trn1ΔH8-PEPTIDES.	182
FIGURE 79 : ÉTUDE DE LA STABILITE DES CHIMERES Trn1ΔH8-PEPTIDE.	183
FIGURE 80 : CRISTALLISATION DE LA CHIMERE Trn1ΔH8-HPV16_PEP.	184
FIGURE 81 : CRISTALLISATION DE LA CHIMERE Trn1ΔH8-HPV18_PEP.	185
FIGURE 82 : PURIFICATION DES PROTEINES SUMO-REV_NLS ET SUMO-HPV16_NLS.	187
FIGURE 83 : TESTS DE CO-CRISTALLISATION DU COMPLEXE Trn1ΔH8/SUMO-REV_NLS.	188
FIGURE 84 : TESTS DE CO-CRISTALLISATION DU COMPLEXE Trn1ΔH8/SUMO-HPV16_NLS.	188
FIGURE 85 : PURIFICATION DE LA PROTEINE ADAR1ΔNLS.	191
FIGURE 86 : CRISTALLISATION D'ADAR1ΔNLS.	192
FIGURE 87 : STRUCTURE GENERALE D'ADAR1ΔNLS.	195
FIGURE 88 : COMPARAISON DES CHAINES DE LA STRUCTURE CRISTALLINE D'ADAR1ΔNLS.	196
FIGURE 89 : COMPARAISON DE L'EMPILEMENT CRISTALLIN DES CHAINES DE LA STRUCTURE CRISTALLINE D'ADAR1ΔNLS.	197
FIGURE 90 : COMPARAISON DES CHAINES A (EN MAGENTA) ET CHAÎNE B (EN VERT) DE LA STRUCTURE CRISTALLINE AVEC LA STRUCTURE RMN EN SOLUTION (EN ORANGE) (CODE PDB (2MDR)) D'ADAR1ΔNLS.	198

FIGURE 91 : COMPARAISON DES HELICES αN DES CHAINES A (EN MAGENTA) ET B (EN VERT) DE LA STRUCTURE CRISTALLINE ET DE LA STRUCTURE EN SOLUTION (EN ORANGE) (CODE PDB (2MDR)) D'ADAR1 Δ NLS.	199
FIGURE 92 : TESTS DE CO-CRISTALLISATION DU COMPLEXE ADAR1 Δ NLS/ARN-I.	201
FIGURE 93 : OPTIMISATION DES TESTS DE CO-CRISTALLISATION DU COMPLEXE ADAR1 Δ NLS/ARN-I.	202
FIGURE 94 : TESTS DE CO-CRISTALLISATION DU COMPLEXE ADAR1 Δ NLS/ARN-II.	202
FIGURE 95 : TESTS DE CO-CRISTALLISATION DU COMPLEXE ADAR1 Δ NLS/ARN-III.	203
FIGURE 96 : COMPOSITION DE L'UNITE ASYMETRIQUE DU MODELE DU COMPLEXE ADAR1 Δ NLS/ARN-III.	206
FIGURE 97 : STRUCTURE GENERALE DU COMPLEXE ADAR1 Δ NLS/ARN-III.	208
FIGURE 98 : COMPARAISON DES CHAINES A DE LA STRUCTURE LIBRE ET EN COMPLEXE D'ADAR1 Δ NLS.	208
FIGURE 99 : REPRESENTATION DE L'INTERACTION CANONIQUE ENTRE ADAR1 Δ NLS ET L'ARN-III.	209
FIGURE 100 : REPRESENTATION DE L'INTERACTION CANONIQUE ADAR1 Δ NLS/ARN-III AU NIVEAU DE LA REGION 1.	210
FIGURE 101 : REPRESENTATION DE L'INTERACTION CANONIQUE ADAR1 Δ NLS/ARN-III AU NIVEAU DE LA REGION 2.	211
FIGURE 102 : REPRESENTATION DE L'INTERACTION CANONIQUE ADAR1 Δ NLS/ARN-III AU NIVEAU DE LA REGION 3.	212
FIGURE 103 : REPRESENTATION DE L'INTERACTION ADAR1 Δ NLS/ARN-III IMPLIQUANT L'HELICE αN .	213
FIGURE 104 : SCHEMA DE L'ASSEMBLAGE DE L'ARN-III ET DE SA RECONNAISSANCE CANONIQUE PAR ADAR1 Δ NLS.	214
FIGURE 105 : SCHEMA RECAPITULATIF DES PRINCIPAUX AXES ETUDIES.	225

Liste des tableaux

TABEAU 1 : TABLEAU REGROUPANT LES PROTEINES DE LA FAMILLE DES KARYOPHERINES β .	15
TABEAU 2: TABLEAU REGROUPANT LES DIFFERENTS CARGOS DE MAMMIFERES DE Trn1.	21
TABEAU 3 : TABLEAU REGROUPANT LES STRUCTURES CRISTALLOGRAPHIQUES DE COMPLEXE Trn1 ET DIFFERENTS PY-NLS.	29
TABEAU 4 : TABLEAU REGROUPANT LES DIFFERENTES CONSTRUCTIONS DE LA TRANSPORTINE 1 ET ADAR1.	63
TABEAU 5 : TABLEAU REGROUPANT LES DIFFERENTES CONSTRUCTIONS CHIMERIQUES.	74
TABEAU 6 : TABLEAU REGROUPANT LES DIFFERENTS PEPTIDES NON-PY-NLS SYNTHETIQUES.	105
TABEAU 7 : TABLEAU REGROUPANT LES DIFFERENTS PEPTIDES NON-PY-NLS RECOMBINANTS.	105
TABEAU 8 : TABLEAU REGROUPANT LES DIFFERENTES CONSTRUCTIONS CHIMERIQUES (Trn1 ET LES DIFFERENTS NON-PY-NLS).	108
TABEAU 9 : TABLEAU REGROUPANT LES DIFFERENTS ARN SYNTHETIQUES ETUDIES.	116
TABEAU 10 : LES STATISTIQUES DU JEU DE DONNEES DE Trn1 Δ H8.	126
TABEAU 11 : DETERMINATION DU COEFFICIENT DE MATTHEWS POUR LA STRUCTURE CRISTALLINE DE Trn1 Δ H8.	127
TABEAU 12 : LES STATISTIQUES D'AFFINEMENT DE LA STRUCTURE DE Trn1 Δ H8.	128
TABEAU 13 : TABLEAU COMPARATIF DES TROIS STRUCTURES CRISTALLOGRAPHIQUES DE Trn1.	130
TABEAU 14 : TABLEAU COMPARATIF DES FACTEURS B DES STRUCTURES DE Trn1.	131
TABEAU 15 : SYNTHESE DES EXPERIENCES SEC-MALS DE Trn1 Δ H8.	150
TABEAU 16 : SYNTHESE DES EXPERIENCES SEC-MALS DE Trn1FL.	154
TABEAU 17 : SYNTHESE DES EXPERIENCES DE Trn1FL EN ABSENCE ET EN PRESENCE D'ADAR1LONG.	159
TABEAU 18 : TABLEAU REGROUPANT LES DIFFERENTS PEPTIDES SYNTHETIQUES ETUDIES.	175
TABEAU 19 : SEQUENCES DES NLS UTILISES POUR LA CONSTRUCTION DES CHIMERES.	180
TABEAU 20 : TABLEAU REGROUPANT LES STATISTIQUES (XDS) DU JEU DE DONNEES D'ADAR1 Δ NLS.	193
TABEAU 21 : TABLEAU DE PROBABILITE DU NOMBRE DE MOLECULES PAR UNITE ASYMETRIQUE DE LA STRUCTURE D'ADAR1 Δ NLS (DETERMINEE PAR LA METHODE MATTHEWS (185)).	193
TABEAU 22 : TABLEAU REGROUPANT LES STATISTIQUES (REFMAC ET MOLPROBITY) D'AFFINEMENT DE LA STRUCTURE CRISTALLINE D'ADAR1 Δ NLS.	194
TABEAU 23 : TABLEAU REGROUPANT LES STATISTIQUES (XDS ET STARANISO) DU JEU DE DONNEES D'ADAR1 Δ NLS EN COMPLEXE AVEC L'ARN-III.	205
TABEAU 24 : TABLEAU REGROUPANT LES STATISTIQUES (REFMAC ET MOLPROBITY) D'AFFINEMENT DE LA STRUCTURE CRISTALLOGRAPHIQUE D'ADAR1 Δ NLS EN COMPLEXE AVEC L'ARN-III.	207

Liste des structures

Structure cristallographique de Trn1 Δ H8 :

Figure 41, page 131

Modèle structural de Trn1FL généré par modes normaux :

Figure 64, page 161

Modèle structural du complexe Trn1/ADAR1 généré par modes normaux et modélisation en corps rigides : Figure 66, page 163

Structure cristallographique du dsRBD3 d'ADAR1 :

Figure 87, page 195

Structure cristallographique du complexe dsRBD3/ARNdb :

Figure 97, page 208

Liste des principales abréviations

°C/K: Degré Celsius/Kelvin

A, U, G, C: Adénine, Uracile, Guanine, Cytosine

Å: Ångström

ADAR: Adenosine deaminase acting on RNA

ARN / ARNdb: Acide RiboNucléique / Acide RiboNucléique double brin

CC_{1/2}: Coefficient de corrélation calculé à partir du pourcentage de corrélation entre deux intensités aléatoires

D_{max}: Distance maximale

DO: Densité optique

dsRBD: Double-stranded RNA binding domain

DTT: Dithiothréitol

E. coli: *Escherichia coli*

GDP: Guanosine Diphosphate

GTP: Guanosine Triphosphate

HEPES: Acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique

HRV 3C: Recombinant type 14 3C protease from human rhinovirus

I/σ(I): Rapport signal sur bruit

IPTG: Isopropyl β-D-1-thiogalactopyranoside

ITC: Titration calorimétrique isotherme

K_D: Constante de dissociation

LB: lysogeny broth

PAA: Poly (Acrylic acid-co-maleic) Acid

PCR: Polymerase Chain Reaction

PDB: Protein Data Bank

PEG : PolyEthylene Glycol

pH: potentiel hydrogène

pI: point isoélectrique

Rfacteur/Rlibre: coefficients d'accord entre les facteurs de structure observés et calculés

R_g: Rayon de giration

RMN: Résonance magnétique nucléaire

SAXS: Small Angles X Ray Scattering

SDS-PAGE: Sodium Dodécyl Sulfate-Polyacrylamide Gel Electrophoresis

SEC: Size exclusion chromatography

SEC-MALLS: Size exclusion chromatography-Multi Angle Laser Light Scattering

TEV: Tobacco Etch Virus

Trn1: Transportine 1

Tris: 2-amino-2-hydroxyméthyl-1,3-propanediol

UV: Ultraviolet

Introduction

Introduction

1 Le transport nucléo-cytoplasmique

Le transport nucléocytoplasmique est un processus cellulaire par lequel des macromolécules comme les protéines ou les ARN sont transférées du noyau au cytoplasme des cellules eucaryotes de manière bidirectionnelle. Ce mécanisme s'effectue au travers des canaux qui perforent la double membrane du noyau.

La caractéristique déterminante de la cellule eucaryote est la présence d'un compartiment séparé du reste du contenu cellulaire, qui contient le matériel génétique, on l'appelle le noyau. Ainsi, la compartimentation du matériel génétique dans le noyau entraîne la séparation spatiale et temporelle de la réplication et de la transcription de l'ADN dans le noyau, de la synthèse protéique dans le cytoplasme. De ce fait, des milliers de macromolécules transitent entre le noyau et le cytoplasme à travers la double membrane qui entoure le noyau, l'enveloppe nucléaire. En effet, les ARN transcrits sont exportés du noyau pour atteindre les ribosomes dans le cytoplasme où ils seront traduits en protéines. Et, à l'inverse, les protéines comme les histones ou les facteurs de transcription sont importés du cytoplasme vers le noyau pour atteindre l'ADN. Cette compartimentation permet non seulement une haute régulation de ces différents processus mais aussi, nécessite un système de transport de macromolécules sélectif entre le noyau et le cytoplasme (1–5). La sélectivité du transport nucléocytoplasmique repose en partie sur la structure des pores nucléaires permettant les échanges entre les deux compartiments, et dépend également d'une variété de récepteurs de transport nucléaire, des protéines qui reconnaissent les cargos destinés à passer les pores nucléaires.

Dans cette introduction, qui présente le transport nucléocytoplasmique, nous allons dans un premier temps, détailler l'organisation de la cellule eucaryote en présentant ses différentes caractéristiques. Nous nous intéresserons ensuite aux pores nucléaires permettant le passage des macromolécules. Enfin, nous discuterons des mécanismes et acteurs de l'import et de l'export de macromolécules au travers des pores nucléaires.

1.1 La cellule eucaryote

La cellule, l'unité fonctionnelle et structurelle du vivant, est la plus petite structure qui présente les propriétés du vivant. Un être humain est constitué d'environ 100.000 milliards de cellules, et chacune de ces cellules est une petite unité vivante. On reconnaît deux grands types de cellules : les cellules eucaryotes et les cellules procaryotes.

1.1.1 L'organisation structurale de la cellule eucaryote

Les cellules eucaryotes sont des cellules dont la taille varie en moyenne entre 10 et 100 micromètres. Elles sont entourées par une membrane plasmique délimitant un espace intérieur, l'espace intracellulaire d'un espace extérieur, l'espace extracellulaire. L'espace intracellulaire contient le cytoplasme et le noyau (Figure 1A). En effet, la cellule eucaryote possède un vrai noyau qui va contenir le matériel génétique sous forme d'ADN entouré par l'enveloppe nucléaire. Le cytoplasme contient le cytosquelette, différents organites, et le cytosol, qui est un milieu semi-liquide où est centralisée la majorité des activités métaboliques de la cellule. Les organites sont des compartiments faits de membrane présentant des fonctions particulières, comme le réticulum endoplasmique où se fait notamment la synthèse et la maturation des protéines sécrétées par la cellule. On compte également des organites comme les mitochondries, l'appareil de Golgi et les lysosomes. De plus, la cellule eucaryote possède un cytosquelette qui permet de donner la forme de la cellule et l'organisation des organites. Le réseau filamenteux formant le cytosquelette est composé notamment de microfilaments et de microtubules (Figure 1A).

Le noyau est délimité par l'enveloppe nucléaire qui contient des pores permettant des échanges entre le noyau et le cytoplasme dans les deux sens. Le noyau n'est défini dans la cellule que lorsque celle-ci n'est pas en division. Associé de ses protéines, l'ADN s'organise en structures appelées chromosomes. En dehors des périodes de divisions cellulaires, les chromosomes ont la forme d'un amas diffus appelé chromatine. Le noyau permet donc de séparer physiquement l'information génétique contenue dans les chromosomes, des réactions chimiques qui se produisent dans le cytoplasme.

Le cytoplasme est constitué du matériel biologique contenu entre la membrane plasmique cellulaire et l'enveloppe nucléaire. C'est un milieu complexe, organisé et dynamique. Au sein du cytoplasme se déroulent en permanence des réactions biochimiques du métabolisme qui permettent le renouvellement constant des composants moléculaires et cellulaires.

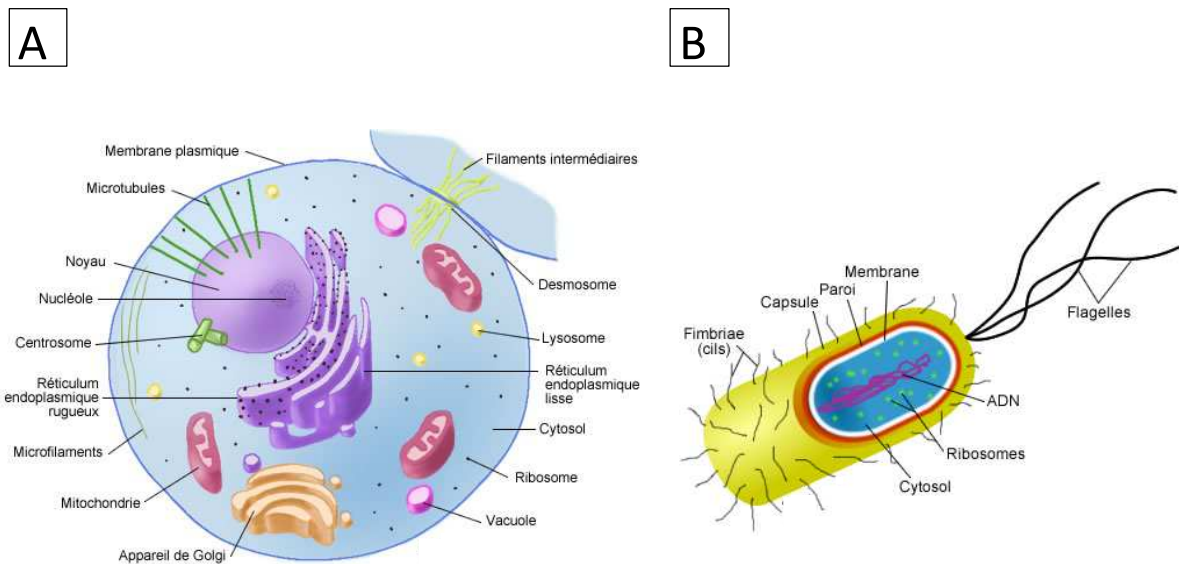


Figure 1 : Organisation structurale de la cellule.

(A) Schéma d'une cellule eucaryote présentant le noyau cellulaire comportant le nucléole et différents organites ainsi que des ribosomes. (B) Schéma d'une cellule procaryote comportant le matériel génétique sous forme d'ADN dans le cytoplasme et les ribosomes.

Grâce au cytosquelette, le cytoplasme permet de maintenir la forme et la cohérence de la cellule. Il joue également un rôle dans le transport des substances se trouvant dans la cellule.

A la grande différence de la cellule eucaryote, la cellule procaryote ne possède pas de noyau mais contient un compartiment unique, le cytoplasme qui comprend l'information génétique sous forme de molécules d'ADN qui peuvent être linéaires ou circulaires (Figure 1B). De ce fait, les processus de réplication, de transcription et de traduction ne sont pas compartimentés et se font donc directement dans le cytoplasme. De plus, les cellules procaryotes ne possèdent ni organites ni cytosquelette.

La compartimentation stricte entre noyau et cytoplasme dans les cellules eucaryotes nécessite la mise en place de phénomènes de transport contrôlés entre le noyau et le cytoplasme. Ce transport s'effectue au travers des pores nucléaires qui perforent l'enveloppe nucléaire.

1.1.2 Les pores nucléaires : structure et transport

Dans les cellules eucaryotes, l'enveloppe nucléaire est une barrière qui permet de séparer le noyau du cytoplasme. Cette enveloppe permet néanmoins des échanges de macromolécules entre le noyau et le cytoplasme à travers des pores nucléaires qui y sont encastrés. Ces pores constituent le seul site d'échange bidirectionnel entre le noyau et le cytoplasme (2, 4–8).

1.1.2.1 Les caractéristiques des pores nucléaires

Un pore nucléaire est formé d'une structure élaborée, le complexe de pore nucléaire (NPC – *Nuclear Pore Complex*). C'est un large complexe perforant la double membrane nucléaire. Il apparaît sous forme cylindrique avec un arrangement selon une symétrie octogonale. De plus, il présente une masse moléculaire pouvant aller jusqu'à plusieurs centaines de mégadaltons selon les espèces. Par ailleurs, l'architecture des complexes de pores nucléaires est conservée suivant les espèces (2, 6–10).

Généralement, les NPC présentent les éléments structuraux suivants : le pore central qui se trouve à la jonction des membranes interne et externe du noyau, les anneaux nucléaires et cytoplasmiques qui sont ancrés au niveau du pore central, le panier nucléaire, et les filaments cytoplasmiques qui sont des éléments périphériques provenant des anneaux nucléaires et cytoplasmiques (Figure 2) (7–13). En réalité, ces filaments sont des domaines intrinsèquement désordonnés, dépourvus de structure secondaire, et hautement flexibles.

Un NPC est formé par l'assemblage de plusieurs copies d'une trentaine de protéines distinctes appelées nucléoporines. Les nucléoporines forment une sous-unité qui se répète à l'identique pour former la structure octogonale des NPC. Le faible nombre de composants protéiques des pores nucléaires comparé à sa masse moléculaire reflète les répétitions et le haut degré de symétrie dans l'assemblage des NPC (7–11, 13).

De manière générale, les nucléoporines contiennent des domaines non globulaires riches en résidus phénylalanines (F) et glycines (G), appelés motifs FG (ces motifs peuvent être du type FxFG, GLFG ou PAFG) (9, 14). Les domaines FG peuvent contenir jusqu'à une cinquantaine de motifs FG. Par ailleurs, les nucléoporines peuvent être classées en deux grands groupes : les nucléoporines structurales et les nucléoporines contenant les domaines FG (FG-Nups). Les nucléoporines structurales forment le support des pores nucléaires et forment des points d'ancrage pour les nucléoporines contenant les domaines FG. Il existe également des nucléoporines transmembranaires. Les FG-Nups varient principalement selon différents aspects : les sites au niveau desquels les FG-Nups sont ancrés dans le NPC et la séquence précise dans laquelle les domaines FG sont intégrés. Les domaines FG retrouvés dans le pore central sont des domaines intrinsèquement désordonnés qui toutefois pourraient engager plusieurs interactions entre eux de manière cohésive formant ainsi les mailles d'un

réseau (9, 11, 15, 16). Un réseau de maille sert de barrière entre le noyau et le cytoplasme, mais permet aussi des échanges de macromolécules entre les deux compartiments (7, 8, 17).

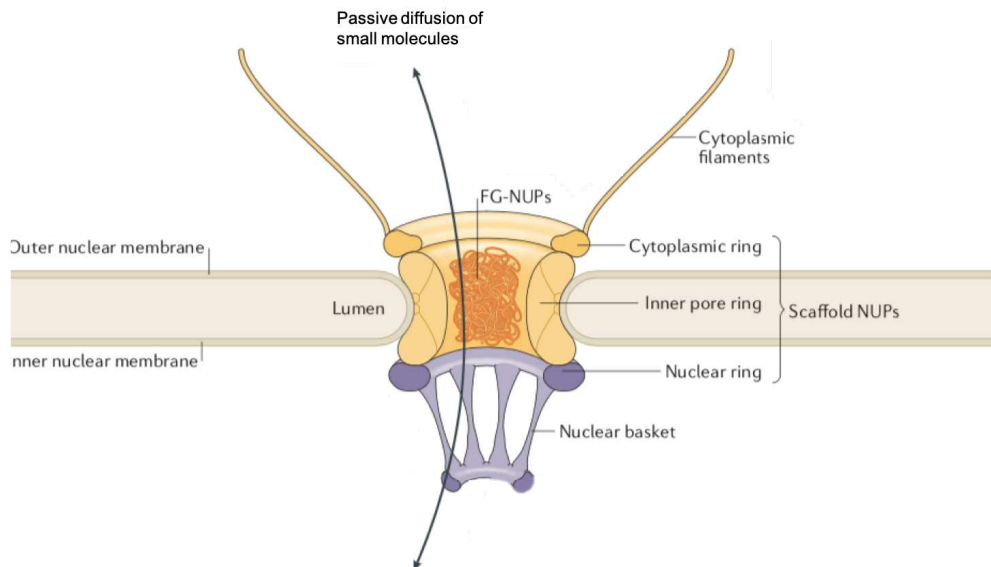


Figure 2 : Caractéristiques structurales des pores nucléaires.

Représentation schématique d'une moitié du complexe de pore nucléaire (NPC) ancré entre les deux membranes externe et interne du noyau (**Outer nuclear membrane and Inner nuclear membrane**). Les NPC sont composés de nucléoporines (**FG-NUPs**). Les éléments structuraux clés du NPC sont l'anneau cytoplasmique (**Cytoplasmic ring**), le pore central (**Inner pore ring**) et l'anneau nucléaire (**Nuclear ring**) qui constituent les nucléoporines de support (**Scaffold Nups**), ainsi que des éléments périphériques comme le panier nucléaire (**Nuclear basket**) et les filaments cytoplasmiques (**Cytoplasmic filaments**). Les NPC permettent le transport de molécules dans les deux directions entre le noyau et le cytoplasme. Ils sont perméables à des molécules de petite taille par diffusion passive. D'après Beck & Hurt (9).

En effet, les nucléoporines contenant les domaines FG au niveau du pore central forment un réseau complexe (Figure 2) dont la nature est encore débattue présentement. Les FG-Nups sont essentiels pour assurer l'effet barrière des NPC (14). En absence de FG-Nups, les NPC perdent leur sélectivité et deviennent ultra perméables (18). De plus, le réseau formé par les FG-Nups au niveau du pore central des NPC constitue une barrière qui exclut les molécules qui ne sont pas destinées pour le transport nucléocytoplasmique. Toutefois, ce réseau permet le passage de petites molécules ($< 20\text{-}40\text{ kDa}$) par diffusion passive, de manière bidirectionnelle entre le noyau et le cytoplasme. En effet, ce réseau possède en son centre un canal de diffusion par lequel la diffusion passive de petites molécules est possible. La diffusion passive devient moins efficace avec l'augmentation de la masse des molécules (autour de $20\text{-}40\text{ kDa}$). Par ailleurs, les

NPC permettent la translocation rapide des cargos préalablement reconnus par des récepteurs de transport nucléaire (NTR – *Nuclear Transport Receptors*) qui ont la capacité de pouvoir interagir avec les FG-Nups présents dans le pore central (9, 19–22). Les cargos transportés par les NTR peuvent être diverses macromolécules comme des protéines ou des ARN.

1.1.2.2 La perméabilité des pores nucléaires

A ce jour, il n'existe pas de consensus sur le mécanisme permettant le passage de la barrière formée par les FG-Nups dans le pore central. Différents modèles ont été proposés pour décrire l'organisation des FG-Nups au niveau du pore central permettant d'assurer la sélectivité du transport nucléocytoplasmique *via* les complexes de pores nucléaires (9, 21). Ainsi, le modèle de l'hydrogel (*selective phase model*) (18), le modèle du *virtual gating* (22), le modèle de réduction de dimension (*reversible collapse*) (23), le modèle du *karyopherin-centric* (24), le modèle *forest* (25) et le modèle du *ring cycle* (26) ont été proposés. Actuellement, seuls les modèles de l'hydrogel et du *virtual gating* résistent de façon assez convaincante aux données expérimentales. Les modèles de l'hydrogel (*selective phase model*) et du *virtual gating* supposent que les FG-Nups sont capables de pouvoir construire un réseau avec des propriétés biophysiques spécifiques permettant le passage des complexes cargo-NTR (Figure 3). Ces deux modèles sont brièvement présentés ci-dessous.

D'un côté, le modèle du *virtual gating* présume que la barrière formée par les FG-Nups qui s'oppose au passage des macromolécules, est une barrière entropique. Considérant ainsi que les macromolécules sont exclues des NPC car elles n'interagissent pas avec les domaines FG. Ces derniers se présentent sous forme de filaments non cohésifs. Cependant, cette barrière se détruit lors du passage des macromolécules uniquement lorsque ces dernières sont associées aux NTR du fait de leur affinité avec les FG-Nups (9, 18, 21, 22). Le complexe cargo-NTR peut ainsi transiter à travers le NPC en se liant progressivement aux FG-Nups. Par contre, ce modèle sous-entend que les domaines FG non cohésifs suffisent pour réaliser la sélectivité des NPC.

D'un autre côté, le modèle de l'hydrogel ou *selective phase model* admet que les interactions entre les domaines FG sont essentielles et doivent se faire de manière complète et cohésive. Il peut alors se former un réseau 3D de type tamis moléculaire qui va se comporter comme une barrière perméable. Ce modèle suppose que les domaines FG forment un hydrogel grâce à des interactions hydrophobes cohésives survenant entre les motifs FG qui constituent ces domaines (18). Le passage des molécules dont la taille est supérieure aux mailles de ce réseau

est exclu, et comme avec le modèle du *virtual gating*, les complexes cargo-NTR peuvent cependant passer la barrière car des interactions sont formées entre les FG-Nups et le NTR pendant le passage de ces complexes (Figure 4A). Ces interactions inhibent les interactions entre les FG-Nups des domaines FG permettant ainsi le passage des complexes. De ce fait, ces complexes peuvent transiter en dissolvant localement le réseau, favorisant ainsi le passage des complexes à travers le NPC (Figure 4B). Ainsi, le réseau de maille formé par les FG-Nups constitue un hydrogel qui se comporte comme une barrière « adaptative » qui se referme autour des complexes cargo-NTR tout en restant cohésif pendant la translocation (18, 21).

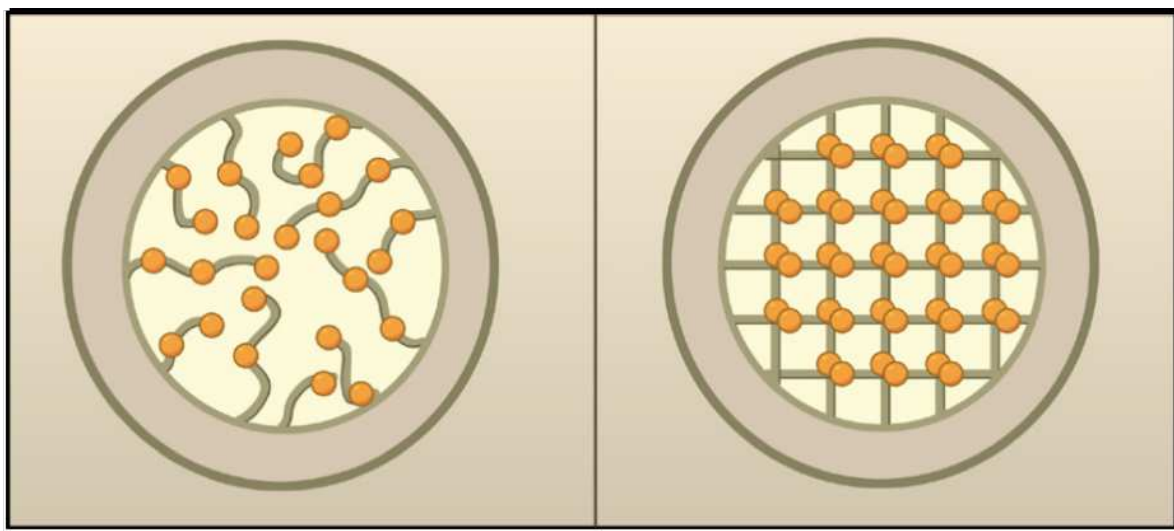


Figure 3 : Schématisation de l'organisation du réseau formé par les domaines FG.

Deux modèles permettant d'expliquer la sélectivité établie par les FG-Nups au sein des complexes de pores nucléaires (NPC). **(A gauche)** Schéma correspondant au modèle du **virtual gating**, les domaines FG sont en forme de filaments, remplissent le pore central et forme une barrière non cohésive. Les motifs FG en orange interagissent avec le récepteur de transport nucléaire en complexe avec le cargo permettant ainsi leur translocation à travers le NPC. **(A droite)** Schéma correspondant au modèle de l'hydrogel, les domaines FG sont en forme de filaments et interagissent de manière cohésive les uns avec autres *via* les motifs FG en orange créant ainsi un réseau en 3D qui exclut les molécules dont la taille est supérieure à celle des mailles du réseau. D'après Weis (21).

De plus, le modèle de l'hydrogel prend en compte le fait que les NTR ont des sites de liaison pour les motifs FG au niveau de leur surface et que ce sont ces motifs FG qui constituent les points de contact pour la formation du réseau (Figure 4). Enfin, le modèle de l'hydrogel est en accord avec des observations expérimentales qui montrent que les interactions cohésives des motifs FG et des FG-Nups sont nécessaires pour maintenir une barrière perméable sélective du NPC, et que les FG-Nups s'auto-assemblent spontanément en hydrogel (18).

Cependant, les domaines FG se présentant comme des filaments ou des polymères en brosse non cohésifs (comme les chaînes de polyéthylène glycol) sont capables de repousser efficacement les macromolécules en surface. Ce qui va dans le sens du modèle du *virtual gating*. Néanmoins, l'hydrogel formé par des domaines FG cohésifs (modèle de l'hydrogel) est plus efficace que les polymères en brosse non cohésifs. Ceci peut être expliqué par le fait que ces polymères en brosse non cohésifs forment une barrière uniquement au niveau de leur point d'ancrage. Plus on s'éloigne de ces points d'ancrage, moins ils constituent une barrière efficace puisqu'ils sont libres et solubles (Figure 3). A l'inverse, l'hydrogel se forme grâce à l'auto-assemblage des domaines FG par des interactions hydrophobes au niveau des motifs FG de manière cohésive (18, 22). Cette cohésion permet de former une barrière efficace indépendamment du point d'ancrage (Figure 4).

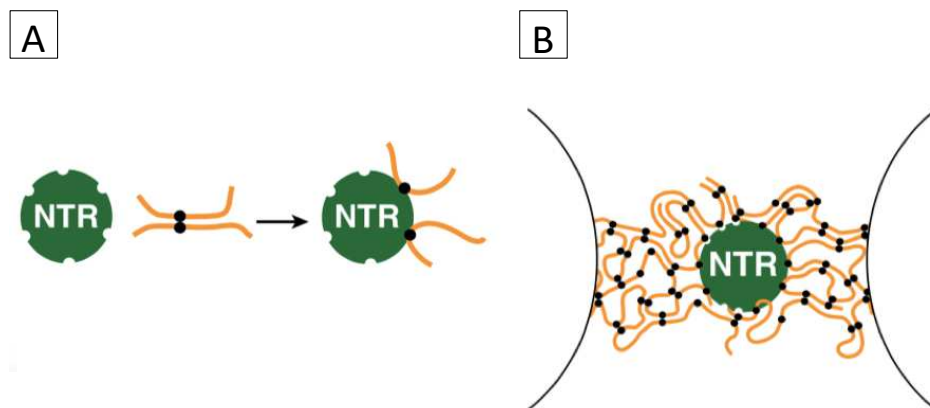


Figure 4 : Diffusion facilitée des macromolécules selon le modèle de l'hydrogel.

(A) Le modèle de l'hydrogel admet que les récepteurs de transport nucléaire (NTR) présentent sur leur surface des points de contact avec les motifs FG (en noir) présents au niveau des FG-Nups (en orange). (B) Le passage des macromolécules au sein du réseau 3D du modèle de l'hydrogel se fait grâce aux interactions entre le NTR et les motifs FG, ainsi le complexe transite en dissolvant localement et progressivement le réseau. D'après Hülsmann *et al* (18).

Ainsi, parmi les différents modèles pouvant expliquer l'organisation du réseau des domaines FG et le passage des molécules par les pores nucléaires c'est le modèle de l'hydrogel (*selective phase model*) caractérisé par des brosses de polymères présentant des interactions hydrophobes cohésives *via* les motifs FG formant l'hydrogel qui semble le mieux décrire les phénomènes de translocation sélective à travers les NPC (9, 21) (Figure 4).

Dans cette sous-partie, nous avons vu que les NTR, en interagissant localement avec les nucléoporines des pores nucléaires permettent le passage des macromolécules dans les deux sens entre le noyau et le cytoplasme. Ce mécanisme en lui-même ne consomme pas d'énergie.

Il fonctionne grâce à l'existence de gradients de concentrations associés à la protéine Ran entre les deux côtés de l'enveloppe nucléaire. L'établissement des gradients de concentrations associés à la protéine Ran consomme de l'énergie et permet d'assurer que la liaison cargo-NTR a bien lieu dans le compartiment souhaité et la dissociation dans l'autre compartiment, assurant ainsi la directionnalité du transport nucléo-cytoplasmique.

1.2 Le cycle de la protéine Ran

1.2.1 Les caractéristiques de la protéine Ran

La protéine Ran est une petite protéine d'environ 25 kDa qui appartient à la superfamille des protéines Ras, une famille de petites protéines G monomériques. Les protéines G sont des GTPases aussi appelées « protéine liant le GTP ». La protéine Ran présente une structure très conservée au cours de l'évolution. En effet, il existe environ 80 % de similarité entre la protéine humaine Ran et son homologue chez les levures (27).

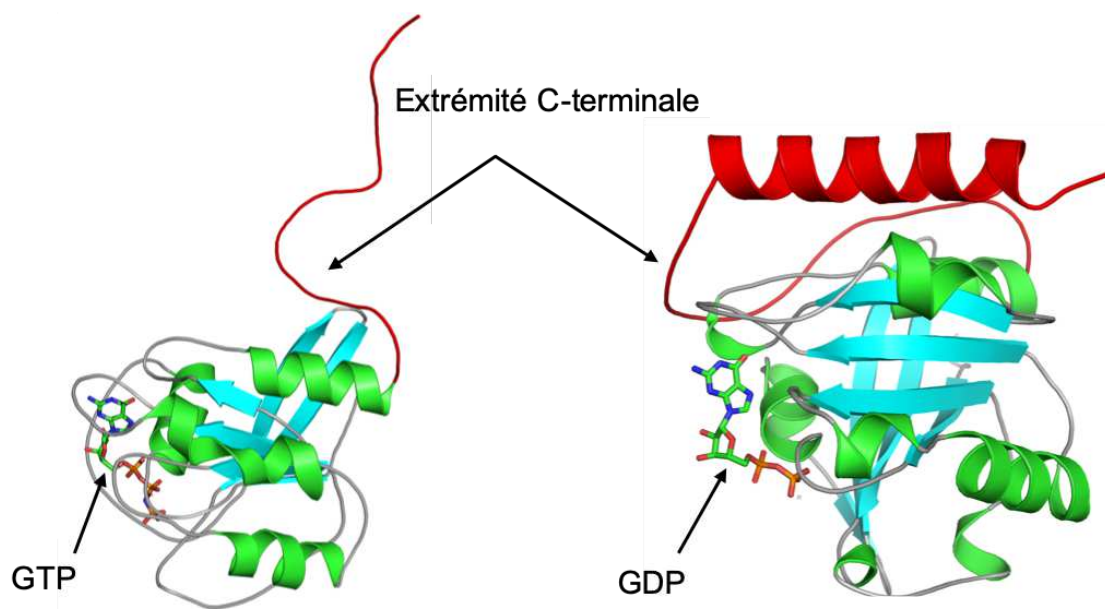


Figure 5 : Représentation de la structure de la protéine Ran.

Représentation en cartoon (code PDB RanGTP : 1QBK et RanGDP : 1BYU) illustrant le changement conformationnel majeur que subit la protéine Ran à la suite de l'hydrolyse du GTP (à gauche) en GDP (à droite). Après l'hydrolyse du GTP en GDP la région non structurée (en rouge) dans l'état RanGTP se replie en une hélice α dans l'état RanGDP au niveau de l'extrémité C-terminale de la protéine Ran.

La protéine Ran présente une activité GTPasique, c'est-à-dire qu'elle est capable de fixer une molécule de GTP puis de l'hydrolyser en GDP. Dans ces cas, elle prend respectivement les

noms de Ran sous sa forme liée au GTP (RanGTP) et Ran sous sa forme liée au GDP (RanGDP) (28–30). Il est important de noter que l'hydrolyse de la molécule de GTP en GDP s'accompagne d'un changement de conformation important de la protéine Ran, qui existe donc sous deux formes structurales correspondant aux deux états nucléotidiques RanGTP et RanGDP (21, 26, 28). La protéine Ran interagit de façon différente avec différents NTR en fonction de son état GTP ou GDP. Suite à l'hydrolyse du GTP en GDP, la protéine Ran subit divers changements de conformation au niveau de sa structure tridimensionnelle. L'un des changements majeurs est le remodelage de l'extrémité C-terminale de Ran (31). Cette région est non structurée lorsque la protéine est sous sa forme RanGTP, et elle se replie en une hélice α après hydrolyse du GTP sous la forme RanGDP (Figure 5). Cette hélice empêche la forme RanGDP de pouvoir se lier aux mêmes substrats que RanGTP (32).

1.2.2 La distribution asymétrique des formes de la protéine Ran

Un point essentiel dans les processus de transport nucléocytoplasmique est qu'il existe une distribution asymétrique des formes de la protéine Ran dans le noyau et le cytoplasme qui permet de contrôler la directionnalité du transport (4, 7, 29, 33–37). En effet, dans la cellule, la protéine RanGTP est présente en grande concentration dans le noyau et en plus faible concentration dans le cytoplasme. A l'inverse, la protéine RanGDP est retrouvée en grande quantité dans le cytoplasme (28, 29, 36, 37). La compartimentation de RanGTP et RanGDP dépend de la localisation des protéines qui régulent les états nucléotidiques GTP/GDP de Ran.

Les cellules contrôlent le cycle de Ran grâce à deux facteurs protéiques (4, 28). D'une part, un facteur d'échange nucléotidique appelé RanGEF (*Ran-Guanine nucleotide Exchange Factor*) ou RCC1 (*Regulator of chromosome Condensation 1*) catalyse la substitution du GDP par le GTP (35). Ce facteur RCC1 est localisé dans le noyau au sein duquel il est associé à la chromatine, et permet donc la formation de RanGTP uniquement dans le noyau. D'autre part, l'hydrolyse de RanGTP nécessite l'aide de la protéine RanGAP (*RanGTPase-activating protein*) (38). En effet, l'hydrolyse du GTP par la protéine Ran est une réaction très lente, le temps de demi-vie du RanGTP est de plusieurs heures sans l'aide de la protéine RanGAP. Cette protéine est localisée dans le cytoplasme et le facteur RanGAP stimule ainsi l'activité GTPasique de Ran et ramène donc la protéine sous sa forme RanGDP dans le cytoplasme (Figure 6) (4, 17, 27, 39). Cette action du facteur RanGAP a lieu du côté cytoplasmique de l'enveloppe nucléaire et nécessite des cofacteurs de type *Ran-Binding Domains* ou RanBD.

Comme l'hydrolyse intrinsèque du GTP par Ran est une réaction lente, l'établissement d'une distribution asymétrique de RanGTP et RanGDP de part et d'autre de l'enveloppe nucléaire est possible par le contrôle de la concentration des facteurs RCC1 et RanGAP dans le noyau et le cytoplasme, respectivement (4, 17, 35, 38, 40). Ainsi la concentration de RanGTP est élevée dans le noyau (environ 4.5 μ M) et faible dans le cytoplasme (environ 8 nM), où la protéine est majoritairement sous sa forme RanGDP (4, 5, 17, 27). C'est cette différence de concentration qui permet de garantir la bonne directionnalité du transport nucléocytoplasmique, comme nous le verrons dans la partie suivante.

Après l'hydrolyse du GTP dans le cytoplasme, la forme RanGDP est importée dans le noyau par un NTR qui lui est dédié, le facteur NTF2 (*Nuclear Transport Factor 2*) (4, 5, 41). NTF2 reconnaît spécifiquement la protéine Ran dans sa forme liée au GDP. Une fois dans le noyau, RanGDP se dissocie du facteur NTF2, et la molécule de GDP liée à la protéine Ran est alors remplacée par une molécule de GTP. Ces deux étapes sont réalisées par le facteur nucléaire RCC1 (Figure 6).

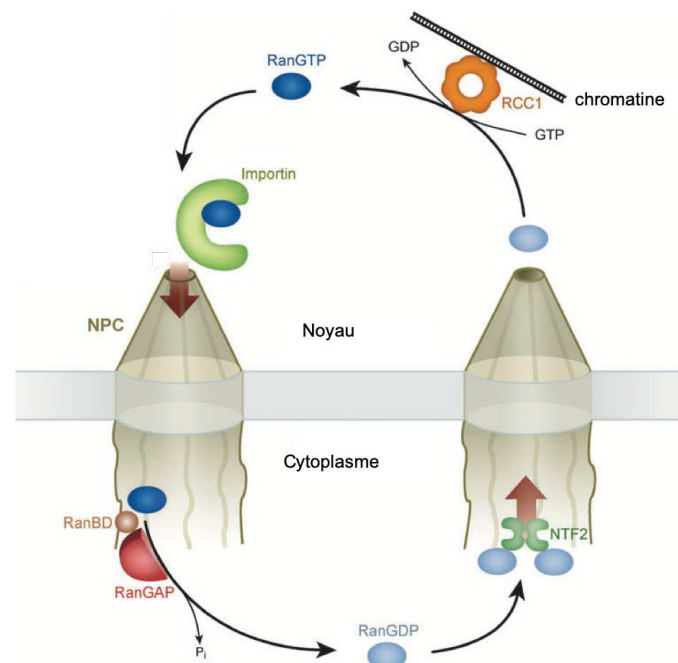


Figure 6 : Représentation schématique du cycle de la protéine Ran.

D'une part, la forte concentration cytoplasmique de la forme RanGDP est maintenue par le facteur RanGAP. Avec l'aide de RanBD, le facteur RanGAP hydrolyse la forme RanGTP entrée dans le cytoplasme en RanGDP ($GTP = GDP + P_i$). D'autre part, la forte concentration nucléaire de RanGTP est maintenue par le facteur RCC1, régulateur de condensation des chromosomes 1, qui permet de substituer la molécule de GDP en GTP de RanGDP qui entre dans le noyau. D'après Cook *et al* (5).

Comme il a été vu dans cette sous-partie, la protéine Ran joue un rôle essentiel dans la directionnalité du transport nucléocytoplasmique grâce à la distribution asymétrique de sa forme RanGTP. Cependant, le transport nucléocytoplasmique est aussi connu pour être spécifique, et cette spécificité est apportée par les NTR. Ces derniers ont un rôle particulier dans la diffusion facilitée des cargos, ils reconnaissent spécifiquement les cargos avant de les transporter d'un compartiment à un autre. L'une des caractéristiques qui définit un NTR est sa capacité à pouvoir interagir avec les pores nucléaires, en particulier avec les motifs FG caractéristiques des nucléoporines permettant ainsi la translocation des cargos au niveau des complexes de pores nucléaires. La majorité des récepteurs de transport nucléaire appartient à une grande famille de protéines : les karyophérines β (4, 5, 17, 28, 42).

1.3 La famille des karyophérines β

Les protéines de la famille des karyophérines β (Kap β) comptent pour la majorité des récepteurs de transport nucléaire impliqués dans le transport nucléocytoplasmique (43). Cette famille comporte des récepteurs d'import nucléaire appelés importines, des récepteurs d'export nucléaire appelés exportines, et également des récepteurs bidirectionnels qui effectuent le transport dans les deux sens (4, 5, 17, 42, 44). Le nombre de protéines des Kap β varie suivant les espèces mais chez l'Homme, on compte une vingtaine de membres dont dix importines, sept exportines et deux récepteurs bidirectionnels bien que ces derniers soient les moins caractérisés de cette famille (Tableau 1). Ce sont des protéines de relativement grande taille, dont le poids moléculaire est autour de 100 kDa.

A côté de la famille des Kap β , il existe aussi celle des karyophérines α (Kap α) qui compte six membres chez l'Homme (17, 45). De manière générale, ce sont des adaptateurs des Kap β , elles servent d'intermédiaire pour la reconnaissance des cargos (17, 28). Par exemple, la karyophérine β également dénommée importine β , qui est l'importine la mieux caractérisée à ce jour, utilise la karyophérine α , également dénommée importine α comme adaptateur (46). Il faut noter que l'importine β est toutefois capable de reconnaître certains de ses cargos sans adaptateur (5, 17, 44).

1.3.1 L'organisation structurale des karyophérines β

Structuralement, les karyophérines β sont caractérisées par une architecture en super-hélice composée d'une répétition successive de motifs HEAT (*Huntingtin*, *Elongation factor 3*,

protein phosphatase 2A, and yeast *p13-kinase TOR1*) (5, 47). Chaque motif HEAT se replie en une paire d'hélices α antiparallèles. Le nombre de HEAT (une vingtaine) varie en fonction des protéines (44, 48, 49). Les motifs s'empilent quasi parallèlement donnant aux protéines une structure en super-hélice. La structure en super-hélice confère à ces protéines deux arches (Figure 7) : une arche N-terminale avec une surface concave interne (en général entre les HEAT 1 à 8) et une arche C-terminale avec également une surface concave interne (en général entre le HEAT 9 et l'extrémité C-terminale) (44, 50). La surface concave de l'arche N-terminale est le site de liaison du RanGTP, et constitue la région la plus conservée au sein de la famille des karyophérines β . La surface concave de l'arche C-terminale correspond au site de liaison des cargos et constitue une région moins conservée. En effet, la conservation des karyophérines β est faible, on observe environ 15-20 % de similarité de séquence au sein des membres de cette famille, ce qui permet aux différents membres de reconnaître des substrats variés et spécifiques (4, 17, 42, 51, 52).

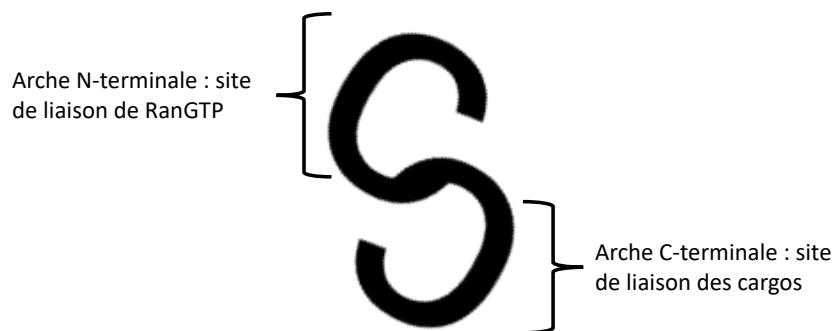


Figure 7 : Schéma simplifié de la structure en super-hélice des karyophérines β .

Grâce à leur composition en HEAT, les karyophérines adoptent une structure en super-hélice qui leur confère deux arches N- et C-terminales. L'arche N-terminale est la région la mieux conservée chez les karyophérines β , c'est le site de liaison du RanGTP. L'arche C-terminale représente la région la moins conservée où se lient les différents cargos. D'après Chook & Günter (51).

Les analyses structurales des membres la famille des karyophérines β en complexe avec différents cargos ont révélé que des molécules distinctes peuvent interagir avec différents sites de l'arche C-terminale d'une même protéine. Ainsi, les karyophérines semblent avoir différents sites de reconnaissance et une large gamme de cargos potentiels. En effet, différentes structures de la karyophérine β (importine β) liée à différents cargos ont suggéré que les NTR sont versatiles dans la reconnaissance de leurs cargos (5, 44, 50). L'architecture en super-hélice des

Kap β leur confère une plasticité intrinsèque qui leur permet d'adopter des conformations complémentaires en fonction des cargos avec un mécanisme de type ajustement induit.

1.3.2 Les fonctions des karyophérines β

Fonctionnellement, les karyophérines β partagent la capacité de pouvoir interagir avec la protéine RanGTP et d'également interagir avec les composants des pores nucléaires *via* les motifs FG des nucléoporines (16, 53, 54). Comme indiqué plus haut, la protéine Ran dans sa forme RanGTP est fortement concentrée dans le noyau, et peu concentrée dans le cytoplasme, et cette différence de concentration détermine la capacité des karyophérines β à se lier ou à se dissocier des cargos qu'elles transportent. Les complexes importines-cargos *via* les signaux de transport nucléaire se forment dans le cytoplasme et se dissocient dans le noyau suite à la liaison de RanGTP. Les exportines se lient aux cargos *via* les signaux de transport nucléaire en présence de RanGTP pour former un complexe cargo-exportine-RanGTP dans le noyau. La libération du cargo se fait suite à l'hydrolyse de RanGTP en RanGDP dans le cytoplasme. De plus, les karyophérines β libres ou liées à des cargos, et avec ou sans RanGTP sont capables de traverser les NPC dans les deux sens (5, 17, 53, 55, 56). La directionnalité est ainsi définie par des réactions d'association/dissociation des complexes cargo-NTR dans le noyau et le cytoplasme, qui sont régulées par le cycle de la protéine Ran. De plus, les karyophérines β reconnaissent spécifiquement les cargos en fonction de leurs signaux de transport nucléaire suivant leur destination, faisant de ce phénomène de transport un processus spécifique.

Les fonctions des kap β dans le transport nucléo-cytoplasmique peuvent s'étendre à d'autres rôles en lien avec la régulation des voies de transport auxquelles elles participent. C'est le cas de la Transportine 1 qui est impliquée dans la voie de signalisation rédox impliquant le facteur de transcription FOXO4 (57). En effet, la Transportine 1 (Tableau 1) est capable de former des ponts disulfures avec le facteur FOXO4 et de l'importer dans le noyau sans interaction avec son signal de localisation en cas de fort stress oxydant.

Par ailleurs, les kap β ont également un rôle dans la régulation des microtubules et la mitose (42). L'importine β peut se comporter comme un inhibiteur direct de protéines essentielles à la polymérisation des microtubules (58). Durant l'interphase, l'importine β permet l'import nucléaire de la protéine d'appariement nucléaire mitotique NuMA. Par contre

lors de la phase de mitose, l'interaction entre l'importine β et la protéine NuMA entraîne l'inhibition de son activité dans la polymérisation micro-tubulaire perturbant ainsi la mitose.

KARYOPHERINES β	CARGOS
IMPORTINES	
IMPORTINE β / KARYOPHERINE β1	Plusieurs cargos connus, des protéines portant des NLS dits classiques reconnus par l'importine α , ...
TRANSPORTINE 1 / KARYOPHERINE β2	Plusieurs cargos connus, des protéines portant des PY-NLS (hnRNP A1) et de protéines portant des NLS autre que des PY-NLS (ADAR1, histones, ...)
TRANSPORTINE SR1	Les protéines SR, ...
TRANSPORTINE SR2	HuR, ...
IMPORTINE 4	Les protéines ribosomiques, les histones, ...
IMPORTINE 5 / KARYOPHERINE β3	Les protéines ribosomiques, les histones, ...
IMPORTINE 9	Les protéines ribosomiques, les histones, ...
IMPORTINE 7	Les protéines ribosomiques, RTC de l'HIV, ...
IMPORTINE 8	SRP19
IMPORTINE 11	UbcM2, rpl12
EXPORTINES	
CMR1 / EXPORTINE 1	Des protéines avec des NES hydrophobiques, riche en leucines.
EXPORTINE-T	tRNA
CAS / EXPORTINE 2	Importine α
EXPORTINE 4	eIF5A, Pho4
EXPORTINE 5	Les précurseurs des micro ARN (pre-miRNA), tRNA, eEF-1A, ILF3, ...
EXPORTINE 6	Actine, profiline
EXPORTINE 7	P50Rho-GAP
IMPORTINES / EXPOTINES	
IMPORTIN 13	Rbm8, Ubc9, Pax6 (import) ; eIF1A (export)
NON CARACTERISEES	
RANBP6	Non définis
RANBP17	Non définis

Tableau 1 : Tableau regroupant les protéines de la famille des karyophérines β .

Chez l'Homme, il existe au moins 20 membres de la famille des karyophérines β . On y compte des importines (orange), des exportines (bleu), des protéines étant à la fois des importines et des exportines (violet), et des protéines qui à ce jour demeurent non caractérisées (vert). D'après Pemberton & Paschal (59).

1.3.3 Les signaux de transport d'import et d'export nucléaire

Les karyophérines β reconnaissent spécifiquement les cargos *via* différents signaux qui déterminent leur déplacement, elles reconnaissent des signaux de localisation nucléaire (NLS) favorisant l'import nucléaire ou des signaux d'export nucléaire (NES) favorisant l'export nucléaire (4, 5, 17, 44, 53, 55, 56).

Les NLS sont en général des séquences d'acides aminés chargés positivement. On distingue les NLS dits « classiques » des NLS « non classiques » (Figure 8). Les NLS classiques regroupent

les NLS monopartites et les NLS bipartites (17, 60, 61). Les NLS monopartites sont de courtes séquences riches en lysines et arginines. Les NLS bipartites sont des clusters de NLS monopartites séparés par une douzaine d'acide aminés. Les NLS dits non classiques n'ont pas de caractéristiques distinctives dans leur séquence d'acides aminés permettant de les décrire par une séquence consensus (55, 62). Toutefois, certains NLS non classiques possèdent des caractéristiques bien définies, comme c'est le cas pour les PY-NLS, qui possèdent un motif hydrophobe conservé proline tyrosine en C-terminal (5, 17, 44, 62).

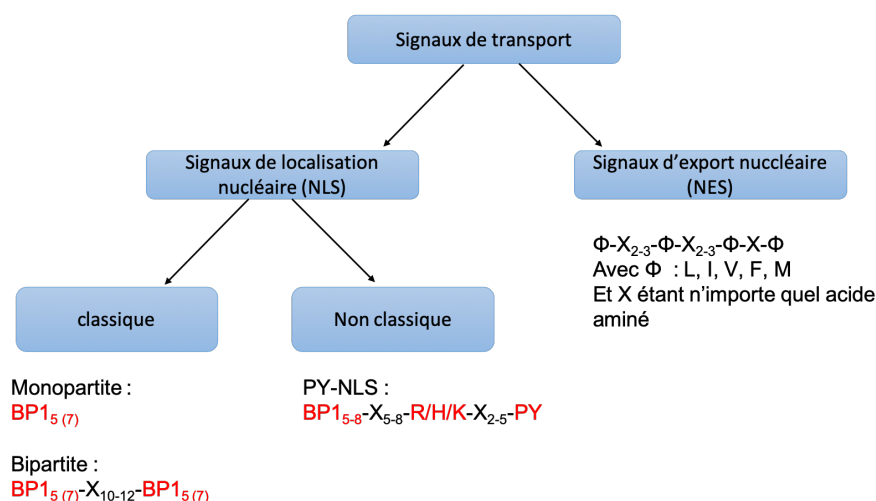


Figure 8 : Classification simplifiée des différents signaux de transport reconnus par les karyophérines β.

Les signaux de localisation nucléaire peuvent être classés comme ceci : les NLS dits classiques avec d'une part les monopartites : de courts peptides ayant entre 5 et 7 acides aminés, et riches en acides aminés chargés, lysines (K), arginines (R) ; et d'autre part les NLS bipartites : des clusters de peptides d'acides aminés basiques séparés par une douzaine d'acide aminé non conservé; les NLS non classiques quant à eux n'ont pas de caractéristiques distinctives dans leur séquence d'acide aminés cependant on peut y classer les PY-NLS, des NLS caractérisés et reconnus par la Transportine 1. Les NES sont des séquences riches en acides aminés encombrants comme les leucines avec une séquence consensus. Ces acides aminés encombrants permettent une structuration particulière du NES qui est essentielle à sa reconnaissance. D'après Fu *et al* (17)

Les NES sont pour la plupart des séquences hydrophobes riches en acides aminés encombrants comme les leucines (63). Ces NES riches en leucine ont une séquence consensus Φ -X₂₋₃- Φ -X₂₋₃- Φ -X- Φ (Avec Φ = L, I, V, F, M et X étant n'importe quel acide aminé) (Figure 8) (17). Les acides aminés encombrants et hydrophobes comme la leucine et l'isoleucine sont essentiels pour la reconnaissance par les NTR, mais également pour maintenir une structuration fonctionnelle. En effet, le contexte du NES et la structuration sont la clé pour le fonctionnement des NES. Par exemple, le NES minimal de la protéine REV du VIH a une plus faible capacité de liaison à son partenaire CRM1 que la protéine entière. Le NLS de REV ne peut se lier à

CRM1 qu'en adoptant une structuration particulière qui nécessite un contexte particulier (17, 62).

1.3.4 Les exportines et l'export nucléaire

Les karyophérines β impliquées dans l'export nucléaire sont les exportines. Les exportines vont reconnaître dans le noyau les NES portés par les cargos à exporter. Cette reconnaissance se fait en présence de RanGTP. Le complexe formé par ces trois éléments est transporté du noyau vers le cytoplasme *via* les NPC. Le passage à travers les NPC est possible grâce à l'affinité des exportines avec les motifs FG des nucléoporines présentes dans le pore central. Une fois dans le cytoplasme, la libération du cargo se fait suite à l'hydrolyse de RanGTP en RanGDP (5, 17, 44). En effet, la forme RanGDP est incapable de se lier aux karyophérines β , et l'hydrolyse du GTP au sein de la forme RanGTP entraîne la dissociation du complexe RanGDP/karyophérines β . Par la suite, le niveau de RanGTP est reconstitué dans le noyau par le ré-import de RanGDP en complexe avec le facteur NTF2 (4, 5, 41), et l'échange de nucléotide de la forme RanGDP par le facteur RCC1 (35, 38) (Figure 6).

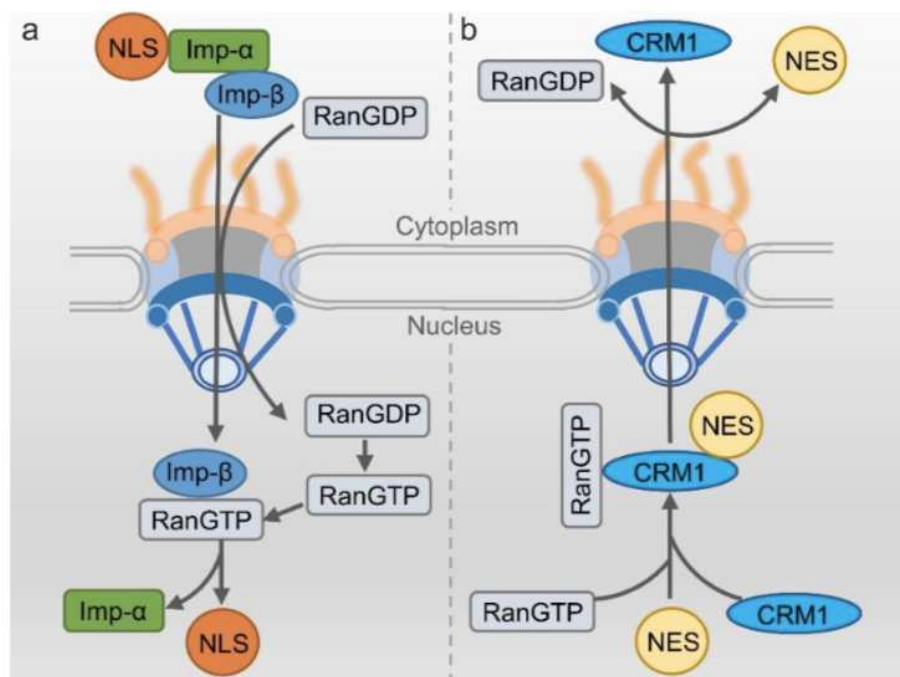


Figure 9 : Schéma simplifié du processus de transport nucléocytoplasmique médié par les karyophérines β .

(a) L'import nucléaire de cargos portant des NLS classiques reconnus par l'importine β *via* l'importine α . (b) L'export nucléaire des NES riche en leucine reconnus par la protéine CRM1 en présence de RanGTP. D'après Fu *et al* (17)

Les exportines reconnaissent principalement leurs cargos *via* les NES riches en leucine (5, 17, 44, 61). Par exemple, l'exportine CRM1 est responsable de l'export d'un grand nombre de protéines présentant des NES hydrophobes riches en leucine (Tableau 1) (64, 65). Dans le noyau, CRM1 se lie dans un premier temps à la protéine RanGTP, s'ensuit la liaison du cargo à exporter *via* le NES. L'ensemble traverse les pores nucléaires grâce aux interactions formées entre la protéine CRM1 et les motifs FG des nucléoporines. Une fois dans le cytoplasme, l'hydrolyse de RanGTP entraîne le relargage du cargo (Figure 9).

1.3.5 Les importines et l'import nucléaire

Les karyophérines impliquées dans l'import nucléaire sont les importines. Les importines reconnaissent leurs cargos de manière directe ou de manière indirecte par le biais d'un adaptateur (5, 17, 44, 51, 53). Le phénomène d'import nucléaire est quasi-identique à celui de l'export nucléaire, excepté que la reconnaissance du cargo se fait *via* les NLS dans le cytoplasme sans RanGTP où sa concentration est plus faible. Et, la libération des cargos importés se fait par la liaison de RanGTP dans le noyau où sa concentration est plus élevée (4, 5, 17, 27, 64). Après l'arrivée dans le noyau et l'association avec RanGTP, le complexe importine-cargo se dissocie et le complexe importine-RanGTP retourne dans le cytoplasme où l'hydrolyse de RanGTP en RanGDP permet la libération de l'importine pour un autre cycle de transport.

Certains récepteurs de transport nucléaire ne peuvent pas reconnaître des substrats sans l'aide d'un adaptateur. Par exemple, l'importine β reconnaît les cargos portant les NLS classiques par le biais de l'importine α (4, 5, 17, 44, 67–69). L'importine α reconnaît le NLS classique porté par le cargo à importer, puis l'importine β se lie à l'importine α *via* son domaine de liaison à l'importine β (IBB – Importin β *binding*) en N-terminal. L'importine β permet le passage du complexe ainsi formé à travers les NPC en interagissant avec les motifs FG des nucléoporines. Le mécanisme d'import dit classique fait référence à ce mécanisme d'import médié par l'importine β en association avec l'importine α (Figure 9). Dans un premier temps, l'importine α reconnaît le NLS porté par le cargo dans le cytoplasme, puis l'ensemble se lie à l'importine β . Le complexe formé par le cargo, l'importine β , et l'importine α est ensuite transporté du cytoplasme vers le noyau *via* les NPC. Dans le noyau, RanGTP se lie à l'importine β ce qui entraîne la dissociation du complexe et donc la libération du cargo. Le complexe importine β /RanGTP peut retourner dans le cytoplasme où RanGTP sera hydrolysé en RanGDP

conduisant à la libération de l'importine β . En parallèle, l'importine α est exportée vers le cytoplasme par un autre NTR de la famille des karyophérines β , l'exportine CAS (70) (Tableau 1).

Dans cette première partie, nous avons vu les principaux acteurs impliqués dans les processus de transport nucléocytoplasmique. C'est un phénomène de diffusion facilitée qui permet des échanges entre le noyau et le cytoplasme. La majorité des voies de ce transport implique les karyophérines β qui sont capables d'effectuer l'import nucléaire des cargos du cytoplasme vers le noyau et l'export nucléaire des cargos du noyau vers le cytoplasme, et ce de manière spécifique. Le sens du transport effectif des cargos est régulé par la petite protéine Ran, qui est présente dans la cellule sous différents états. Seule la forme RanGTP est capable de se lier aux karyophérines β , permettant ainsi d'établir un contrôle de la directionnalité du transport nucléocytoplasmique.

Contrairement au mécanisme d'import dit classique impliquant l'importine β en association avec l'importine α (Figure 9), les mécanismes d'import impliquant certains membres de la famille des karyophérines β sont parfois incomplètement caractérisés. C'est le cas de la Transportine 1, qui reconnaît notamment une classe de NLS non classique, les PY-NLS (Figure 8 et Tableau 1). Même si certains aspects de la reconnaissance des PY-NLS par la Transportine 1 sont bien caractérisés, dans leur ensemble, les bases de la spécificité de reconnaissance de ses cargos par la Transportine 1 sont mal connues. Cette importine fait ainsi l'objet d'études structurales, biochimiques et cellulaires, afin de mieux comprendre les mécanismes moléculaires de son fonctionnement, et les bases moléculaires de la reconnaissance de ses cargos.

2 La Transportine 1

Dans la sous-partie précédente, les différents acteurs du transport nucléaire ont été présentés, notamment les karyophérines β qui représentent la majorité des récepteurs de transport nucléaire impliqués dans le processus de transport nucléocytoplasmique. La Transportine 1 est une protéine qui fait partie des karyophérines β , elle joue un rôle majeur dans le phénomène d'import nucléaire. On lui compte plus d'une trentaine de cargos et elle présente des particularités par rapport aux autres membres de la famille des karyophérines β . Ces particularités seront présentées dans cette partie.

2.1 La Transportine 1 : une importine des karyophérines β

La Transportine 1 (Trn1) est un membre de la famille des karyophérines β (Tableau 1). Comme tout autre NTR, c'est une importine qui reconnaît divers cargos, les nucléoporines des NPC et la forme RanGTP de la petite protéine Ran (2, 4, 5, 17, 44).

Tout d'abord, la Transportine 1 a été identifiée par trois groupes indépendants comme étant le récepteur d'import nucléaire de la protéine hnRNP A1 (heterogeneous nuclear RiboNucleoProtein A1) (71–74). Comme la séquence NLS de hnRNP A1 peut également médier l'export nucléaire de hnRNP A1 (75), *Pollard et al* (71), ont dénommé la nouvelle karyophérine identifiée la Transportine 1, pensant qu'elle pouvait à la fois importer hnRNP A1 dans le noyau et l'exporter vers le cytoplasme. Cette hypothèse a été abandonnée après avoir observé que la liaison de RanGTP à Trn1 entraînait la dissociation du complexe Trn1/hnRNP A1 (36, 76). Cependant, la terminologie Transportine 1 est restée, mais cette protéine est aussi connue sous les noms de karyophérine β 2 ou importine β 2.

Comme les autres membres de la famille des karyophérines β , la Transportine 1 présente une structure en super-hélice comportant deux arches N- et C-terminales successives qui se chevauchent. Malgré des structures quasi identiques, Trn1 présente une similarité de séquence de seulement 24 % avec l'importine β (46, 77–79). Elle possède une plus forte similarité au niveau de l'arche N-terminale où se fait la liaison de RanGTP et diffère au niveau de l'arche C-terminale où se fait la liaison des différents cargos. Elle présente une similarité de séquence de l'ordre de 20 % avec le reste des membres de la famille des karyophérines β . Comme toutes les importines, Trn1 reconnaît ses cargos dans le cytoplasme où la concentration en RanGTP est faible, les importe dans le noyau *via* les pores nucléaires, et les relargue dans le noyau en présence de RanGTP.

De plus, la Transportine 1 est connue pour importer en majorité des protéines qui ont la particularité de lier des ARN, mais aussi des facteurs de transcription (Tableau 2) (44, 80). Effectivement, plus d'une vingtaine de protéines impliquées dans la maturation des ARN messagers ont été identifiées comme étant des cargos de Trn1, notamment les protéines de la famille des hnRNP, comme hnRNP A1, D, F, M (Tableau 2). La Transportine 1 reconnaît directement les NLS portés par ces divers cargos sans adaptateur (5, 53, 80).

Cargos avec motif PY (PY-NLS)	Caractéristiques	Cargos sans motif PY (non PY-NLS)	Séquence du NLS	Expériences
CLK3	bPY	Ad2 adenovirus core protein V	Non défini	P-D, IPCc, IPCk
CPSF6	hPY	Ad2 adenovirus core protein VII	82-198	P-D, IPCc, IPCk
Cycline T1	hPY	ADAR1	dsRBD3 708-801	P-D/MS, CoIP, IPCk, IPCc
EWS	hPY	CD44 (ICD)	671-697	P-D/MS, CoIP, IPCc
FUS	hPY	DDX3	Non défini	P-D/MS,
GLI2, GLI3	hPY	FOS	81-160	P-D/MS, P-D, IPCc, siRNA
HCC1	bPY	FOXO4	Autour de la cystéine Cys239	Non caractérisé
HCMV UL79	hPY	H2A	Non défini	P-D, IPCk
HEXIM1	bPY	H2B	Non défini	P-D, IPCk
hnRNP A1, A2, A3	hPY	H3	11-27	P-D, IPCk
hnRNP AB	hPY	H4	Non défini	P-D, IPCk
hnRNPD	hPY	REV (VIH1)	35-46	IPCk, P-D
hnRNPD L/JKTBP	hPY	E6 (HPV16)	121-151	P-D, IPC
hnRNPF	hPY	L2 (HPV16)	1-13	P-D, IPC
hnRNPH1	hPY	L2 (HPV18)	1-12	P-D, (IPC)
hnRNPM	bPY	HSP70	Non défini	P-D/MS
Huntingtin	bPY	HSP90	Non défini	P-D/MS
HuR	hPY	JMJD5	134-151	CoIP
NXF1	hPY	JUN	250-334	P-D, IPC
PABPN1	bPY	NMP - ALK	Non défini	P-D/MS
PQBP-1	bPY	PLK1	396-433	P-D/MS, P-D
QKI-5	hPY	RPL23A	32-74	P-D, IPCk
RB15B	bPY	RPS7	Non défini	IPCk
SAM68	hPY	RPL5	Non défini	IPCk
SOX14	bPY	RPL7	1-54	PD, siRNA
TAF15	hPY	SRP19	Non défini	P-D/MS, P-D, IPCk
WBS16	hPY	TAF1A / TAF148	400-450	P-D
YBX1	bPY	TCP-1-g	Non défini	P-D/MS, P-D
TAP	hPY	U1 snRNP A	Non défini	IPC
RPL4	bPY	U2 snRNP A	Non défini	IPC

Tableau 2: Tableau regroupant les différents cargos de mammifères de Trn1.

Trn1 reconnaît une large gamme de cargos dont certains portent un motif PY constituant la classe des PY-NLS, mais certains des cargos de Trn1 ne possèdent pas ce motif PY (non-PY-NLS), et donc ne font pas partie de cette classe des PY-NLS. Dans la classification des PY-NLS, on retrouve les PY-NLS dits hydrophobes (hPY) et les PY-NLS dits basiques (bPY). L'interaction de Trn1 avec les non-PY-NLS a été mise en évidence par différentes techniques pour la plupart : P-D (test de pull-down), P-D/MS (test de pull down couplé à la spectrométrie de masse), IPCk (test d'import par Trn1), IPCc (test de compétition d'import avec un substrat bien connu), CoIP (co-immunoprécipitation). D'après Twyffels *et al* (80).

2.2 L'organisation structurale de la Transportine 1

Au niveau structural, comme toutes les karyophérines β , la Transportine 1 est composée de plusieurs motifs HEAT qui lui donnent cette structure en super-hélice (Figure 10). La première structure de Trn1 a été résolue par *Chook & Blobel* (65) il y a une vingtaine d'années. Il s'agit de la structure de Trn1 en complexe avec un analogue non hydrolysable de RanGTP (RanGppNHp). Elle a permis de révéler des particularités structurales de Trn1. Depuis, plusieurs autres structures de la Transportine 1 libre ou en complexe avec divers substrats ont également été résolues. La structure de Trn1 libre réalisée en utilisant la cristallographie par *Cansizoglu & Chook* (81) a permis de montrer que Trn1 se compose de 20 motifs HEAT. Les motifs HEAT sont connectés les uns aux autres grâce à de petites boucles ou de petites hélices sauf au niveau du HEAT H8, où on retrouve une grande boucle acide d'environ une soixantaine d'acides aminés (dont 27 Asp/Glu sur 65 résidus) (62), appelée boucle H8 qui permet de connecter les deux hélices du HEAT H8. Cette boucle H8 est conservée au sein des Karyophérines β mais elle est plus longue chez Trn1 (65 résidus chez Trn1 contre 12 résidus chez l'importine- β) (5, 46, 52, 79, 82). La Transportine 1 s'organise donc en deux arches orthogonales successives qui se chevauchent. L'arche N-terminale comprend les HEAT H1 à H13 et l'arche C-terminal les HEAT H9 à H20 (Figure 10) (5, 47, 79, 83). La résolution des différentes structures de Trn1 a permis de mieux comprendre son fonctionnement dans les mécanismes d'import notamment grâce aux structures de Trn1 en complexe avec RanGTP ou en complexe avec différents NLS de ses cargos (Tableau 3).

De plus, ces différentes structures ont permis de mieux appréhender son organisation structurale. En utilisant différentes techniques de superposition de structures, *Cansizoglu et Chook* (81), ont pu identifier trois segments majeurs de la super-hélice de Trn1 pour décrire sa flexibilité intrinsèque, lui permettant de s'adapter en fonction des différents cargos qu'elle reconnaît. Cette description a aussi été faite pour l'importine β (50, 81, 84). D'un point de vue structural, Trn1 peut ainsi être divisée en trois grands segments (81): les deux premiers segments correspondent aux régions comprises entre les HEAT H8-H13 et H14-H18. Ces deux segments constituent le site de liaison des différents substrats de Trn1 au niveau de la face concave de l'arche C-terminale de la protéine. Ils sont respectivement appelés site A et site B (Figure 10). Le troisième segment correspond à la région formée par les HEAT H1 à H8, et constitue le site de liaison de RanGTP au niveau de la surface concave de l'arche N-terminale. Ce troisième segment (HEAT H1-H8) peut être sous-divisé en deux segments plus petits. Le

segment formé par les HEAT H1-H4 qui est intrinsèquement flexible. Et le segment formé par les HEAT H5-H8 possède une conformation qui change uniquement suite à la liaison de RanGTP (81, 84).

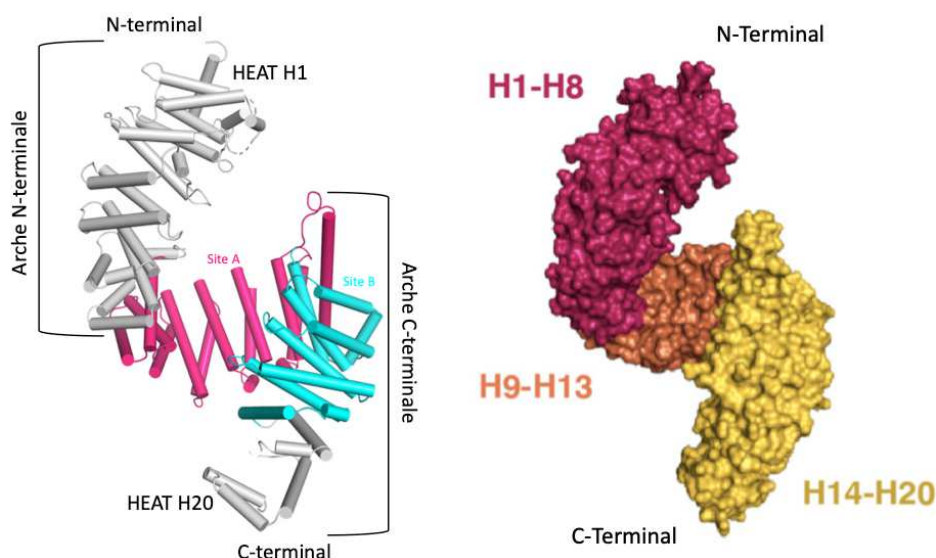


Figure 10 : Organisation structurale de la Transportine 1.

(**Gauche**) Représentation en cartoon (code PDB 2H4M) montrant la structure en super-hélice de la Transportine 1 (Trn1). Trn1 se compose d’une vingtaine de motifs HEAT qui s’assemblent de manière antiparallèle les uns par rapport aux autres lui conférant une structure en super-hélice. Cette structure en super-hélice de Trn1 s’arrange en deux arches successives, l’arche N-terminale où se fait la reconnaissance de RanGTP et l’arche C-terminale où sont reconnus les différents cargos. L’arche C-terminale est constituée des sites A (en fuschia) et B (en cyan) qui représentent la surface d’interaction avec les PY-NLS. (**Droite**) Représentation en surface (code PDB 2H4M) montrant les trois régions qui permettent de décrire la flexibilité conformationnelle de Trn1. La région comprenant les HEAT H1-H8 est en rouge, H9-H13 en orange et H14-H20 en jaune.

Ces études ont également montré une hétérogénéité conformationnelle entre différentes molécules de Trn1 au sein du cristal (81). La Transportine 1 possède ainsi une flexibilité intrinsèque, du fait de son organisation structurale basée sur la répétition des motifs HEAT. En effet, la répétition des motifs HEAT et leur organisation permettent de produire un effet « ressort » sur la structure de Trn1. En fonction de la structure considérée, notamment en fonction de son état libre ou en complexe avec des cargos, Trn1 est plus ou moins étirée au niveau de ses extrémités N- ou C-terminales, ce qui donne *in fine* une structure plus ou moins allongée. Dans cette étude, les modifications les plus marquées se situent au niveau de l’arche N-terminale (HEAT H1-H8), et la région centrale (HEAT H9-H13) est dans la majorité des cas plus homogène. Cependant, dans une autre analyse structurale, le segment formé par les HEAT

H14-H20 apparaît comme celui présentant la plus grande flexibilité intrinsèque (75, 79, 82) (Figure 10).

2.3 La reconnaissance des partenaires de la Transportine 1

La Transportine 1 pour effectuer le mécanisme de transport nucléaire, doit pouvoir se lier à différents partenaires pour importer un cargo dans le noyau. Une fois le cargo reconnu dans le cytoplasme, Trn1 doit transiter les NPC en interagissant avec les FG-Nups qui les constituent. De plus, la libération du cargo n'est possible qu'après fixation de RanGTP dans le noyau.

2.3.1 La reconnaissance de RanGTP par la Transportine 1

La structure résolue par *Cansizoglu et Chook* (81) a permis de montrer que la liaison de RanGTP se fait au niveau de l'arche N-terminale et plus précisément au niveau des segments comprenant les HEAT H1-H4 et H7-H8 mais également au niveau de la grande boucle acide H8 (Figure 12A) (79). Cette liaison entre Trn1 et RanGTP implique majoritairement des interactions hydrophobes (5, 44).

Les régions « *switch I* » (résidus 30-47) et « *switch II* » (résidus 65-80) de RanGTP, qui ont la particularité de subir des changements conformationnels lors du passage entre les différents états nucléotidiques de la protéine Ran (5, 31, 85), sont reconnues au niveau du segment H1- H8 de Trn1 (Figure 11B) (44, 79, 86). Plus précisément, ces régions sont reconnues au niveau du petit segment H1-H4. Comme observé avec l'importine β , la région basique (139-142) située à l'opposé des régions « *switch* » est impliquée dans des interactions électrostatiques avec le HEAT H7 (66). De plus, des résidus basiques retrouvés au sein du domaine G de RanGTP forment des interactions électrostatiques avec les chaînes latérales des résidus polaires chargés de la boucle H8, mais également des interactions hydrophobes avec les chaînes latérales des résidus apolaires et aromatiques de la boucle H8. Environ un tiers des résidus de la boucle H8 ont des contacts directs avec RanGTP (44, 79).

De façon intéressante, l'organisation structurale de Trn1 en complexe avec RanGTP a montré une structure en super-hélice moins allongée que lorsque Trn1 est libre ou en complexe avec des cargos. Toutefois, les changements observés suite à la liaison de RanGTP restent mineurs. La liaison de RanGTP à Trn1 entraîne des changements de conformation au niveau de la boucle acide H8. Cette dernière qui est quasi entièrement désordonnée dans la structure de Trn1 libre, devient structurée suite à la liaison de RanGTP (79).

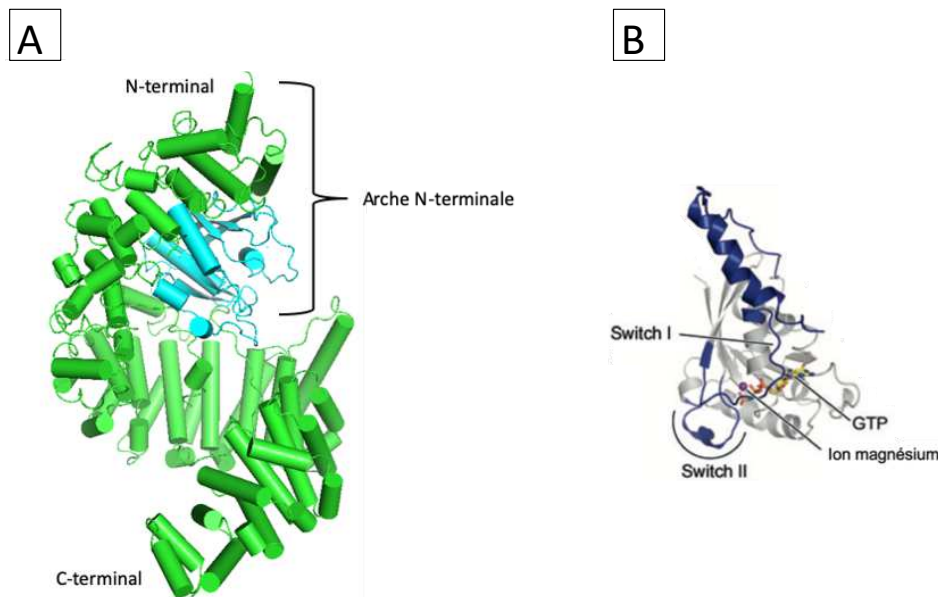


Figure 11 : Reconnaissance de RanGTP par Trn1.

(A) Représentation en cartoon illustrant l'interaction entre la Transportine 1 (Trn1) (en vert) et RanGTP (en cyan), code PDB : 1QBK. La liaison de RanGTP se fait dans la surface concave de l'arche N-terminale de Trn1. Cette liaison implique de manière générale des interactions hydrophobes entre le domaine G de RanGTP et l'arche N-terminale de Trn1. Toutefois des interactions électrostatiques sont également observées entre des résidus basiques du domaine G de RanGTP et des résidus acides de la boucle acide H8 de Trn1. (B) Représentation en cartoon de RanGTP illustrant les régions d'interaction (en bleu) de RanGTP avec les karyophérines β dont Trn1. Les boucles comprenant les régions « switch I » et « switch II » de la protéine Ran sont les régions principales d'interaction avec les karyophérines β . Ces boucles ont la particularité de subir des changements conformationnels lors du passage entre les différents états nucléotidiques RanGTP/RanGDP. D'après Cook *et al* (5).

La dissociation du cargo de Trn1 se fait dans le noyau suite à la fixation de RanGTP (86). Et, c'est l'interaction entre le domaine G de la protéine RanGTP et la boucle H8 qui en est responsable. En effet, des changements surviennent dans l'orientation des hélices des HEAT H1 à H8 lors de la fixation de RanGTP au niveau de l'arche N-terminale. La boucle H8 est alors déplacée en direction de l'arche C-terminale. Le déplacement de la boucle H8 en direction du site de reconnaissance des NLS des cargos entraînerait la dissociation du cargo (Figure 12) (77). L'interaction de la boucle H8 avec le site de liaison des cargos est dépendante de la présence de RanGTP. En effet, en absence de RanGTP, l'interaction de la boucle H8 qui a une charge globale négative n'est pas favorable avec l'arche C-terminale. En revanche, en présence de RanGTP, l'interaction des résidus basiques du domaine G de RanGTP avec les résidus acides de la boucle H8 pourrait perturber les interactions électrostatiques entre Trn1 et le NLS du cargo. Ce qui entraînerait la dissociation du cargo, et donc son relargage (62, 77, 79, 86).

Il est important de noter que, contrairement à l'importine β , la boucle H8 a toujours été considérée comme ne participant pas à la reconnaissance des NLS par Trn1 (77). En effet, plusieurs structures impliquant des mutants de Trn1 pour lesquels la boucle H8 a été délétée ont pu être résolues en complexe avec différents NLS. Ceci montre que Trn1 délétée de la boucle H8 est toujours capable de se lier à ses substrats. Pour certaines structures, il avait été nécessaire d'enlever cette longue boucle flexible pour arriver à cristalliser Trn1 en complexe avec des substrats. C'est le cas par exemple de la première structure de Trn1 résolue en complexe avec un NLS substrat, le NLS de hnRNP A1 (62).

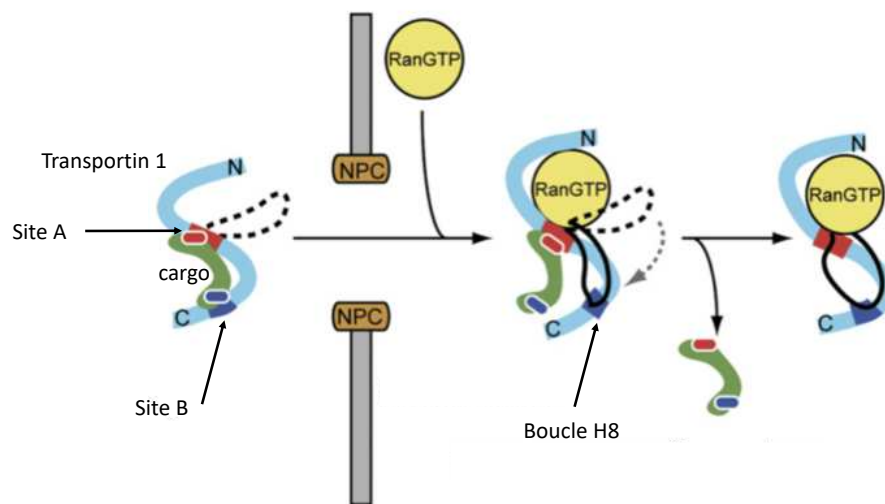


Figure 12 : Dissociation des cargos de Trn1.

Schéma illustrant le mécanisme de dissociation des cargos de la Transportine 1 (Trn1) qui implique la boucle acide H8. Une fois le complexe Trn1/cargo arrivé dans le noyau, la liaison de RanGTP au niveau de l'arche N-terminale crée des interactions avec la boucle H8 qui change de conformation et se déplace en direction du site de liaison des cargos au niveau de l'arche C-terminale. Ce qui favorise la dissociation du cargo, et donc son relargage dans le noyau. Ce relargage débute avec la dissociation du NLS au niveau du site B (HEAT H14-H18) et le relargage total du cargo survient suite à la dissociation du NLS au niveau du site A (HEAT H8-H13). D'après Imasaki *et al* (77).

2.3.2 La reconnaissance des différents substrats de la Transportine 1

La Transportine 1 reconnaît en majorité des protéines se liant aux acides nucléiques notamment des protéines impliquées dans la maturation des ARN (62). Toutefois elle compte une large gamme de cargos allant des protéines ribosomiques aux facteurs de transcription (Tableau 2). Elle avait d'abord été identifiée comme étant le récepteur d'import nucléaire de la protéine hnRNP A1. Depuis, plusieurs autres de ses substrats ont été révélés notamment les

ribonucléoprotéines D, M ou JKTBP, la protéine REV du VIH1 mais aussi le facteur de transcription HuR et bien d'autres (80) (Tableau 2).

2.3.2.1 La reconnaissance des PY-NLS par la Transportine 1

Contrairement à l'importine β , la Transportine 1 ne reconnaît pas les NLS dits classiques mais une autre classe de NLS appelée les PY-NLS (62, 77). C'est une classe de NLS qui est aujourd'hui bien caractérisée et dont le mécanisme de reconnaissance par Trn1 est bien connu. Les PY-NLS sont des peptides qui comptent entre 20 et 30 acides aminés avec un désordre structural intrinsèque. Ils sont également caractérisés par une charge globale positive, et une faible conservation de séquence excepté un ensemble de motifs hydrophobes ou basiques faiblement conservés en N-terminal et un motif PY très conservé en C-terminal pour proline (P) tyrosine (Y) (Figure 8), d'où l'appellation de PY-NLS.

2.3.2.1.1 Les surfaces d'interactions entre la Transportine 1 et les PY-NLS

Toutes les structures de Trn1 en complexe avec des PY-NLS ont montré que les PY-NLS étaient reconnus au sein du même site de fixation, au niveau de la face concave interne de l'arche C-terminale de Trn1. Néanmoins, les PY-NLS peuvent être classés en deux groupes, les PY-NLS hydrophobes et les PY-NLS basiques (62) (Tableau 2). Les PY-NLS hydrophobes présentent plusieurs résidus hydrophobes au niveau de leur extrémité N-terminale, tandis que les PY-NLS basiques vont présenter plusieurs résidus basiques au niveau de leur extrémité N-terminale. La Transportine 1 est ainsi capable de reconnaître ces extrémités N-terminales hydrophobes ou basiques au niveau du même site de fixation. En effet, la reconnaissance des PY-NLS par Trn1 est d'une grande variabilité. L'arche C-terminale de Trn1, à laquelle se lie les PY-NLS, est riche en acides aminés chargés négativement et contient aussi des résidus hydrophobes, lui permettant de reconnaître un grand nombre de séquences de NLS différentes (44, 62, 77).

Au sein de la face concave de l'arche C-terminale de Trn1, le NLS porté par le cargo est reconnu dans le sens opposé à celui de la super-hélice de Trn1. Les différentes structures de Trn1 avec des PY-NLS ont montré que leur reconnaissance se fait au niveau d'un site de liaison commun qui peut être divisé en deux sous-sites : le site A au niveau des HEAT H9-H13 et le site B au niveau des HEAT H14-H18 (Figure 13).

Le motif PY se trouvant au niveau de l'extrémité C-terminale des PY-NLS est reconnu au niveau du site A de Trn1. Peu importe la nature du PY-NLS, hydrophobe ou basique, l'extrémité N-terminale est reconnue au niveau du site B de Trn1. Néanmoins, l'interaction du motif PY au niveau du site A est essentielle à la reconnaissance du NLS par Trn1 et l'interaction au niveau du site B permet de renforcer l'affinité du NLS pour Trn1 (44, 62, 77).

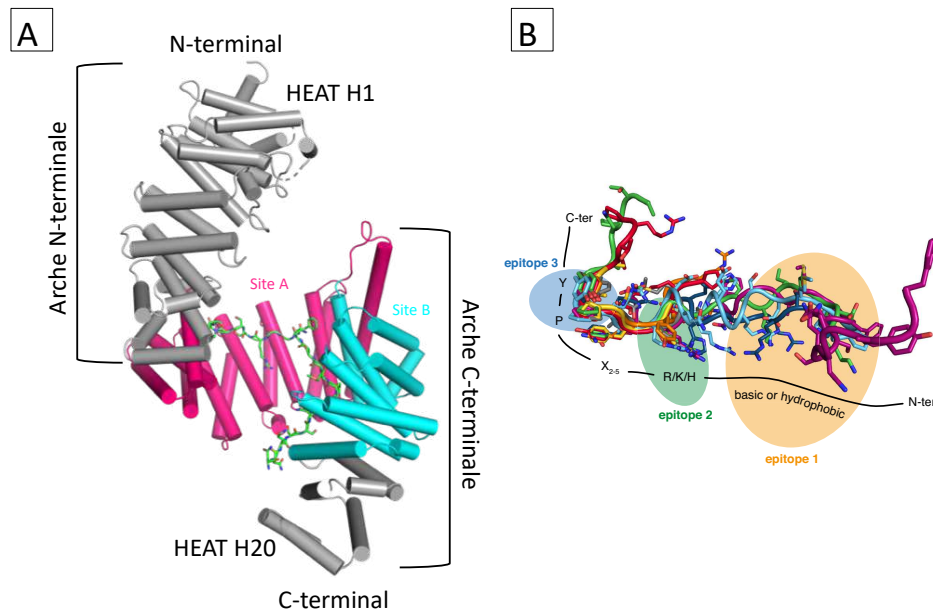


Figure 13 : Surfaces d'interaction des PY-NLS.

(A) Représentation illustrant l'interaction entre la Transportine 1 (Trn1) et le NLS (en vert) de hnRNP A1 (code PDB : 2H4M). La liaison du NLS se fait dans la surface concave de l'arche C-terminale de Trn1. Cette liaison implique de manière générale des interactions hydrophobes entre le motif PY du NLS et le site A (en rose) de l'arche C-terminale de Trn1. L'extrémité N-terminale du NLS est reconnue au niveau du site B (en cyan) de l'arche C-terminale de Trn1. La liaison du NLS se fait dans le sens opposé à celui de la super-hélice de Trn1. (B) Superposition des PY-NLS en complexe avec Trn1 (non représentée ici) qui adoptent tous la même conformation dans la poche d'interaction. Les régions de contact au niveau des différents PY-NLS (hnRNP A1 en rose, FUS en cyan, hnRNP M en vert, hnRNP D en bleu, TAP en rouge, HCC1 en orange, RPL4 en gris, Nab2 en jaune) sont appelées les épitopes 1 (en orange), 2 (en vert) et 3 (en bleu).

A ce jour, plusieurs structures de la Transportine 1 avec différents PY-NLS ont été résolues par cristallographie aux rayons X (Figure 14). Les différentes structures de Trn1 en complexe avec les PY-NLS retrouvées sur la PDB correspondent aux PY-NLS suivants : les protéines hnRNP A1 (2H4M) (62), hnRNP M (2OT8) (87), TAP (2Z5K, 2Z5M) (77), hnRNP D (2Z5N) (77), JKTB (2Z5O) (77), Nab2 (4JLQ) (88), FUS (4FDD, 4FQ3, 5YVG, 5YVI, 5YVH) (89) (90), HCC1 (4OO6) (44), RPL4 (5TQC) (91) (Figure 13).

2.3.2.1.2 La reconnaissance des PY-NLS hydrophobes

On distingue les structures de Trn1 en complexe avec les PY-NLS hydrophobes, qui correspondent aux structures suivantes : les protéines hnRNP A1 (62), hnRNP D (77), et JKTBP (77), le facteur d'export nucléaire TAP (77), et la protéine FUS (89). De manière générale, pour les PY-NLS dits hydrophobes, les interactions se font au niveau du site B (HEAT H14-H18) pour les motifs hydrophobes en N-terminal, et au niveau du site A (HEAT H9-H13) pour le motif PY en C-terminal (44) (Figure 13) (Tableau 3).

Dans la structure de Trn1 en complexe avec hnRNP A1 (62), le NLS se lie au niveau du segment de Trn1 comprenant les HEAT H8-H20, et occupe ainsi une grande partie de l'arche C-terminale avec des interactions au niveau du site A et du site B (Figure 13). Dans cette structure, le NLS se présente comme un peptide linéaire sans structure secondaire. Deux motifs du NLS de hnRNP A1 forment des interactions indispensables à la liaison de Trn1 : le motif PY en C-terminal et le motif hydrophobe en N-terminal. La recherche de ce type de motifs et d'interactions potentielles par des méthodes bio-informatiques a permis d'identifier d'autres PY-NLS pouvant être reconnus par Trn1 (62).

	PY-NLS	Type	Zone d'interaction	Résolution (Å)	Année	Code PDB	Refs.
Transportine 1	hnRNP A1 (M9) NLS	hPY	HEAT H8-H20	3.05	2006	2H4M	(62)
Transportine 1	hnRNP M NLS	bPY	HEAT H8-H16	3.10	2007	2OT8	(87)
Transportine 1	hnRNP D NLS	hPY	HEAT H8-H18	3.20	2007	2Z5N	(77)
Transportine 1	TAP NLS	hPY	HEAT H8-H13	2.60	2007	2Z5K	(77)
Transportine 1	JKTB NLS	hPY	HEAT H8-H3	3.20	2007	2Z5O	(77)
Transportine 1	FUS NLS	hPY	HEAT H8-H18	2.30	2012	4FDD	(89)
Transportine 1	NAB 2 NLS	bPY	HEAT H8-H12	3.05	2013	4JLQ	(88)
Transportine 1	HCC1 NLS	bPY	HEAT H8-H14	2.70	2014	4OO6	(44)
Transportine 1	RPL4 NLS	bPY	HEAT H8-H14	3.00	2017	5TQC	(91)
Transportine 1	FUS NLS	bPY	HEAT H8-H16	4.05	2018	5YVG	(90)

Tableau 3 : Tableau regroupant les structures cristallographiques de complexe Trn1 et différents PY-NLS.

Les différentes structures retrouvées sur la PDB de la Transportine 1 en complexe avec des PY-NLS hydrophobes (hPY) et basiques (bPY). Ces différentes structures ont été résolues par cristallographie aux rayons X (la meilleure résolution étant de 2.30 Å).

En ce qui concerne la structure de Trn1 en complexe avec le NLS de hnRNP D (77), les interactions ont lieu au niveau du segment correspondant aux HEAT H8-H18 alors que les NLS des protéines TAP et JKTBP sont reconnus au niveau du segment H8-H13 (Figure 14). De façon similaire, le NLS de la protéine FUS occupe l'arche C-terminale de Trn1 au niveau des sites A et B. Mais contrairement aux autres PY-NLS qui adoptent des conformations étendues sans structure secondaire, la partie centrale du NLS de FUS forme une petite hélice de 2,5 tours. Le NLS de FUS interagit avec Trn1 *via* des interactions hydrophobes au niveau de ses extrémités N- et C-terminales, et par des interactions électrostatiques au niveau de sa partie centrale structurée (89).

2.3.2.1.3 La reconnaissance des PY-NLS basiques

Les structures de Trn1 en complexe avec les PY-NLS basiques (Tableau 3) comprennent les structures suivantes : les protéines hnRNP M (87), Nab2 de la levure *Saccharomyces cerevisiae* (88), la protéine HCC1 (44), et la protéine RLP4 (91).

Concernant la structure du complexe de Trn1 avec la protéine Nab2, bien que Nab2 porte un motif PL à la place du motif PY, le NLS de Nab2 est également reconnu au niveau de l'arche C-terminale de Trn1 exactement comme sont reconnus les PY-NLS possédant le motif PY (Figure 14). Le motif PL est ainsi reconnu comme un variant du motif PY, et même si le NLS de Nab2 ne comprend pas strictement le dipeptide PY, il est tout de même classé dans la famille des PY-NLS (88).

Le complexe formé par Trn1 et la protéine HCC1 (44) montre que le NLS de la protéine HCC1 est reconnu au niveau du segment comprenant les HEAT H8-H14, interagissant bien au niveau des sites A et B (Figure 14). Par contre le NLS de la protéine hnRNP M (87) est reconnu au niveau du segment comprenant les HEAT H8-H16 alors que la protéine hnRNP A1 occupe le segment comprenant les HEAT H8-H20. Cette différence est due à la différence de taille entre les deux NLS. En effet, contrairement à la protéine hnRNP A1 dont le NLS occupe toute l'arche C-terminale de Trn1, la protéine hnRNP M n'occupe qu'une partie de cette arche. Ceci est donc dû à la différence de taille qu'il existe entre les deux NLS (Figure 15).

La structure la plus récente de Trn1 en complexe avec un PY-NLS est la structure avec la protéine RPL4 (91). Comme pour les précédents PY-NLS, le PY-NLS de la protéine RPL4 est reconnu au niveau de la même poche d'interaction avec des interactions au niveau des différents sites A et B, et plus précisément au niveau du segment comprenant les HEAT H8-H14.

De manière générale, la Transportine 1 reconnaît le motif PY des PY-NLS avec les mêmes résidus au niveau de son site A (comprenant les HEAT H8-H13). Pour les motifs hydrophobes ou basiques en N-terminal la reconnaissance se fait de manière plus variée (44, 62, 77).

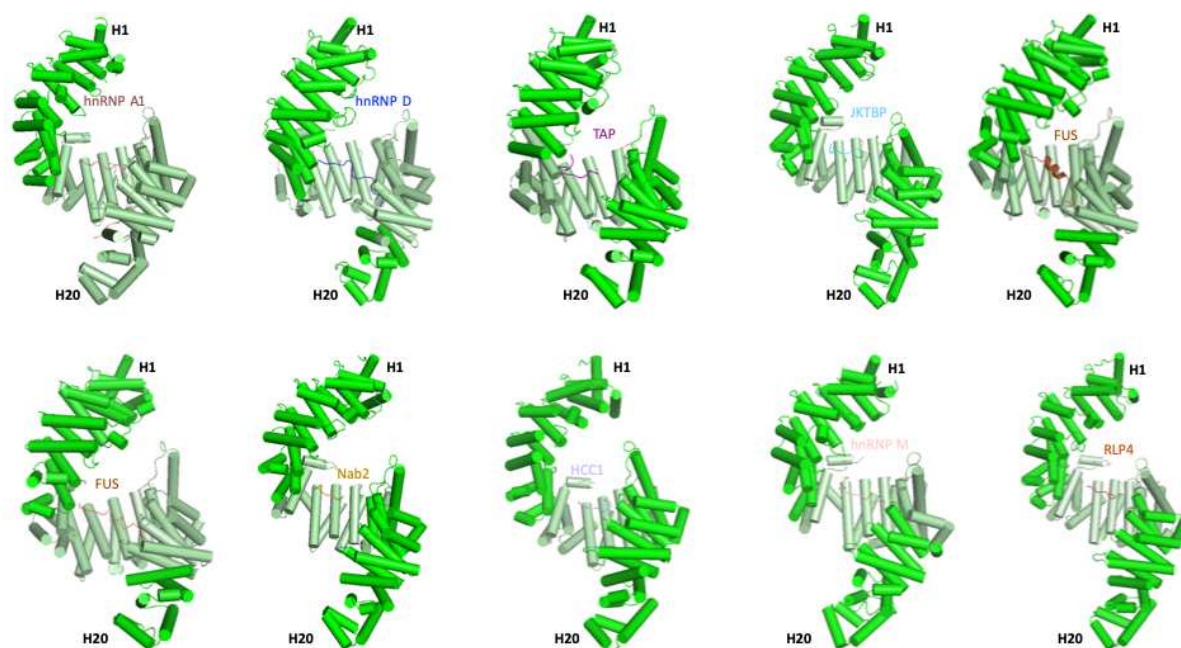


Figure 14 : Structures cristallographiques de la reconnaissance des PY-NLS par Trn1.

Représentation en cartoon illustrant les différentes structures de la Transportine 1 (en vert) en complexe avec les PY-NLS (linéaire) disponibles sur la PDB. Les complexes représentés avec leur code PDB : hnRNP A1 (2H4M), hnRNP D (2Z5N), TAP (2Z5K), JKTBP (2Z5O), FUS (4FDD) avec / sans l'hélice, Nab2 (4JLQ), HCC1 (4OO6), hnRNP M (2TO8), RPL4 (5TQC). D'après Christie *et al* (44).

2.3.2.1.4 Les règles de reconnaissance des PY-NLS.

En analysant la structure de la Transportine 1 en complexe avec le PY-NLS de hnRNP A1, une étude pionnière a proposé différentes règles permettant de décrire la reconnaissance des PY-NLS par la Trn1 (62). Plus précisément, ces règles ont été établies grâce à l'analyse structurale de Trn1 en complexe avec le NLS de hnRNP A1 et à des analyses de mutations et d'alignement concernant des séquences reconnues précédemment comme des NLS de Trn1. Ces règles ont par la suite pu être utilisées pour identifier de nouveaux NLS de Trn1 (62). Ces nouveaux PY-NLS ont été caractérisés par des expériences d'interaction avec Trn1 et/ou des expériences d'import nucléaire.

Bien que les PY-NLS reconnus par la Transportine 1 soient souvent riches en acides aminés chargés positivement, et ce quel que soit leur classe (bPY ou hPY – Figure 15), et qu'ils comprennent souvent un grand nombre de glycines, ils ne partagent pas une similarité de

séquence suffisante qui permettrait de les décrire simplement grâce à une séquence consensus. Pour cette raison, les PY-NLS ne sont pas décrits par une séquence consensus, mais par une série de règles communes (62).

FUS	----	G	P	G	K	M	D	S	R	G	E	H	R	Q	D	R	E	R	P	Y	----														
TAP	-----	V	A	M	S	D	A	Q	D	G	P	R	V	R	Y	N	P	Y	T	T	R	P													
RPL4	-----																M	L	R	L	N	P	Y	----											
Nab2	-----																T	R	F	N	P	L	A	----											
hnRNP M	-----																E	K	N	K	R	G	G	N	R	F	E	P	Y	A	N	P	T		
hnRNP D	-----																K	V	S	R	R	-	G	G	H	Q	N	S	Y	K	P	Y	----		
JKTBP	-----																G	K	A	S	R	G	G	-	G	N	H	Q	N	N	Y	Q	P	Y	----
hnRNP A1	NNQSSNFG	P	M	K	G	G	-	N	F	G	G	R	S	S	G	P	Y	----																	
HCC1	-----																R	G	R	Y	R	S	P	Y	----										
Site de liaison	Site B (HEAT H14-H20)										Site A (HEAT H8-H13)																								

Figure 15 : Alignement de séquences de neuf PY-NLS connus.

Alignement de séquences de neuf PY-NLS dont les structures sont disponibles à ce jour sur la PDB. Pour les différents PY-NLS, trois résidus consensus en rouge (R/H, P, Y) sont reconnus au niveau du site A de Trn1 et un résidu basique en bleu ou hydrophobe en vert (V/A/P/K) est reconnu au niveau du site B de Trn1.

La première règle est que les PY-NLS sont des extensions d'au moins trente acides aminés qui n'ont pas de structure propre lorsqu'ils ne sont pas liés à la Transportine 1. Ils sont structuralement désordonnés à l'état libre et peuvent ainsi adopter des conformations qui s'adaptent à la surface d'interaction de Trn1.

La deuxième règle est que les PY-NLS doivent avoir une charge globale positive. En effet, comme les poches d'interaction des PY-NLS sont globalement chargées négativement, les PY-NLS doivent être chargés positivement pour assurer une bonne interaction avec Trn1 (62, 77).

La troisième règle correspond à la présence de séquences faiblement conservées. En effet, malgré la faible similarité de séquence entre les différents PY-NLS, il est quand même possible de retrouver des régions conservées, principalement au niveau des extrémités N- et C terminales (Figure 15). La première région concerne la région centrale du NLS qui correspond à la région retrouvée à environ 10 à 15 acides aminés de l'extrémité C-terminale du NLS. Cette région est soit riche en acides aminés hydrophobes (pour les hPY), soit riche en acides aminés basiques (pour les bPY). La deuxième région concerne l'extrémité C-terminale, où l'on observe un motif PY conservé. De plus, deux à cinq résidus en amont du motif PY, on

retrouve toujours un acide aminé chargé positivement. La troisième règle est donc qu'un PY- NLS possède la séquence consensus (R/K/H)-X₂₋₅-PY en C-terminal (62, 77).

Pour faciliter la description de la reconnaissance des PY-NLS par Trn1, les différentes interactions présentes le long du PY-NLS ont été regroupées en trois épitopes distincts (Figure 13 B). Ainsi, le motif hydrophobe ou basique en N-terminal constitue l'épitope 1. Le résidu basique du motif (R/K/H)-X₂₋₅-PY constitue l'épitope 2. Et le motif PY constitue l'épitope 3. Les contributions énergétiques de ces épitopes à la liaison avec Trn1 sont quasiment indépendantes étant donné qu'ils sont séparés par des petits espaceurs flexibles (92). De plus, la contribution énergétique de chaque épitope peut varier d'un PY-NLS à un autre, avec dans certains cas une contribution prépondérante de l'épitope 3 et dans d'autres de l'épitope 1. Dans tous les cas, c'est la combinaison des affinités apportées par les trois épitopes qui permet d'établir l'affinité adéquate aux mécanismes d'import par Trn1 (92).

2.3.2.2 La reconnaissance des non-PY-NLS par la Transportine 1

Comme présenté dans la sous-partie précédente, la Transportine 1 est connue pour importer des cargos portant des NLS de la classe des PY-NLS, et leur reconnaissance est bien caractérisée (62, 77, 80). Cependant, Trn1 est également capable de reconnaître et d'importer dans le noyau, des cargos portant des NLS qui ne contiennent pas le motif PY, et donc qui ne répondent pas aux règles précédemment citées, on les appellera les non-PY-NLS (80, 93, 94). Le mécanisme de reconnaissance de ces non-PY-NLS par Trn1 est très mal connu. Ces non- PY-NLS appartiennent à diverses protéines, dont des protéines virales comme la protéine REV du VIH1, des protéines ribosomiques comme la protéine RPL5, la protéine ADAR1 et les histones comme l'histone H3 (Tableau 2).

Récemment, la structure de Trn1 en complexe avec un non-PY-NLS situé au niveau de la queue de l'histone H3 a été résolue par cristallographie (95). Cette structure montre que le NLS de H3 se lie au niveau du même site de fixation que les PY-NLS avec une affinité et des interactions similaires. Les mêmes résidus de Trn1 impliqués dans la reconnaissance des PY-NLS sont également engagés dans la reconnaissance de ce non-PY-NLS (Figure 16). La reconnaissance au niveau de l'épitope 1 en N-terminal compense la perte du motif PY au niveau de l'épitope 3. En effet, le NLS de H3 contient et utilise seulement deux des trois épitopes caractéristiques des PY-NLS : les épitopes 1 et 2. Ainsi le NLS de H3 n'interagit pas avec Trn1 au niveau du site

d'interaction du motif PY, ce site reste vacant. Les interactions au niveau des épitopes 1 et 2 semblent compenser la perte d'interaction correspondant à l'épitope 3 (95).

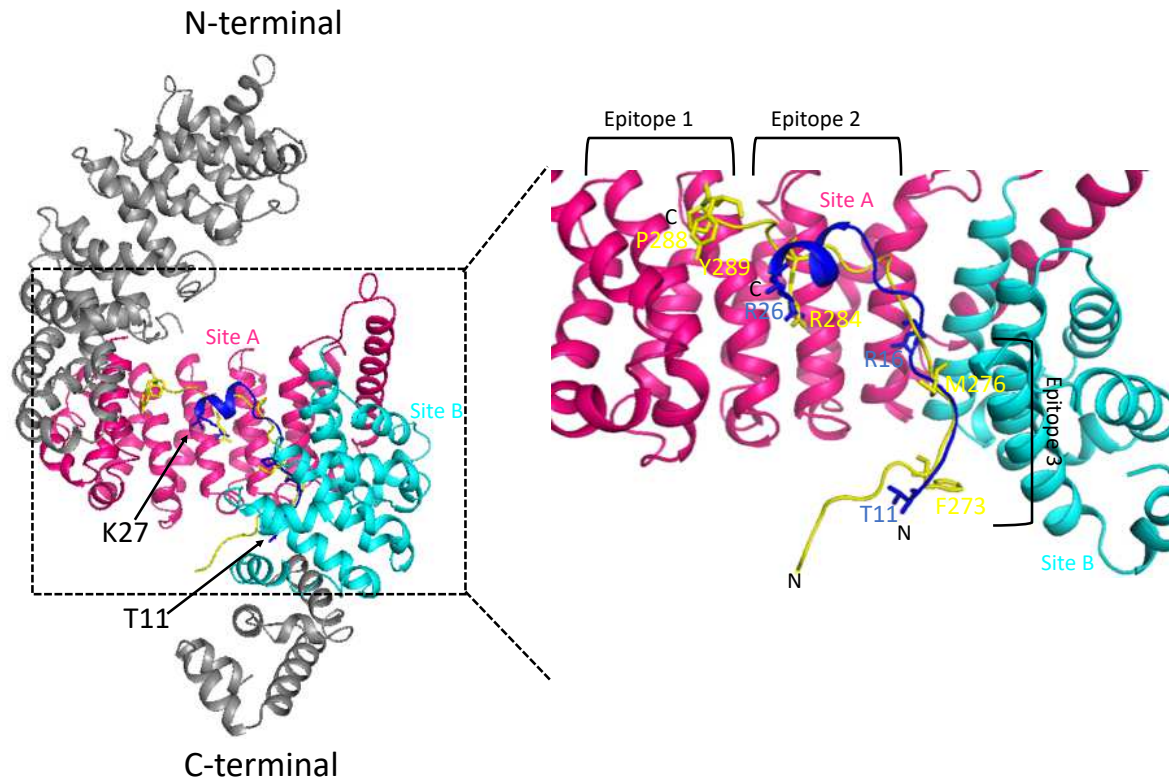


Figure 16 : Reconnaissance d'un non-PY-NLS par Trn1.

Représentation en cartoon illustrant la structure du complexe entre la Transportine 1 (avec le site A (en rose) et le site B (en cyan)) et la superposition du non-PY-NLS de l'histone H3 (en bleu) et du PY-NLS de hnRNP A1 (en jaune) dont les codes PDB sont respectivement 5J3V et 2H4M. Le non-PY-NLS de H3 est reconnu dans la même poche où est également reconnu le PY-NLS de hnRNP A1. D'après Soniat *et al* (95).

Les interactions entre la Transportine 1 et le NLS de H3 comprennent des interactions électrostatiques, polaires et hydrophobes. Ces interactions impliquent les chaînes latérales plutôt que les chaînes principales, ce qui est également le cas pour les PY-NLS. Seuls les résidus 11 à 27 (T11-K27) de la queue de l'histone H3 sont reconnus. Les résidus 11 à 19 (T11-Q19) adoptent une conformation étendue sans structuration alors que les résidus 20 à 27 (L20-K27) forment une hélice α à 2 tours comme observé dans la structure du PY-NLS de la protéine FUS (Figure 14 et Figure 16).

L'épitope 1 du NLS de H3 est formé par les résidus 11 à 18 (T11-K18), riches en acides aminés basiques. L'épitope 2 correspond au résidu R26 du segment 20-27 au niveau de l'hélice (95) (Figure 16).

La structure de Trn1 en complexe avec le NLS de H3 correspond à la première structure résolue d'un complexe de Trn1 avec un non-PY-NLS. Dans ce cas précis, la structure montre des similarités avec les PY-NLS notamment au niveau des épitopes 1 et 2. Cependant cette structure n'éclaire pas sur le mécanisme global de reconnaissance des non-PY-NLS par Trn1 car les non-PY-NLS reconnus par Trn1 sont très variés (80, 94) (Tableau 2). Il faudrait avoir différentes structures de Trn1 en complexe avec d'autres non-PY-NLS pour pouvoir mieux comprendre le mécanisme de reconnaissance des non-PY-NLS par Trn1 et éventuellement établir des règles de reconnaissance pour ces non-PY-NLS comme cela a été fait pour les PY-NLS.

Une étude récente de Bourgeois *et al* (96) suggère que la Transportine 1 est également responsable de l'import nucléaire de la protéine CIRBP (Cold-inducible RNA-binding protein) *via* un non-PY-NLS. La protéine CIRBP porte un non-PY-NLS puisqu'elle ne porte pas de motif PY, il s'agirait d'un NLS jusqu'ici non caractérisé qui ne fait pas partie de la classe des NLS dits classiques. En effet, le NLS de la protéine CIRBP reconnu par Trn1 porte une séquence riche en résidus arginines-glycines (-glycine) correspondant au motif RG/RGG. Ce NLS serait reconnu au niveau du même site de fixation que les PY-NLS mais cette fois-ci avec la participation de la boucle acide H8 (96). Cette étude montre que l'affinité de Trn1 pour le motif RG/RGG est 5 fois moins importante en cas de délétion de la boucle H8. Par ailleurs ces motifs RG/RGG sont également retrouvés dans les séquences des PY-NLS, comme il est le cas pour la protéine FUS (97).

C'est la première étude qui fait référence à la participation de la boucle H8 dans la reconnaissance des cargos par Trn1. En effet, pour l'ensemble des PY-NLS caractérisés, leur reconnaissance n'implique pas la boucle H8. Cette dernière est connue pour participer au relargage des cargos. Cependant, pour d'autres membres de la famille des karyophérines, comme pour l'importine β , l'implication de la boucle H8 dans la reconnaissance des cargos a déjà été révélée.

En lien avec son activité d'import nucléaire, Trn1 est impliquée dans la régulation de la formation de granules d'ARN (90, 98–101). En réponse au stress cellulaire, la protéine FUS reste séquestrée dans le cytoplasme et son accumulation aboutit à la formation de granules de stress. Cette formation de granules de stress est la conséquence directe d'un import nucléaire insuffisant par Trn1. La formation de granules de stress par la protéine FUS peut devenir pathologique dans les neurones, et donner lieu à des pathologies comme la sclérose latérale

amyotrophique ou maladie de Charcot (ALS – Amyotrophic lateral sclerosis) ainsi que la dégénérescence lobaire frontotemporale (FTLD – Frontotemporal lobar degeneration) (78, 88, 97, 98). En condition de stress, la protéine CRIBP est relocalisée dans le cytoplasme au lieu du noyau, ce qui entraîne la formation de granules de stress. Ainsi, un défaut d'import de la protéine CRIBP par Trn1 suite à la méthylation des arginines au niveau de son NLS riche en arginines (motif RG/RGG) entraîne la formation de granules de stress (96).

Dans cette deuxième partie qui a présenté la description de la Transportine 1, nous avons pu voir les caractéristiques de ce récepteur d'import nucléaire. La Transportine 1 est impliquée dans l'import nucléaire de nombreux et différents substrats du cytoplasme vers le noyau. Parmi ces différents substrats, on compte la classe des PY-NLS dont la reconnaissance aujourd'hui est bien caractérisée. La Transportine 1 permet également l'import nucléaire de NLS moins bien caractérisés dont le mécanisme de reconnaissance n'est pas bien connu, les non-PY-NLS. L'enzyme d'édition ADAR1, pour laquelle le NLS atypique a été déterminé, fait partie de ces non-PY-NLS reconnus par Trn1 (102).

3 L'import nucléaire d'ADAR1 par la Transportine 1

La Transportine 1 a été identifiée comme étant l'unique récepteur d'import nucléaire pour la protéine ADAR1. Comme pour tout phénomène d'import nucléaire, ADAR1 est reconnue par Trn1 grâce à son NLS dans le cytoplasme et l'ensemble est transloqué *via* les pores nucléaires dans le noyau où le relargage d'ADAR1 se fait suite à la liaison de la protéine RanGTP. Par ailleurs, les détails moléculaires de la reconnaissance de la protéine ADAR1 par Trn1 ne sont pas connus, puisque le NLS d'ADAR1 ne répond pas aux critères de définition des PY-NLS présentés dans la partie précédente, et ne fait donc pas partie de la classe des PY-NLS. Dans cette troisième et dernière partie, nous allons nous intéresser à la protéine ADAR1, et notamment à son import nucléaire par Trn1.

3.1 La famille des protéines ADAR

La protéine ADAR1 est une enzyme qui fait partie de la famille des enzymes ADAR (*adenosine deaminase acting on RNA*) ou adénosine désaminases agissant sur l'ARN double brin (ARNdb). Ces enzymes sont connues pour être impliquées dans l'édition de l'ARN double brin. Le rôle des protéines ADAR est d'ajuster et d'optimiser différents processus biologiques (103–105).

3.1.1 L'édition des ARN par les enzymes ADAR

Tout d'abord, l'édition des ARN est un processus biologique post-transcriptionnel qui permet d'apporter des modifications dans les séquences d'ARN (103, 105, 106). Ainsi, la séquence d'ARN obtenue suite à ce processus ne correspond pas à la séquence du gène initial. On peut donc considérer l'édition comme une forme de régulation post-transcriptionnelle. De plus, l'édition des ARN peut survenir au cours de la transcription mais également après la transcription. Aussi, ces modifications post-transcriptionnelles par le biais de conversion d'un nucléotide en un autre, peuvent conduire à des modifications au niveau des codons lorsque celles-ci surviennent dans des séquences codantes. Par conséquent, l'édition peut conduire à des altérations au niveau de la séquence de certaines protéines, et donc *in fine* apporter une diversification dans les fonctions des protéines.

Le phénomène d'édition a été découvert initialement dans les mitochondries des trypanosomes il y a plus de trente ans maintenant (107). Il s'agit de l'édition des cytidines en uridines dans des séquences d'ARN. L'édition des cytidines en uridines dans les ARN messagers (ARNm) est catalysée par les APOBEC (*apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like*) qui sont une famille de cytidines désaminases. De plus, ce type d'édition a depuis été retrouvé dans différents autres compartiments comme le noyau, et les chloroplastes chez les plantes. Il existe différents types de mécanismes d'édition, on compte l'édition concernant la désamination des cytidines en uridines et la désamination des adénosines en inosines (101, 103, 106, 107).

L'édition des adénosines en inosines dans des séquences d'ARNdb est le processus d'édition qui est le plus observé dans le règne animal. Cette conversion est catalysée par une classe d'enzymes, les désaminases de la famille des protéines ADAR. Dans ce processus d'édition, l'adénosine est convertie en inosine par une réaction de désamination hydrolytique de la base adénine (105, 110, 111) (Figure 17).

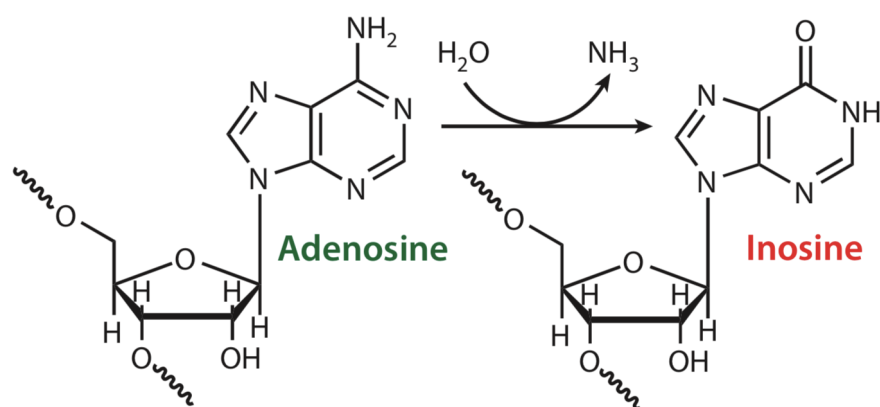


Figure 17 : Réaction de désamination hydrolytique de l'adénosine en inosine.

L'édition des adénosines en inosines est réalisée par les enzymes ADAR qui catalysent la réaction de désamination hydrolytique qui permet de modifier la base adénine de l'adénosine. D'après Nishikura *et al* (105).

De plus, le phénomène d'édition est chez certaines espèces un phénomène biologique essentiel. En effet, l'inactivation de certains gènes ADAR peut être létal, c'est le cas notamment du gène ADAR1 chez les mammifères (112). Les études menées par Wang *et al* (112) ont indiqué que l'expression d'ADAR1 est importante dans le développement embryonnaire de la souris. Les embryons de souris (ADAR1^{+/−}) présentent de faible niveau d'expression de l'enzyme ADAR1 dès les premières semaines de développement, ce qui entraîne un faible taux d'édition dans le système hématopoïétique causant un défaut de prolifération/différenciation des érythrocytes. En conséquence, une absence d'érythrocytes matures entraîne une hypoxie dans les tissus et donc une éventuelle mort de l'embryon.

3.1.2 Les protéines de la famille ADAR

Les enzymes ADAR ont été initialement identifiées chez les œufs et les embryons de *Xenopus laevis* (110, 111). Chez les mammifères, l'enzyme ADAR1 a tout d'abord été découverte (113), puis les enzymes ADAR2 (114, 115) et ADAR3 (114).

Les enzymes ADAR ont été caractérisées chez quasiment tous les métazoaires, mais elles sont absentes chez les protozoaires, les levures, les plantes et les champignons (105, 116). Les protéines ADAR sont très conservées chez les animaux, des nématodes à l'homme (103, 115, 116) (Figure 18). Toutefois le phénomène d'édition d'adénosine en inosine est présent également chez les protozoaires et les plantes, mais dans ces cas, d'autres familles de désaminases interviennent. En effet, il existe, à côté de la famille des ADAR, une autre famille de désaminase, les ADAT (adenosine deaminase acting on tRNA). Les ADAT partagent une

homologie de séquence avec les ADAR. Les ADAT sont très conservées chez les eucaryotes, elles sont conservées des levures à l'homme (119, 120). Et, contrairement aux ADAR, les ADAT présentent des orthologues chez les bactéries. Les enzymes ADAT sont impliquées dans l'édition des adénosines en inosines au niveau de la région anticodon des ARN de transfert (ARNt) (105, 119, 121).

3.1.2.1 L'organisation structurale des protéines ADAR

Les différentes enzymes ADAR au sein de la même espèce, ou appartenant à des espèces différentes, possèdent une organisation structurale en domaine commune. En effet, ces enzymes possèdent une à trois copies de domaines de liaison à l'ARNdb (dsRBD – double-stranded RNA-binding domain) suivant les espèces (Figure 18). Les dsRBD diffèrent en fonction des espèces non seulement par leur nombre mais aussi par leur position. On observe une différence d'espacement entre les différents dsRBD selon les espèces.

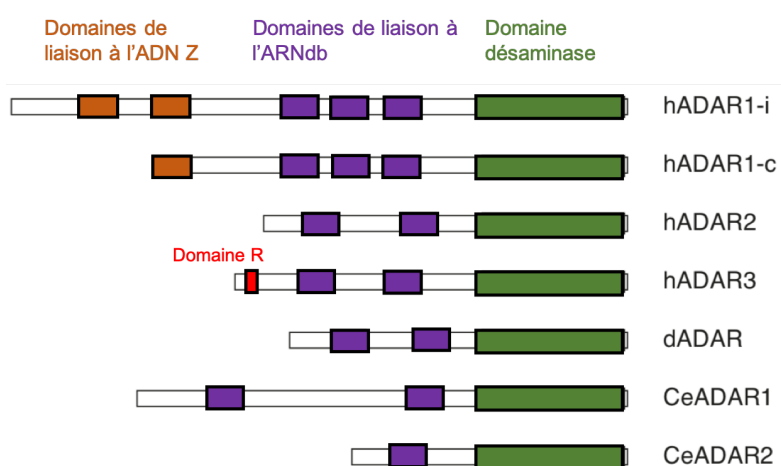


Figure 18 : Organisation structurale des membres de la famille des protéines ADAR.

Les membres de la famille des enzymes ADAR sont conservés des nématodes à l'homme. Trois membres sont retrouvés chez les vertébrés (ADAR1, ADAR2 et ADAR3), un membre est répertorié chez la drosophile (dADAR) et deux membres sont comptés chez *C. elegans* (CeADAR1 et CeADAR2). Les membres de la famille présentent un domaine désaminase (vert) très conservé en C-terminal et une à trois copies de domaines de liaison à l'ARN double brin (violet). En sus, l'enzyme ADAR1 présente un à deux domaines de liaison à l'ADN Z (orange) en N-terminal. Cette dernière est retrouvée sous deux isoformes : une isoforme inducible (hADAR1-i) et une isoforme constitutive (hADAR1-c). D'après Barraud *et al* (122).

Ce sont ces domaines dsRBD qui font que les enzymes ADAR reconnaissent et modifient les ARN présentant des régions en ARNdb. Certaines enzymes ADAR présentent des domaines qui leur sont uniques, par exemple, la protéine ADAR1 comptent une à deux copies de domaines de liaison à l'ADN Z en N-terminal, et la protéine ADAR3 se compose d'un domaine

de liaison à l'ARN simple brin riche en arginine (domaine R) en N-terminal (Figure 18). Le domaine catalytique (domaine désaminase) de ces enzymes qui permet de réaliser l'édition des adénosines en inosines est retrouvé dans la partie C-terminale (103–105).

3.1.2.2 La localisation des protéines ADAR

Chez l'homme, contrairement aux enzymes ADAR1 et ADAR2 qui possèdent une activité catalytique de désamination (113–115), l'enzyme ADAR3 ne présente pas d'activité catalytique démontrée. Elle possède cependant la même organisation structurale en domaine que les deux autres membres de la famille ADAR, et notamment un domaine désaminase (123) (Figure 18). Il est intéressant de noter que ADAR3 n'est exprimée que dans le cerveau (124), et que même si l'origine de son inactivité apparente n'est à ce jour pas élucidée, il se pourrait que les substrats sur lesquels elle est active dans le cerveau n'aient toujours pas été identifiés.

À l'inverse, les enzymes ADAR1 et ADAR2 sont exprimées dans de nombreux tissus et modifient un très grand nombre de substrats (113–115). De manière générale, chez les mammifères, les enzymes ADAR sont principalement localisées dans le noyau cellulaire. Plus précisément, l'isoforme ADAR1-c (Figure 18) et l'enzyme ADAR2 sont localisées dans le nucléole et cette localisation dépend des dsRBD. La localisation des enzymes ADAR dans le nucléole se fait vraisemblablement grâce à la présence d'une grande concentration d'ARNdb dans le nucléole, notamment du fait de la présence des ARN ribosomiques (ARNr) (104).

L'import nucléaire des enzymes ADAR2 et ADAR3 est réalisé par la voie classique impliquant le couple importine α /importine β (125), tandis que l'import nucléaire de l'enzyme ADAR1 est réalisé par une voie non classique impliquant la Transportine 1 (126).

3.1.3 La sélectivité et la spécificité de l'édition par les enzymes ADAR

Les substrats des enzymes ADAR présentent une grande diversité structurale ce qui résulte en un large répertoire de cibles pour ces enzymes. Les enzymes ADAR sont capables de modifier différents ARNdb de séquences variées. L'édition des adénosines en inosines catalysée par les ADAR peut être dans certains cas très spécifique et dans d'autres cas peut survenir de manière quasi aléatoire (103–105, 108).

3.1.3.1 La spécificité des enzymes ADAR

Les enzymes ADAR ont une spécificité structurale, elles peuvent lier des ARNdb mais pas les ARN simple brins (127). Ces enzymes ADAR ont des cibles d'origine endogènes ou virales (103, 128). Mais elles vont uniquement agir sur des ARN présentant des régions formant des structures en ARNdb. Il peut s'agir de séquences codantes, des introns, des régions UTR (*untranslated regions*) 5' et 3'. En effet, la majorité des sites d'édition des adénosines en inosines dans des séquences d'ARNdb se trouve dans des régions non codantes comme les rétrotransposons (103, 105, 108, 109) (Figure 19). Les rétrotransposons de type SINE et LINE sont les cibles les plus fréquentes. Ces rétrotransposons ou éléments transposables sont des séquences d'ADN endogènes mobiles qui constituent environ 40 % du génome humain (129). Les éléments transposables les plus actifs chez l'homme sont des rétrotransposons de type non- LTR (*long terminal repeat*), qui peuvent être soit des éléments longs (LINE : long interspersed repetitive element) ou des éléments courts (SINE : small interspersed repetitive element). Les éléments SINE sont divisés en deux sous classes, les séquences « Alu » et « SVA ». Ces éléments transposables ont un rôle dans les phénomènes d'instabilité génomique (129, 130).

À côté de ces phénomènes d'édition des régions non codantes, certains précurseurs d'ARNm codants (pré-ARNm) sont des cibles d'édition d'adénosine en inosine (Figure 19). Dans ces cas, les sites d'édition se trouvent en général au niveau des jonctions entre les exons et les introns où des structurations de type ARNdb en épingle à cheveux peuvent se former entre la séquence codante de l'exon et une séquence imparfaitement complémentaire, appelée séquence complémentaire au site d'édition (« *editing complementary sequence* ») au niveau de l'intron adjacent (103–105, 108). Les précurseurs de micro-ARN (pré-miARN) constituent également des cibles d'édition d'adénosines en inosines. Dans ce cas de figure, le processus d'édition peut réguler la maturation des pré-miARN, et même influencer sur la cible finale du miARN (131, 132).

Lorsque qu'elle a lieu dans des séquences codantes, l'édition d'ARN peut entraîner des altérations de fonctions pour des gènes ciblés. Une diversification des protéines, et entre autres, la formation d'isoformes peut être obtenue. Effectivement, les inosines sont reconnues comme étant des guanosines par la machinerie cellulaire de traduction, car elles ont la capacité de former deux liaisons hydrogènes avec les cytidines ce qui peut entraîner des changements de signification des codons. Cependant, dans le cas où l'édition survient dans des séquences non

codantes, les conséquences peuvent aller de l'altération de la stabilité de l'ARN à l'altération du transport des ARN cellulaires ou viraux (103, 105).

3.1.3.2 La sélectivité des enzymes ADAR

Le phénomène d'édition peut dans certains cas être un processus très sélectif, où une adénosine unique est précisément éditée, ou peut se faire de manière non sélective, ce qui conduit à la désamination non sélective de nombreuses adénosines dans un substrat donné (103–105, 108). Les substrats des enzymes ADAR peuvent être des petites séquences d'ARNdb avec un minimum d'une vingtaine de paires de bases environ, ou de longues séquences d'ARNdb de plusieurs centaines de paires de bases (133). Les adénosines au sein de ces longues séquences d'ARN sont éditées de manière non sélective, ce qui entraîne l'édition d'environ 50 % des adénosines de ces séquences. À l'inverse, dans les petites séquences d'ARNdb, contenant environ 20-30 paires de bases, et qui présentent souvent des mésappariements de bases, des renflements ou des boucles, l'édition des adénosines en inosines se fait en général de manière beaucoup plus spécifique (134).

Bien que les enzymes ADAR peuvent se lier de manière non spécifique aux ARNdb, ces enzymes présentent une légère préférence de séquence pour la désamination d'une adénosine en particulier par rapport à une autre (103, 122). Ce concept de préférence a été suggéré par des études sur les sites d'édition des ARNm des sous-unités des récepteurs de glutamate notamment la sous-unité GluR-2 (ou GluR-B). Ces études ont révélé une préférence quant à la localisation des adénosines dans les cibles de substrats des enzymes ADAR1 et ADAR2, notamment les bases nucléotidiques qui les entourent (103, 122, 135, 136).

En effet, les enzymes ADAR sont connues pour être responsable de l'édition de la sous-unité GluR-2 (des récepteurs AMPA) au niveau de trois sites (114, 115, 137). L'édition du premier site conduit à la modification d'une glutamine en une arginine (site Q/R) tandis qu'au niveau du deuxième site, l'édition entraîne la modification d'une arginine en une glycine (site R/G). L'édition du troisième site situé au niveau de l'intron n'entraîne pas de changement au niveau de la protéine. Ces trois sites sont modifiés par différents membres de la famille ADAR. Le site Q/R est principalement modifié par ADAR2 alors que le site R/G peut être édité par les deux enzymes, ADAR1 et ADAR2. Le site situé au niveau de l'intron est principalement édité par ADAR1. Ceci suggère que les enzymes ADAR sont capables de pouvoir discriminer des adénosines parmi des substrats similaires (103, 135, 136). L'enzyme ADAR1 a une préférence

pour des adénosines ayant un voisin en 5' comme suit : $U = A > C > G$ alors que l'enzyme ADAR2 a une préférence pour des adénosines ayant un voisin en 5' comme suit : $U \approx A > C = G$. L'enzyme ADAR2 a aussi une préférence pour les adénosines ayant un voisin en 3' comme suit : $G = U > C = A$ (135).

Le nucléotide qui s'apparie à l'adénosine à éditer, a une influence sur l'édition. Il existe une préférence pour la cytidine (les mésappariements AC dans le site GluR-2 R/G) plutôt que pour les uridines (les appariements AU comme dans le site GluR-2 Q/R).

De plus, il a été proposé que la plupart des réactions d'édition chez les mammifères est catalysée par ADAR1. ADAR1 agit essentiellement dans des sites répétitifs alors que ADAR2 agit préférentiellement dans des sites non répétitifs (138).

Par ailleurs, la sélectivité des enzymes ADAR dépend d'une part du domaine désaminase et d'autre part des dsRBD. Les dsRBD contribuent à l'affinité de liaison et à la sélectivité (137, 138), mais la différence de sélectivité entre les différentes protéines ADAR semble plutôt provenir des différences au niveau de leurs domaines désaminases (136, 141).

La capacité de pouvoir discriminer certaines adénosines par rapport à d'autres, c'est-à-dire d'avoir une seule ou peu d'adénosine éditée dans la séquence cible définit le concept de spécificité (103, 122, 142). La présence de structuration dans la séquence cible comme des renflements ou des mésappariements est fréquente chez les substrats des enzymes ADAR. La présence de structuration étant conservée dans l'évolution, cela indique que la structuration des ARN est très importante dans la spécificité de l'édition. Les séquences d'ARNdb aux environs des sites où l'édition est réalisée de manière spécifique sont aussi particulièrement conservées, ceci indique que la séquence joue un rôle dans la sélectivité de l'édition par les enzymes ADAR. Ainsi, la structure et la séquence de l'ARN du site d'édition détermine la spécificité/sélectivité de l'édition des ADAR (103, 122, 136).

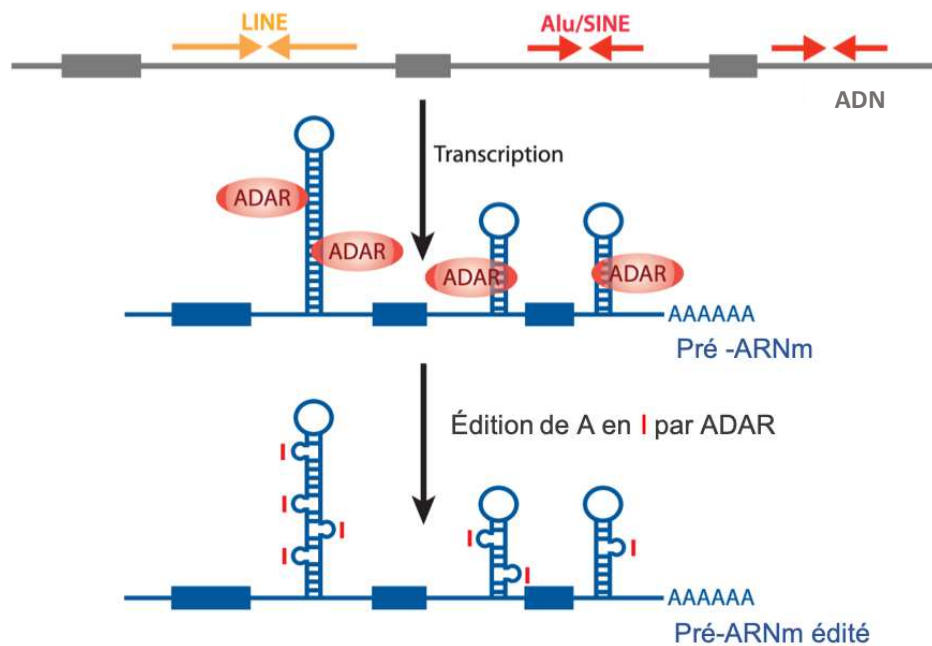


Figure 19 : Edition d'adénosine (A) en inosine (I) dans des séquences non codantes d'ARNdb.

Les séquences non codantes de type LINE ou SINE, présentes dans les introns ou les régions UTR forment des séquences d'ARN double brins qui deviennent des substrats pour l'édition d'adénosine en inosine catalysée par les enzymes de la famille ADAR. D'après Nishikura *et al* (105).

3.2 La protéine ADAR1

Bien que la protéine ADAR1 fût la première protéine de la famille ADAR identifiée (111, 141), elle n'est pourtant pas celle dont les mécanismes conduisant à l'édition sont les mieux maîtrisés. En effet, il y a peu de connaissances sur la reconnaissance des substrats par ADAR1 alors que beaucoup d'études ont été faites sur celle d'ADAR2, notamment sur le domaine désaminase (136, 144, 145).

3.2.1 L'organisation structurale d'ADAR1

La protéine ADAR1 existe sous deux isoformes, une isoforme dite inductible, qui correspond à une version longue d'environ 150 kDa, appelée ADAR1p150 (ou ADAR1-i), et une isoforme dite constitutive, qui correspond à une version courte d'environ 110 kDa, nommée ADAR1p110 (ou ADAR1-c). En effet, un des trois promoteurs qui contrôle la transcription du gène *ADAR1* est un promoteur inductible par des interférons, notamment les interférons α et γ (IFN- α/γ), et l'ARNm qui en résulte donne l'isoforme ADAR1p150. Les deux autres promoteurs sont constitutifs et ils dirigent la transcription directe des ARNm qui donnent la synthèse de l'isoforme ADAR1p110 qui résulte d'un épissage alternatif (146). De plus, la

synthèse d'ADAR1p150 se fait en présence IFN- α/γ , et son expression est augmentée en conditions inflammatoires dans les lymphocytes et les macrophages (147). Même si les deux isoformes possèdent la même activité enzymatique, elles présentent des différences de localisation et de composition en domaine structuraux.

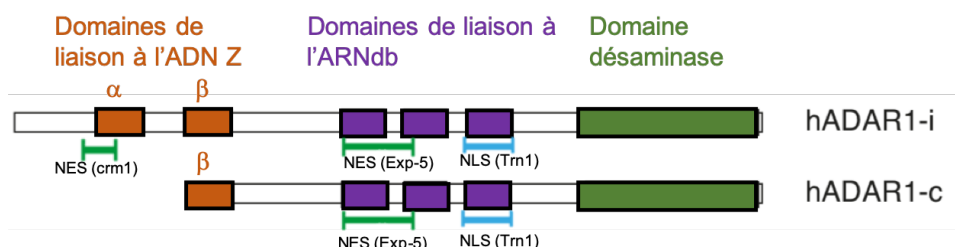


Figure 20 : Organisation structurale de l'enzyme ADAR1.

L'enzyme ADAR1 présente deux isoformes. Ces deux isoformes diffèrent au niveau de leurs extrémités N-terminales avec pour l'isoforme ADAR1-i, deux domaines Z α et Z β de liaison à l'ADN-Z et l'isoforme ADAR1-c, un domaine Z β de liaison à l'ADN-Z. Les deux isoformes présentent trois copies de domaines de liaison à l'ARN double brin (dsRBD) et un domaine désaminase. On retrouve un NLS au niveau du troisième dsRBD et un NES qui chevauche les deux premiers dsRBD pour chacune des isoformes. Un deuxième NES est retrouvé uniquement au niveau de l'extrémité N-terminale de l'isoforme ADAR1-i. D'après Barraud *et al* (122).

Effectivement, ADAR1p150 a une localisation principalement dans le cytoplasme cellulaire puisqu'elle possède un NES (Figure 20) situé au niveau de son domaine supplémentaire Z α (148). Tandis qu'ADAR1p110 est principalement localisée dans le noyau puisqu'elle ne possède pas ce NES supplémentaire (126). Nous y reviendrons plus en détail dans la sous-partie suivante.

3.2.1.1 Les domaines de liaison aux acides nucléiques Z

L'isoforme ADAR1p150 possède deux versions de domaines de liaison à l'ADN-Z en N-terminal (un domaine de liaison dénommé Z β et un domaine de liaison dénommé Z α) alors que l'isoforme ADAR1p110 ne possède que le domaine de liaison Z β (Figure 20). Seul le domaine Z α possède une activité de liaison aux formes Z des acides nucléiques (ADN-Z et ARN-Z) (122, 149–151). Le domaine Z β n'a pas la capacité de se lier à ces formes Z, car plusieurs résidus essentiels à la reconnaissance des formes Z des acides nucléiques y sont absents (152, 153).

Bien que le rôle biologique du domaine Z α n'ait pas été clairement défini, une fonction possible pourrait être de diriger ADAR1 vers des sites de transcription active. En effet, l'ADN-Z est

stabilisé par un surenroulement négatif de l'ADN, qui apparaît transitoirement en amont d'une ARN polymérase active. De plus, comme ce domaine a une capacité de liaison à l'ARN-Z (151), il pourrait intervenir dans la reconnaissance privilégiée de certaines séquences d'ARN ayant une propension à adopter des formes en ARN-Z et ainsi diriger ADAR1p150 vers ces séquences particulières, par exemple les séquences Alu des éléments SINE (154–156).

De plus, ADAR1p150 joue un rôle important dans la réponse immunitaire en réponse aux infections virales, avec dans certains cas un effet pro-viral et dans d'autres cas un effet antiviral (157). En effet, l'édition du virus de l'hépatite C par ADAR1p150 entraîne l'inhibition du cycle répliatif du virus et permet ainsi son élimination (158). Par contre, ADAR1p150 provoque la persistance du VIH dans l'organisme et le rend plus agressif bien que l'édition ne soit pas la cause directe de cette conséquence (159). Un rôle pour le domaine $Z\alpha$ dans l'implication d'ADAR1p150 dans les infections virales a été suggéré car il pourrait diriger la protéine au niveau des sites de transcription active des virus à ARN (151).

3.2.1.2 Le domaine de désamination

En C-terminal, les deux isoformes d'ADAR1 (ADAR1p150 et ADAR1p110) portent chacune une copie de domaine désaminase qui porte l'activité catalytique des enzymes ADAR.

Le domaine désaminase des protéines ADAR est le domaine le plus conservé au cours de l'évolution. C'est un domaine constitué d'environ 400 acides aminés. Sur ces 400 acides aminés environ 59 % sont identiques entre les protéines ADAR1 et ADAR2 (145) tandis que sur l'ensemble de leur séquence, l'identité d'ADAR1 et d'ADAR2 n'est que de 39 %. La structure du domaine désaminase d'ADAR2 isolé a révélé que le cœur de ce domaine présentait une activité catalytique qui dépend du zinc. Aussi, le cœur de ce domaine est stabilisé par l'inositol hexakisphosphate (IP_6) qui a été retrouvé dans le cœur de cette structure (145, 160).

La résolution de la structure du domaine désaminase en complexe avec un ARNdb a permis de montrer que les enzymes ADAR catalysent la désamination de l'adénosine en inosine sur le carbone en position C6 (Figure 17) par un mécanisme de « *base flipping* » qui consiste en la bascule de l'adénosine à éditer au niveau de la poche du site catalytique (145, 161). De plus, il est considéré que le domaine désaminase peut se comporter comme une unité catalytique indépendante des dsRBD (162).

3.2.1.3 Les domaines de liaison à l'ARN double brin

Les deux isoformes d'ADAR1 possèdent chacune trois domaines dsRBD qui sont des domaines de liaison à l'ARNdb (Figure 20). Les dsRBD donnent la capacité aux enzymes ADAR de reconnaître des ARN hautement structurés. Ces domaines peuvent se lier aux ARN duplex parfaits mais aussi à des ARNdb imparfaits comportant des renflements, des boucles, ou des mésappariements, ainsi qu'à des ARN en épingles à cheveux (103–105, 109). La fonction des dsRBD au sein des enzymes ADAR, serait de guider l'activité catalytique de ces enzymes vers les adénosines cibles à éditer (161).

La reconnaissance des substrats ARN par les enzymes ADAR se fait *via* les domaines de liaison à l'ARNdb de type dsRBD, dont ADAR1 possède trois copies. Un dsRBD est un domaine d'une soixantaine d'acides aminés que l'on retrouve chez les protéines de quasiment tous les protozoaires et métazoaires, et même des virus. Les acides aminés qui le constituent sont de manière générale très conservés le long de la séquence, néanmoins, l'extrémité C-terminale reste la région la plus conservée de ce domaine (163). Ces dsRBD sont présents au sein de différentes protéines avec des fonctions très variées.

Le rôle principal des dsRBD est de reconnaître des ARNdb. Cette reconnaissance se fait principalement grâce à la reconnaissance de la forme des hélices d'ARN de type A. Néanmoins, les dsRBD peuvent également reconnaître des éléments de séquence au sein des structures en ARNdb formées par leurs substrats. Certains dsRBD peuvent également réaliser des interactions protéines-protéines leur permettant ainsi de jouer des rôles dans des processus cellulaires qui ne sont pas directement reliés à la présence d'ARNdb (164, 165).

Différentes structures de dsRBD ont été résolues par cristallographie et par résonance magnétique nucléaire (RMN) (163, 166). Ces structures ont révélé une structure conservée des dsRBD avec une topologie $\alpha 1$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ (Figure 21). Dans cette structure, les deux hélices $\alpha 1$ et $\alpha 2$ viennent se loger contre le feuillet β antiparallèle à trois brins $\beta 1\beta 2\beta 3$. En ce qui concerne la formation du cœur hydrophobe du domaine, on retrouve principalement des chaînes latérales aliphatiques dans les deux hélices $\alpha 1$ et $\alpha 2$ et des cycles aromatiques dans les trois brins $\beta 1$, $\beta 2$ et $\beta 3$ (163, 164).

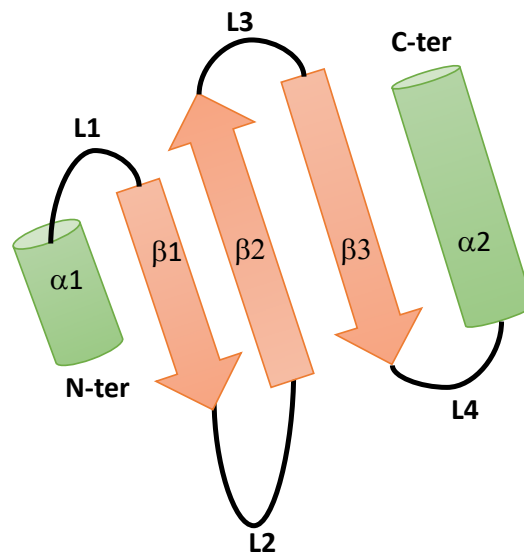


Figure 21 : Structure topologique d'un domaine de liaison à l'ARN double brin (dsRBD).

Les dsRBD présentent une structure très conservée avec une topologie $\alpha 1$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$. Les deux hélices α sont logées contre le feuillet β à trois brins. Chaque structure secondaire est reliée par une boucle, L1-L4. D'après Masliah *et al* (163).

3.2.1.3.1 La reconnaissance canonique

Différentes structures de dsRBD en complexe avec des ARNdb ont révélé le mode de reconnaissance dit canonique d'un ARN par les dsRBD (163, 166). Une reconnaissance canonique correspond à une interaction faisant intervenir trois régions distinctes au niveau du dsRBD (Figure 22). Tout d'abord, la région 1 qui se situe au niveau de l'hélice $\alpha 1$, à l'extrémité N-terminale du domaine, et la région 2 qui correspond à la boucle (L2) reliant les deux premiers brins $\beta 1$ et $\beta 2$, reconnaissent l'ARNdb au niveau de deux sillons mineurs successifs. Ensuite, la région 3 qui comprend la boucle L4 et l'extrémité N-terminale de l'hélice $\alpha 2$, reconnaît le grand sillon situé entre les deux petits sillons reconnus par les régions 1 et 2 (Figure 22).

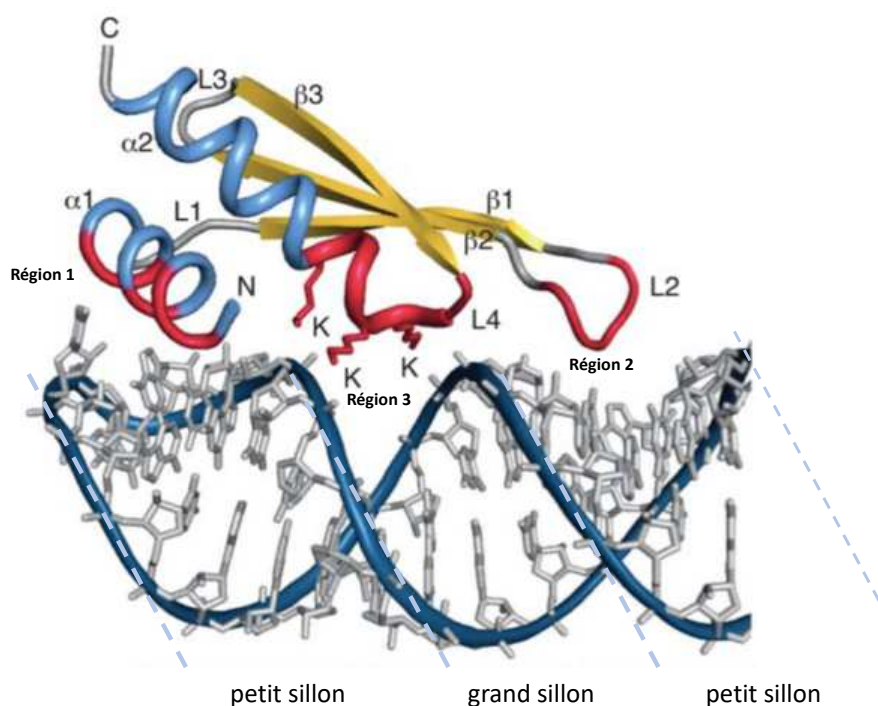


Figure 22 : Reconnnaissance canonique d'un ARNdb par un dsRBD.

Illustration de l'interaction canonique entre un dsRBD (au-dessus) et un ARN double brin (en-dessous). Une interaction canonique entre un dsRBD et un ARN double brin fait intervenir des points de contact entre trois régions distinctes (en rouge) au niveau du dsRBD. La première région est localisée au niveau de l'hélice $\alpha 1$, la deuxième région correspond à la boucle L2 qui permet de relier les brins $\beta 1$ et $\beta 2$, et la troisième région comprend la pointe de l'hélice $\alpha 2$ ainsi que la boucle L4. D'après Masliah *et al* (163).

Cette région 3 possède, un motif riche en lysines très conservé (KKxAK) qui interagit avec le squelette sucre-phosphate de l'ARNdb (167). De plus, l'interaction entre le dsRBD et l'ARNdb se fait par des contacts directs avec les groupes 2'hydroxyles des bases ou du sucre ribose et par des interactions indirectes *via* des molécules d'eau. Ainsi, cette interaction est principalement considérée comme une interaction non-spécifique indépendante de la séquence du dsRBD, qui va plutôt reconnaître la forme de l'hélice A de l'ARNdb. Toutefois, certains dsRBD présentent une grande préférence pour des duplex d'ARN imparfaits ou présentant des tiges-boucles, et d'autres dsRBD vont spécifiquement reconnaître des bases dans le petit sillon mineur *via* des « contacts séquence spécifique » au niveau des régions 1 et 2 (168).

3.2.2 La localisation des isoformes d'ADAR1

Comme indiqué plus haut, les deux isoformes d'ADAR1 diffèrent par leur localisation cellulaire. En effet, ADAR1p150 est essentiellement localisée dans le cytoplasme alors que

ADAR1p110 est principalement localisée dans le noyau. Cette localisation est le résultat complexe de la présence de multiples signaux de transport NES et NLS sur les isoformes d'ADAR1. En effet, ADAR1p150 possède un NES fonctionnel au niveau de son extrémité N-terminale, plus précisément au niveau du domaine Z α (Figure 20). Ce NES est reconnu par l'exportine CRM1 (appartenant à la famille des karyophérines β), et ainsi ADAR1p150 est exportée du noyau par le mécanisme d'export nucléaire présenté plus haut (105, 148). L'isoforme ADAR1p110 ne possède pas quant à elle ce NES en N-terminal. Cependant, cette isoforme présente un NES au niveau des deux premiers dsRBD d'ADAR1, qui est également présent chez ADAR1p150 (Figure 20). L'export nucléaire contrôlé par ce NES est médié par l'exportine-5 (appartenant aussi à la famille des karyophérines β) (126). L'exportine-5 est connue pour intervenir dans l'export nucléaire de petits ARN structurés comme les précurseurs de micro-ARN (pre-miARN) ou des ARNt (169–172). Ainsi, l'exportine-5 n'interagirait pas directement avec ADAR1, mais interviendrait de façon indirecte *via* les ARNdb auxquels les dsRBDs seraient complexés. De cette façon, ADAR1 serait exportée du noyau vers le cytoplasme lorsqu'elle est liée à des ARNdb (126).

Finalement, les deux isoformes d'ADAR1 possèdent un NLS au niveau de la région de leur troisième dsRBD, NLS qui est reconnu par la Transportine 1 et qui permet ainsi l'import d'ADAR1 dans le noyau (123-126). La présence d'au moins un NLS et d'au moins un NES sur chacune des isoformes d'ADAR1 fait que ces protéines effectuent en permanence des navettes entre le noyau et le cytoplasme (102, 173).

Bien que les deux isoformes portent chacune au moins un NES et un NLS, leur localisation principale n'est pas la même. Ceci s'explique par le fait que le NES porté exclusivement par ADAR1p150, NES dépendant de CRM1 (Figure 20) présente une efficacité plus forte que le NES porté par ADAR1p110, NES dépendant de l'exportine-5 et de la liaison aux ARNdb. De plus, l'activité d'import du NLS des deux isoformes est plus faible que l'activité d'export du NES spécifique à ADAR1p150, mais plus forte que l'activité d'export du NES présent dans ADAR1p110. Le résultat final de cette compétition entre les différents NES et NLS fait qu'ADAR1p150 est principalement localisée dans le cytoplasme alors qu'ADAR1p110 est principalement localisée dans le noyau.

L'import nucléaire des deux isoformes d'ADAR1 est assuré par Trn1 qui est l'unique récepteur d'import nucléaire d'ADAR1. Cependant, la région identifiée comme correspondant au NLS

d'ADAR1 ne présente pas de ressemblance avec les substrats préférentiels de Trn1, les PY-NLS, et ne possède par exemple pas de motif PY (126, 174).

Le dsRBD3 d'ADAR1 ne possède pas uniquement des propriétés d'interaction avec Trn1, mais peut également se lier aux ARNdb. Cependant, la liaison du NLS d'ADAR1 à Trn1 est inhibée en présence d'ARNdb. Et comme la Transportine 1 reconnaît le dsRBD3 d'ADAR1 dans le cytoplasme, ceci suggère qu'ADAR1 ne peut être importée dans le noyau qu'en absence d'ARN lié au niveau de ses dsRBD.

3.2.3 Le NLS bimodulaire d'ADAR1

Le NLS de la protéine ADAR1 humaine a été identifié pour la première fois, il y a une vingtaine d'années (102, 173). Grâce à une recherche systématique d'activité de NLS dans la séquence d'ADAR1, Eckman *et al* (102), ont déterminé qu'il existe une activité d'import nucléaire au niveau du dsRBD3 d'ADAR1 (dsRBD3). Le NLS minimal d'ADAR1 est porté par le dsRBD3 et le recouvre entièrement. Puis, la résolution par RMN de la structure du NLS minimal d'ADAR1 ainsi identifié (résidus 708-801), a montré que ce segment était constitué d'un domaine dsRBD structuré au niveau des résidus 716-795, et que ce domaine était flanqué par des extrémités flexibles non-structurées (174) (Figure 23). Le dsRBD3 d'ADAR1 présente la topologie classique des dsRBD, avec les éléments de structure secondaire $\alpha 1$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ (Figure 21), mais possède également une hélice supplémentaire en N-terminal, l'hélice αN (Figure 23). La présence d'hélices supplémentaires avait déjà été observée en N- et C-terminal de certains dsRBD (163, 175).

Le NLS reconnu par la Transportine 1 et porté par cette région autour du dsRBD3 d'ADAR1, ne présente pas les caractéristiques des PY-NLS. En effet, il ne répond pas aux différentes règles qui définissent les PY-NLS, et notamment il ne comprend pas de motif PY.

Le NLS au niveau du dsRBD3 est un NLS atypique puisqu'il recouvre tout le dsRBD3 et compte environ 90 acides aminés ce qui n'est pas commun pour un NLS. De façon plus précise, le NLS n'est pas constitué de l'ensemble du dsRBD en tant que tel, mais il est formé par les deux extensions flexibles situées de part et d'autre du dsRBD3 (Figure 23). Ces deux extensions N- et C-terminales sont positionnées dans une conformation fonctionnelle par le domaine dsRBD structuré, qui joue un rôle de support structural.

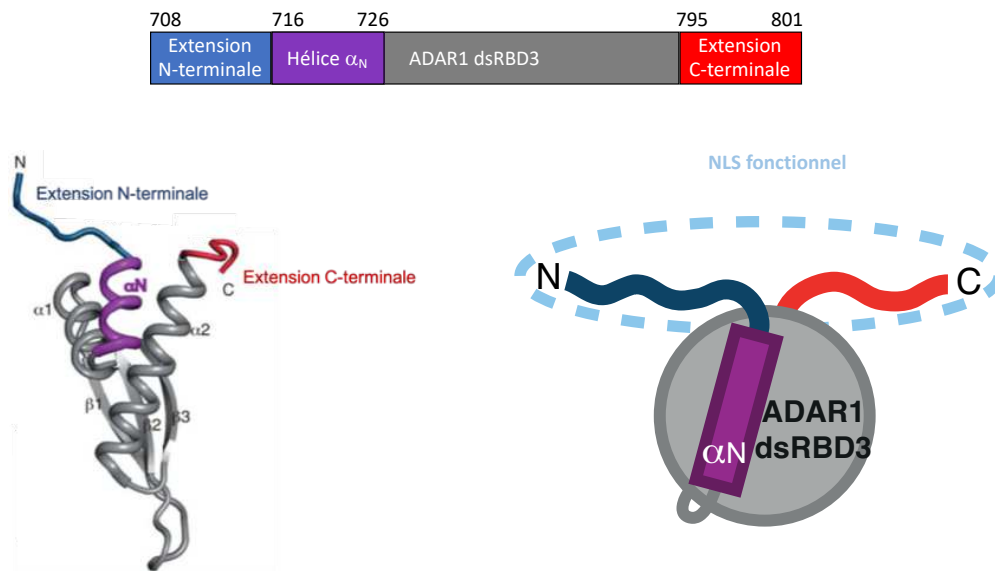


Figure 23 : Structure du dsRBD3 d'ADAR1 portant son NLS bimodulaire.

(**Gauche**) Représentation en cartoon de la structure du dsRBD3. Le dsRBD3 présente une topologie classique des dsRBD, c'est-à-dire un feuillet β à trois brins qui se loge contre les deux hélices α N-terminale et C-terminale (en gris). En plus, le dsRBD3 présente une hélice α supplémentaire en N-terminal (en violet). Ce dsRBD3 forme un NLS bimodulaire qui est constitué des deux extensions N-terminale (en bleu) et C-terminale (en rouge). (**Droite**) Représentation schématique de la structure du dsRBD3. Le dsRBD3 sert de support au NLS bimodulaire, il permet de rapprocher les deux extensions N-terminale et C-terminale formant le NLS dans une conformation fonctionnelle. D'après Banerjee *et al*, Barraud *et al* (164, 174).

De ce fait, le NLS atypique d'ADAR1, formé par les deux extensions flexibles de part et d'autre du dsRBD3, a été qualifié de NLS « bimodulaire » (174). Ces extensions sont essentielles et suffisantes pour l'import nucléaire d'ADAR1, et constituent donc le NLS fonctionnel d'ADAR1 (Figure 23). Ceci a été démontré par l'évaluation des capacités d'import de différentes constructions mutantes et/ou chimériques du NLS d'ADAR1 (174).

L'hélice supplémentaire α_N en N-terminal présente un rôle architectural pour le NLS. En effet, le dsRBD3, avec son hélice supplémentaire, sert de support au NLS atypique d'ADAR1. Il permet de présenter les deux extensions constitutives du NLS dans une conformation fonctionnelle où les deux extensions sont rapprochées du même côté du dsRBD3, ce qui permet alors l'interaction avec la Transportine 1. En effet, l'interaction entre Trn1 et le NLS atypique d'ADAR1 et donc l'import nucléaire sont maintenus, même lorsque le dsRBD3 est remplacé par un autre dsRBD sans relation avec celui d'ADAR1 (174). Le NLS reste également fonctionnel lorsque le dsRBD3 est remplacé par un espaceur (un petit peptide) qui permet de maintenir la conformation fonctionnelle du NLS (164, 174). Ces travaux indiquent que l'interaction entre Trn1 et ADAR1 ne nécessite pas le dsRBD3, mais que ce dernier permet de

mettre les deux extensions formant le NLS du même côté, dans une conformation fonctionnelle. Ceci n'est possible que grâce à la présence de l'hélice additionnelle α_N , car dans le cas d'un dsRBD classique, les extensions flanquantes sont normalement situées de part et d'autre du domaine (Figure 22).

L'activité du NLS d'ADAR1 est dépendante de la présence de l'ARNdb (126, 174). En effet, l'interaction entre le NLS atypique d'ADAR1 et Trn1 est inhibée en présence d'ARNdb, et donc son import nucléaire par Trn1 est aboli. Le modèle actuellement admis pour expliquer comment le NLS atypique d'ADAR1 peut être inactivé en présence d'ARN, et être fonctionnel en absence d'ARN, est présenté dans la sous-partie suivante.

3.2.3.1 Le modèle d'inactivation du NLS en présence d'ARN

L'import nucléaire d'ADAR1 est régulé par la présence d'ARN, potentiellement au niveau du dsRBD3, où se fait également la reconnaissance de Trn1. Cependant, le mécanisme de reconnaissance du NLS d'ADAR1 par Trn1 demeure inconnu. Et les détails moléculaires permettant les interactions entre les deux partenaires ne sont pas connus non plus. Un modèle de reconnaissance d'ADAR1 par Trn1 *via* le NLS bimodulaire dans sa conformation fonctionnelle a toutefois été proposé (174). Ce modèle propose une explication quant à la régulation de l'import d'ADAR1 par Trn1 en présence d'ARNdb.

Dans ce modèle, il est proposé que le dsRBD3 possède deux surfaces d'interaction fonctionnelles distinctes qui ne se chevauchent pas (Figure 24). La première surface d'interaction correspond à l'interface de liaison à l'ARNdb, qui se compose des éléments dits canoniques permettant une interaction canonique entre le dsRBD3 et l'ARNdb. La deuxième surface d'interaction correspond à l'interface fonctionnelle de liaison de Trn1 qui est constituée par les extensions N- et C-terminales qui forment le NLS bimodulaire d'ADAR1 (174) (Figure 24).

Bien que ces deux interfaces soient distinctes, il est proposé que le dsRBD3 ne puisse pas se lier simultanément à l'ARNdb et Trn1. L'interaction entre le dsRBD3, Trn1 et l'ARNdb ne peut pas se faire en même temps, et ce, probablement à cause des problèmes d'encombrement stérique. La reconnaissance simultanée des deux partenaires du dsRBD3 au niveau de leur surface d'interaction respective serait rendue impossible du fait d'incompatibilités stériques. Ce qui expliquerait comment le NLS atypique d'ADAR1 peut être inactivé en présence d'ARNdb.

La liaison d'ADAR1 avec des ARN par le biais de ses dsRBD (dsRBD1-3) favorise son export nucléaire *via* l'exportine-5. Ceci semble indiquer qu'ADAR1 quitte probablement le noyau en étant complexée avec des ARNdb substrats de l'exportine-5, comme les précurseurs des micro-ARN (170, 176). La liaison dans le cytoplasme, de Trn1 à ADAR1 permettrait ainsi la libération de l'ARNdb et *in fine* l'import d'ADAR1 sans ARN dans le noyau par Trn1 (126). Ceci permettrait d'empêcher le ré-import d'ADAR1 liée à des substrats ARN dont la fonction est cytoplasmique, en d'autres termes, d'empêcher des cycles d'export et de ré-import non productifs.

Le modèle d'inactivation du NLS en présence d'ARN repose sur deux hypothèses. La première est que le NLS atypique d'ADAR1 est reconnu par Trn1 au niveau des mêmes sites d'interaction où sont également reconnus les PY-NLS au sein de l'arche C-terminale (Figure 13). La deuxième hypothèse est que la reconnaissance de l'ARNdb se fait de façon canonique au niveau des trois régions d'interactions décrites comme intervenant dans la reconnaissance canonique des ARNdb par les dsRBD (Figure 22).

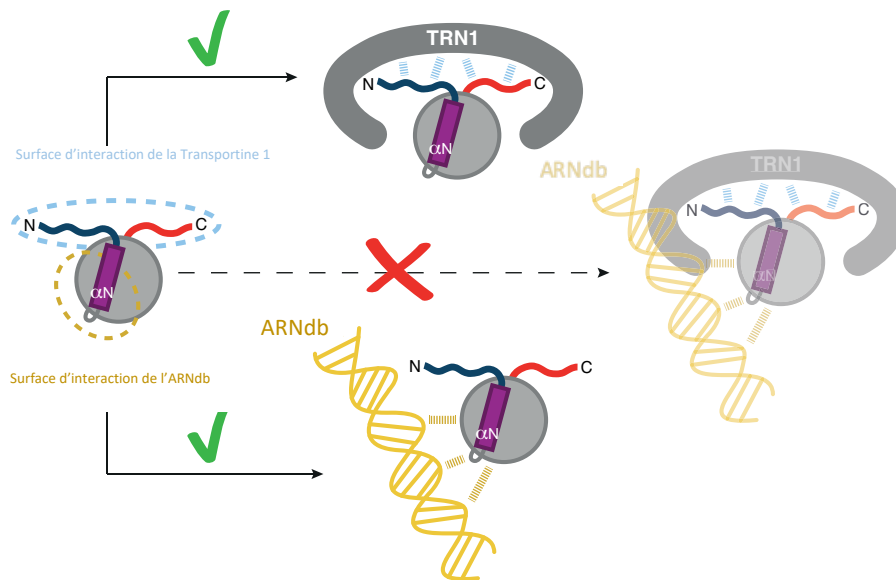


Figure 24 : Le modèle d'inactivation du NLS d'ADAR1 en présence d'ARN.

Schéma illustrant le modèle d'inactivation du NLS d'ADAR1 en présence d'ARN. Ce modèle propose que le dsRBD3 d'ADAR1 possède deux surfaces d'interaction distinctes. Une première surface d'interaction avec l'ARNdb et une deuxième surface d'interaction avec la Transportine 1 (Trn1). Cependant, bien que les deux surfaces soient indépendantes l'une de l'autre, ADAR1 ne peut pas interagir simultanément avec Trn1 et l'ARNdb. Ceci à cause d'incompatibilités stériques entre l'ARNdb et Trn1. D'après Barraud *et al* (174).

3.2.3.1.1 La reconnaissance d'ADAR1 par la Transportine 1

Dans le modèle d'inactivation du NLS en présence d'ARN (Figure 24), il est admis que la liaison du NLS d'ADAR1 avec Trn1 se fait au niveau des mêmes sites de fixation que ceux utilisés par les PY-NLS. Cette première hypothèse repose sur des analyses de mutations et d'alignement de séquences du NLS d'ADAR1 avec certains PY-NLS dont les structures ont déjà été résolues (174).

Deux résidus de Trn1 (les tryptophanes W460 et W730) décrits comme importants pour la reconnaissance des PY-NLS ont été mutés en alanine. Ces mutants W460A et W730A ont perdu leur capacité à importer le NLS minimal d'ADAR1 dans le noyau. Ceci a suggéré que le NLS d'ADAR1 était probablement reconnu au niveau du même site de fixation où sont reconnus les PY-NLS. De plus, à l'aide d'un alignement de séquences de différents PY-NLS dont les structures en complexe avec Trn1 ont été résolues, un modèle structural de l'interaction entre Trn1 et le NLS minimal d'ADAR1 a été construit en plaçant manuellement le NLS minimal d'ADAR1 dans la poche d'interaction des PY-NLS (174). Dans cette modélisation, la position de certains résidus en interaction avec Trn1 a été discutée au regard de la description de la reconnaissance des PY-NLS en termes d'épitopes. Ainsi notamment, il a été proposé que le résidu Arg801 qui est essentiel pour l'import nucléaire d'ADAR1 forme avec le résidu qui le précède un motif ⁸⁰⁰ER⁸⁰¹ qui mimerait l'épitope 3 des PY-NLS, c'est-à-dire le motif PY (voir partie 2.3.2.1.1, en page 27) (174).

3.2.3.1.2 La reconnaissance des substrats ARNdb par ADAR1

Dans le modèle d'inactivation du NLS en présence d'ARN (Figure 24), il est admis que la reconnaissance des ARNdb par le dsRBD3 se fait selon un mode canonique, et que cette interaction canonique ne serait pas possible pour des raisons d'incompatibilités stériques, lorsque le NLS d'ADAR1 est lié à Trn1. En effet, si l'ARNdb est reconnu de manière canonique par le dsRBD3, alors il n'est pas possible pour Trn1 et l'ARNdb de se lier en même temps sur le dsRBD3. Ceci a été visualisé grâce au modèle structural de l'interaction entre Trn1 et le NLS minimal d'ADAR1 construit manuellement (174).

Le modèle d'inactivation du NLS en présence d'ARN repose ainsi sur des données indirectes et des données de modélisation. Toutefois, il est important de conforter ou d'infirmer ce modèle par l'apport de données expérimentales solides et de preuves directes. Il faudrait pour cela montrer que d'une part, la Transportine 1 reconnaît bien le NLS bimodulaire d'ADAR1 au

niveau des poches où sont reconnus les PY-NLS, et que d'autre part, la reconnaissance de l'ARNdb par le dsRBD3 d'ADAR1 se fait bien selon un mode canonique.

Objectifs de la thèse et organisation du manuscrit

Après cette introduction axée sur le transport nucléocytoplasmique qui a permis de décrire les différents acteurs impliqués dans l'import nucléaire de la protéine humaine ADAR1, je vais maintenant présenter les objectifs des travaux de la thèse ainsi que l'organisation du manuscrit.

L'import nucléaire de la protéine ADAR1 est un mécanisme qui est encore aujourd'hui mal compris d'un point de vue moléculaire. Comme vu dans la partie introduction, la protéine ADAR1 est importée dans le noyau par la Transportine 1 (Trn1). Trn1 reconnaît ADAR1 au niveau de son NLS atypique et l'import d'ADAR1 est impacté par la liaison de l'ARN. En présence d'ARN, l'import nucléaire d'ADAR1 est fortement diminué. Bien qu'il soit connu que Trn1 constitue le récepteur d'import nucléaire d'ADAR1, les détails moléculaires de la reconnaissance d'ADAR1 par Trn1 ne sont pas clairement identifiés, d'autant plus qu'ADAR1 ne présente pas les caractéristiques des PY-NLS. Aussi, le mécanisme de régulation de l'import nucléaire en présence d'ARN est également mal caractérisé d'un point de vue structural.

C'est dans ce contexte qu'un modèle qui pourrait expliquer la reconnaissance d'ADAR1 par Trn1 et la régulation de cette reconnaissance en présence d'ARN a été proposé, le modèle d'inactivation du NLS en présence d'ARN (174). Cependant, ce modèle reposait uniquement sur des analyses de mutations et des données de modélisation reposant sur des données indirectes, notamment des alignements de séquences. Le but des travaux de la thèse était de pouvoir valider ce modèle grâce à des données expérimentales. Les expérimentations réalisées devaient permettre d'apporter une compréhension fine au niveau moléculaire du mécanisme d'import d'ADAR1 par Trn1, et notamment sur la reconnaissance de son NLS atypique, et comment cette reconnaissance est régulée en présence d'ARN.

Le premier objectif était de caractériser l'interaction entre la Transportine 1 et ADAR1. Pour cela, nous avons entrepris de résoudre la structure cristallographique de Trn1 en complexe avec ADAR1. Le but était d'analyser le site de fixation d'ADAR1, et de vérifier si la reconnaissance d'ADAR1 se fait au niveau du site d'interaction où sont reconnus les PY-NLS. À défaut d'avoir pu résoudre la structure cristalline du complexe Trn1/ADAR1, nous avons pu apporter un meilleur éclairage sur le mode de reconnaissance du NLS atypique d'ADAR1 grâce à une combinaison de techniques biophysiques. Les techniques de SAXS et de RMN nous ont permis de montrer qu'ADAR1 ne présente pas exactement le même mode de reconnaissance que les

Objectifs de la thèse et organisation du manuscrit

PY-NLS. Notamment, la reconnaissance d'ADAR1 implique la boucle H8 de Trn1 qui est connue pour être impliquée dans le relargage des PY-NLS.

En parallèle de l'étude des mécanismes d'import d'ADAR1, nous avons mené une étude visant à caractériser les mécanismes de reconnaissance d'autres substrats de Trn1 ne portant pas de motif PY, les non-PY-NLS. Pour cela, différents substrats non-PY-NLS reconnus par Trn1 avaient été choisis. L'objectif était de résoudre les structures cristallographiques des complexes de Trn1 avec ces différents substrats. Les différents travaux menés dans cette étude n'ont pas permis d'obtenir de structures cristallographiques, ils ont par contre permis une caractérisation physico-chimique des différents complexes étudiés.

La première partie de ce manuscrit traite du chapitre « Matériels et Méthodes » dans lequel sont présentés les principes des différentes techniques biophysiques qui ont permis la résolution des différentes structures d'une part et qui ont permis d'apporter des informations pour une meilleure compréhension du mécanisme de reconnaissance d'ADAR1 par Trn1 d'autre part. Il s'agit notamment de la cristallographie, du SAXS et de la RMN. Dans cette partie, les différents protocoles d'expression et de purification des différentes protéines ou NLS étudiés durant la thèse sont également présentés. La suite du manuscrit présente d'une part, les résultats obtenus sur l'étude de la caractérisation du complexe Trn1/ADAR1, et d'autre part, l'étude de la caractérisation d'un ensemble de non-PY-NLS, de différents complexes Trn1/non-PY-NLS.

Le deuxième objectif était de mieux comprendre la régulation de la reconnaissance d'ADAR1 par Trn1 en présence d'ARN. Pour ce faire, nous avons entrepris de résoudre la structure cristallographique d'ADAR1 en complexe avec un ARNdb. Le but était d'analyser le mode de reconnaissance de l'ARNdb par ADAR1, de vérifier si cette reconnaissance se fait par un mode canonique. La résolution de la structure cristallographique du dsRBD portant le NLS d'ADAR1 (dsRBD3) en complexe avec un ARNdb a montré que ce domaine dsRBD3 se lie à l'ARNdb de façon canonique sans réarrangement structural majeur. De plus, cette structure cristalline suggère que l'hélice supplémentaire α N du dsRBD3 d'ADAR1 est impliquée dans la reconnaissance de l'ARNdb.

La méthode biophysique utilisée pour la caractérisation du complexe ADAR1/ARNdb, la cristallographie, est présentée dans le chapitre « Matériels et Méthodes ». Dans cette partie, le protocole d'expression et de purification de la construction d'ADAR1 étudiée est également décrit. Les résultats obtenus sur cette étude sont présentés dans le chapitre « Résultats ».

Matériels & Méthodes

Matériels et méthodes

Ce chapitre présente le matériel et les méthodes utilisés durant la thèse. Il décrit d'une part le matériel utilisé et les techniques qui ont permis sa caractérisation biochimique, et d'autre part les méthodes biophysiques utilisées pour répondre aux différents objectifs de la thèse. En effet, au vu des objectifs de la thèse qui consistent à résoudre des structures de différents complexes, j'ai eu recours à différentes techniques biophysiques en fonction des complexes à caractériser. De ce fait, j'ai commencé par des étapes biochimiques qui consistent en l'obtention de protéines purifiées. Les protocoles d'expression et de purification des différents partenaires entrant en jeu dans la formation des différents complexes ont été mis au point. Le but étant d'optimiser le rendement final et le degré de pureté des différents partenaires. Les protéines sont surexprimées de manière recombinante dans *E. coli* selon les différents protocoles décrits ci-après.

I. La reconnaissance d'ADAR1 par Trn1

Cette première partie porte sur la reconnaissance de la protéine d'édition d'ARN ADAR1 par Trn1. Pour mettre en évidence et caractériser l'interaction entre Trn1 et ADAR1, nous avons entrepris l'expression et la purification de deux constructions de Trn1, la version longue et complète appelée Trn1FL (résidus M₉-V₈₉₈), et la version dont la boucle acide H8 est délétée, appelée Trn1ΔH8 (résidus M₉-V₈₉₈ dont les résidus P₃₃₇-D₃₆₇ sont remplacés par l'espaceur GGSGGSG (Figure 25). Concernant la protéine ADAR1, nous avons travaillé avec uniquement le domaine dsRBD3 de la protéine qui porte l'activité d'import nucléaire *via* son NLS bimodulaire reconnu par Trn1. Nous avons réalisé l'expression et la purification de deux constructions du dsRBD3 d'ADAR1 : une construction du dsRBD3 portant le NLS minimal d'ADAR1 (décrit précédemment) appelée ADAR1court (résidus M₇₀₈-R₈₀₁) et une construction du dsRBD3 portant le NLS minimal d'ADAR1 mais cette fois ci avec vingt résidus en amont et seize résidus en aval du NLS minimal appelée ADAR1long (résidus S₆₈₇-R₈₁₇) (Figure 25).

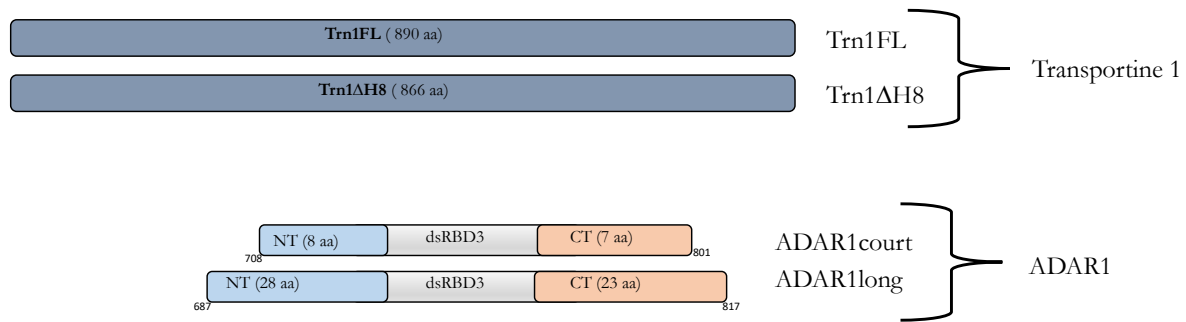


Figure 25 : Les constructions de Trn1 et ADAR1.

Schéma des différentes constructions utilisées pour la caractérisation du complexe Trn1/ADAR1. Trn1 est étudiée sous deux formes Trn1FL et Trn1ΔH8. ADAR1 est aussi étudiée sous deux formes ADAR1court et ADAR1long.

1 Obtention du matériel

L'expression et la purification des deux constructions de Trn1, Trn1FL et Trn1ΔH8, avaient déjà été réalisées à mon arrivée au laboratoire (62, 87). Cependant, il a tout de même été nécessaire d'optimiser les protocoles suivis afin d'augmenter le rendement de la production des deux constructions notamment celui de Trn1ΔH8. Ceci a été possible grâce à un travail de biochimie. On peut noter que la purification de Trn1 est assez contraignante si on considère le grand nombre de cystéine qu'elle compte dans sa séquence (26 cystéines), ce qui perturbe sa stabilité et la rend propice à l'agrégation. Grâce à l'étiquette GST, la purification des constructions de Trn1 se fait en trois étapes, elle compte aussi une étape de coupure de l'étiquette.

Le protocole d'expression et de purification d'ADAR1court a été adapté de protocoles déjà existants (126, 174) et mis au point auparavant au laboratoire. Seul le protocole d'expression et de purification de la construction ADAR1long, que j'ai clonée à mon arrivée au laboratoire a été optimisé. En effet, cette dernière construction (ADAR1long) s'est avérée plus sensible à la protéolyse et à l'agrégation que la première construction (ADAR1court). Différentes conditions de tampons ainsi que plusieurs combinaisons de colonnes de purification ont été testées pour arriver à définir le protocole optimal qui sera décrit plus bas.

De l'expression à la purification, le suivi de la pureté des différentes étapes de production est fait par gel d'électrophorèse, SDS-PAGE (Sodium Dodécyl Sulfate-Polyacrylamide Gel Electrophoresis). Et toutes les étapes de purification par différentes techniques de chromatographie ont été réalisées sur un système ÄKTA pure ou ÄKTA purifier (GE Healthcare) à 4° C. Les absorbances à 280 nm (A_{280}) ont été mesurées en utilisant un

spectrophotomètre NanoDrop (NanoDrop 2000, Thermo Scientific). Les concentrations ($[c]$) des différentes constructions de Trn1 et ADAR1 ont été déterminées grâce à la loi de Beer- Lambert ($A_{280} = \epsilon \times l \times [c]$) en prenant comme coefficient d'extinction, la valeur théorique calculée à partir de la composition en acides aminés par ProtParam (ExPASy) : ϵ Trn1FL = ϵ Trn1 Δ H8 = 98780 L.mol⁻¹.cm⁻¹ et ϵ ADAR1long = ϵ ADAR1court = 9970 L.mol⁻¹.cm⁻¹.

Protéines	PM	pI	ϵ	Vecteur	Etiquette
Trn1FL	101721	4,9	98780	pGEX	GST
Trn1 Δ H8	98499	5,0	98780	pGEX	GST
ADAR1court	10845	9,6	9970	pGEX	GST-His ₆
ADAR1long	14245	8,8	9970	pGEX	GST-His ₆

Tableau 4 : Tableau regroupant les différentes constructions de la Transportine 1 et ADAR1.

Les deux constructions de Trn1 portent une étiquette GST et les deux constructions d'ADAR1 portent une étiquette GST-His₆ (PM = masse moléculaire en Da, pI = point isoélectrique, ϵ = coefficient d'extinction molaire en L.mol⁻¹.cm⁻¹, obtenu avec l'outil ProtParam de ExPASy).

1.1 Production de la protéine Transportine 1

La production des deux constructions de Trn1 commence par une étape de surexpression dans des cellules bactériennes *E. coli* suivie d'une lyse cellulaire par sonication puis de différentes étapes de chromatographie afin d'obtenir du matériel suffisamment pur pour s'engager dans des différentes méthodes de biophysiques décrites plus bas.

1.1.1 Surexpression

1.1.1.1 Souches bactériennes, vecteur plasmidique

Le système pGEX (GE Healthcare) a été utilisé pour la surexpression des différentes constructions de Trn1, permettant de travailler avec des protéines de fusion GST. La séquence ADN codant pour la Transportine 1 humaine (Trn1FL ou Trn1 Δ H8) a été clonée dans le vecteur pGEX-6p-1 (étiquette GST en N-terminal avec un site de coupure à la protéase preScission HRV 3C), et son expression est sous le contrôle du promoteur tac inductible à l'IPTG (IsoPropyl- β - D- 1 - ThioGalactopyranoside). Les cellules chimiquement compétentes BL21 (DE3) codon+ RIL d'*E. coli* sont transformées par choc thermique avec le plasmide pGEX- 6p- 1 contenant le gène d'intérêt (Trn1FL ou Trn1 Δ H8). La préparation consiste à mettre 50 μ L de bactéries chimio-compétentes en contact avec 100 ng de vecteur. Puis, un choc

thermique de 45 s à 42°C est réalisé afin de fragiliser les membranes bactériennes, permettant ainsi au vecteur de les traverser. On y ajoute 300 µL de milieu riche SOC (Super Optimal broth with Catabolite repression), et les bactéries sont ensuite incubées à 37°C pendant 45 min afin de relancer leur métabolisme. Les cellules transformées sont étalées sur gélose LB-agar contenant de l'ampicilline à 100 µg/mL (résistance conférée par le plasmide pGEX-6p-1) du chloramphénicol à 25 µg/mL (résistance intrinsèque des cellules BL21 (DE3) codon+ RIL) permettant la sélection des cellules transformées par le plasmide d'intérêt. Les cellules sont ensuite incubées à 37°C pour la nuit. Les transformations faites suivant ce protocole produisent quelques cinquantaines de colonies pour un étalement de 50 µL.

1.1.1.2 Culture cellulaire

L'expression de Trn1FL est réalisée de la manière suivante : de l'étape précédente, une colonie isolée est prélevée et mise en pré-culture dans 100 mL de milieu LB en présence d'ampicilline (100 µg/mL) et de chloramphénicol (25 µg/mL). Cette pré-culture est incubée une nuit à 37°C sous agitation. La pré-culture sert à ensemercer une culture d'un litre de milieu LB additionné des antibiotiques appropriés en partant d'une densité optique initiale à 600 nm (DO_{600nm}) de 0,05. Ces cultures sont incubées à 37°C sous agitation jusqu'à atteindre une densité de bactéries de $DO_{600nm} = 0,5$. A cette densité, la culture est refroidie à 20°C et l'expression de la protéine est induite par l'ajout d'IPTG (0,2 mM final) lorsque la mesure de la DO_{600nm} se situe entre 0,6 et 0,7. La culture est alors poursuivie à 20°C jusqu'à 20 h après l'induction.

L'arrêt de la culture se fait par centrifugation à 5000 g pendant 20 min (rotor JLA 9-1000, centrifugeuse Beckmann-Coulter Averti JE). Les culots sont repris dans 40 mL du milieu de culture puis à nouveau centrifugés à 5000 g pendant 15 min (rotor JA 12). Enfin, les culots sont congelés et conservés à -80°C.

La surexpression de Trn1ΔH8 est réalisée en suivant le protocole ci-dessus sauf que les cultures sont refroidies à 30°C et la surexpression de la protéine se fait par ajout d'IPTG (0,4 mM final) lorsque la mesure de la DO_{600nm} se situe entre 0,6 et 0,7.

1.1.2 Purification

Les deux premières étapes de purification de Trn1FL et Trn1ΔH8 sont exactement les mêmes, seul le tampon utilisé pour la dernière étape est différent : les deux constructions ayant des solubilités différentes, elles ne demandent pas la même quantité de glycérol.

1.1.2.1 Lyse bactérienne

Le culot bactérien est repris dans 20 mL de tampon de lyse (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 20 % (v/v) glycérol, 2 mM TCEP, 0,5 mM EDTA, 0,2 % (v/v) Triton X-100, 1 mM PMSF) supplémenté par une pastille de cocktail EDTA-free d'inhibiteurs de protéases (ROCHE). Les cellules bactériennes sont lysées à froid par sonication (Sonicator VCX 500, Sonics) : 40 % d'amplitude, 30 cycles de 10 sec et pause de 10 sec entre chaque cycle. L'extrait cellulaire est alors centrifugé à 30000 g pendant 45 min à 4°C (rotor JA20) pour éliminer un maximum de débris cellulaires, et pour récupérer dans le surnageant, la fraction soluble. Cette dernière est donc récupérée, puis filtrée sur une membrane de 0,22 µm avant purification.

1.1.2.2 Chromatographie d'affinité

La première étape de purification de Trn1 est une colonne d'affinité pour l'étiquette GST (GSTprep, FF 16/10, 20 mL, GE Healthcare). La colonne est préalablement équilibrée avec le tampon A (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgoAc, 20 % (v/v) glycérol, 1 mM TCEP). La fraction soluble est alors injectée sur la matrice glutathion et grâce à son étiquette GST, Trn1 va se fixer sur la colonne. Après un lavage de 2 volumes de colonne (CV) de tampon A, la protéine est éluée par le passage du tampon B (Tampon A + 20 mM de glutathion) sur 40 mL à 1 mL/min. Les fractions contenant la protéine d'intérêt sont récupérées. On procède ensuite à la coupure de l'étiquette GST avec la protéase preScission HRV 3C (1 mg de protéase pour 200 mg de protéines). La coupure a lieu à température ambiante pendant une nuit sans agitation.

1.1.2.3 Chromatographie échangeuses d'ions

La purification des fractions rassemblées se poursuit par une deuxième étape de chromatographie échangeuses d'ions. Cette étape consiste à se débarrasser de l'étiquette GST coupée de la protéine. Pour cela, on utilise une colonne échangeuse d'anions de type Q sépharose (HiTrap Q HP, 5 mL, GE Healthcare). L'échantillon est injecté sur la résine de la colonne préalablement équilibrée dans du tampon A. Les deux constructions de Trn1 ont un point isoélectrique (pI) autour de 5 (Tableau 4) ($pI(\text{Trn1FL}) = 4,9$ et $pI(\text{Trn1}\Delta\text{H8}) = 5,0$). Et à pH 7,3 (pH du tampon), contrairement à l'étiquette GST ($pI(\text{GST}) = 6,2$), Trn1 se fixe sur la colonne. Après un lavage de 6 CV de tampon A, on effectue un gradient de 0 à 30 % sur 20 CV de tampon C (tampon A + 1 M NaCl) à 1 mL/min, afin de pouvoir éluer la protéine d'intérêt.

1.1.2.4 Chromatographie d'exclusion stérique

Pour la dernière étape de purification de Trn1FL, afin d'éliminer d'éventuels contaminants résiduels, mais surtout de séparer la fraction correctement repliée de la protéine d'intérêt d'une éventuelle fraction agrégée, on effectue un tamis moléculaire (HiLoad 26/60, Superdex 200, Prep Grade, 320 mL, GE Healthcare). La colonne est préalablement équilibrée avec du tampon de transport (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 1 mM TCEP) et un maximum de 10 mL d'échantillon est injecté. En ce qui concerne Trn1ΔH8, la colonne est préalablement équilibrée avec du tampon A. Les échantillons sont concentrés par centrifugation sur une unité de filtration Amicon de 50 kDa (Millipore) jusqu'à une concentration finale d'environ 10 mg/mL et 20 mg/mL pour Trn1FL et Trn1ΔH8 respectivement, puis congelés dans de la carboglace et conservés à -20°C.

Il faut noter qu'entre chaque étape de la purification, l'équivalent de 2 mM de TCEP est ajouté aux fractions contenant la protéine Trn1 (Trn1FL ou Trn1ΔH8) pour éviter la dégradation de la protéine.

1.1.2.5 Analyse par SDS-PAGE

Après chaque étape de purification, les échantillons (10 µL) sont dénaturés à 95°C pendant 5 min en présence de 100 mM β-mercaptoethanol (agent réducteur) et de 0.25 % SDS (détergent anionique) permettant de dénaturer et de charger uniformément les protéines. Les protéines sont séparées en fonction de leurs poids moléculaires sur un gel SDS-PAGE à 10 % et 12 % (acrylamide) pour respectivement Trn1FL et Trn1ΔH8. La migration se fait à 100 V pendant 1h30. Le gel est ensuite coloré dans une solution de bleu de Coomassie puis rincé dans une solution décolorante pour ne révéler que les bandes de protéines.

1.2 Production de la protéine ADAR1

Comme indiqué plus haut, j'ai travaillé avec deux constructions du dsRBD3 d'ADAR1, ADAR1court et ADAR1long. Pour ADAR1long, j'ai utilisé deux versions de cette construction : ADAR1long-A et ADAR1long-B qui diffèrent par la présence d'une étiquette GST pour ADAR1long-A et d'une étiquette 6 x Histidine (His₆) pour ADAR1long-B. La production des constructions d'ADAR1 commence par une étape de surexpression dans des cellules bactériennes *E. coli* suivie d'une lyse cellulaire par sonication puis de différentes étapes de chromatographie ainsi qu'une étape de coupure de l'étiquette.

1.2.1 Surexpression

1.2.1.1 Souches bactériennes, vecteur plasmidique

Pour ADAR1, le système pGEX a aussi été utilisé pour la surexpression des différentes constructions permettant de travailler avec des protéines de fusion GST. La séquence ADN codant pour la protéine ADAR1 humaine (ADAR1court - résidus M₇₀₈-R₈₀₁) a été auparavant clonée dans le vecteur pGEX-6p-1 modifié (étiquette GST-His₆ en N-terminal avec un site de coupure à la protéase preScission HRV 3C) et son expression est sous le contrôle du promoteur tac qui est induit par l'IPTG.

La séquence ADN codante pour la construction ADAR1long-A (résidus S₆₈₇-R₈₁₇) a été clonée par amplification PCR dans le vecteur d'expression modifié pGEX-6p-1 entre les sites de restriction BamHI et XhoI. La construction contient une étiquette GST suivie d'une deuxième étiquette His₆ (GST-His₆) en N-terminal utilisées pour la purification, et un site de coupure à la protéase preScission HRV 3C. Son expression est sous le contrôle du promoteur tac inductible à l'IPTG.

La séquence ADN codante pour la construction ADAR1long-B (résidus S₆₈₇-R₈₁₇) a été clonée par amplification PCR dans le vecteur d'expression pET28 entre les sites de restriction BamHI et XhoI. La construction contient une étiquette His₆ en N-terminal, un site de coupure à la protéase TEV, et son expression est sous le contrôle du promoteur T7 et est induite par l'IPTG.

Les deux constructions d'ADAR1 clonées dans le système pGEX (ADAR1court et ADAR1long-A) sont exprimées dans les cellules chimiquement compétentes BL21 (DE3) d'*E. coli* transformées par choc thermique comme précédemment indiqué, en utilisant uniquement l'antibiotique ampicilline. Les cellules transformées sont ensuite étalées sur gélose LB-agar contenant de l'ampicilline (100 µg/mL) avant d'être incubées à 37°C pour la nuit.

Par contre la construction ADAR1long-B clonée dans le système pET est exprimée dans les cellules chimiquement compétentes BL21(DE3) codon + RIL d'*E. coli* transformées par choc thermique. La différence porte sur les antibiotiques utilisés, le vecteur pET28 porte une résistance à la kanamycine. Les cellules transformées sont donc étalées sur gélose LB-agar contenant de la kanamycine (50 µg/mL, résistance conférée par le plasmide pET28) et du chloramphénicol (25 µg/mL, résistance intrinsèque des cellules compétentes codon + RIL)

permettant la sélection des cellules transformées par le plasmide d'intérêt. Les cellules sont ensuite incubées à 37°C pour la nuit.

1.2.1.2 Culture cellulaire

L'expression d'ADAR1court et ADAR1long-A se fait comme suit : une colonie isolée obtenue de l'étape précédente est prélevée et mise en pré-culture dans 100 mL milieu LB en présence d'ampicilline (100 µg/mL). La culture est incubée pendant une nuit à 37°C sous agitation. Cette pré-culture sert ensuite à ensemencer 1 litre de milieu LB en partant d'une DO_{600nm} de 0,1. La production de la protéine est déclenchée par l'ajout de l'IPTG à 0,5 mM (final) lorsque la mesure de la DO_{600nm} est à 0,5 et la culture est alors incubée à 37°C pendant 3 h. En ce qui concerne la construction ADAR1long-B, la pré-culture se fait dans 100 mL de LB complémenté par la kanamycine (50 µg/mL) et le chloramphénicol (25 µg/mL), et la culture est commencée à partir d'une DO_{600nm} de 0,05. La culture est ensuite refroidie à 30°C quand elle atteint une DO_{600nm} de 0,5 et l'induction se fait par l'ajout de l'IPTG à 1 mM (final) lorsque la DO_{600nm} mesurée se situe entre 0,6 et 0,7. L'expression se poursuit sur la nuit à 30°C, puis l'arrêt de la culture se fait par centrifugation à 5000 g pendant 20 min. Les culots sont repris dans 40 mL du milieu de culture puis à nouveau centrifugés à 5000 g pendant 15 min. Les culots sont congelés et conservés à -80°C.

1.2.2 Purification

La purification des deux premières constructions, ADAR1court et ADAR1long-A sont réalisées suivant quasi le même protocole, employant les mêmes colonnes et les mêmes tampons bien que celui d'ADAR1long soit adapté à la sensibilité de cette dernière à la dégradation et à l'agrégation. La construction ADAR1long-B quant à elle présente un protocole adapté à son étiquette et aussi à sa sensibilité à la dégradation.

1.2.2.1 Lyse bactérienne

Le culot bactérien est resuspendu dans 20 mL de tampon de lyse (50 mM Tris-HCl pH 8,0, 1 M NaCl, 20 % (v/v) glycérol, 1 mM DTT, 0,5 mM EDTA, 10 mM Imidazole, 0,2 % (v/v) Triton X-100, 1 mM PMSF) supplémenté par une pastille de cocktail EDTA-free d'inhibiteurs de protéases (ROCHE). Ce tampon de lyse est utilisé pour la construction ADAR1-court. Pour les deux versions d'ADAR1-long, le DTT est remplacé par du TCEP à la même concentration. Puis les cellules bactériennes sont lysées comme indiqué pour Trn1, et le lysat est centrifugé et filtré de la même manière.

1.2.2.2 Chromatographie d'affinité

Les échantillons d'ADAR1court sont purifiés par une première étape de chromatographie d'affinité en utilisant une colonne d'affinité contenant une résine d'agarose sur laquelle sont fixés des ions Nickel (HisTrap, HP, 5 mL GE Healthcare). La fraction soluble est chargée sur la colonne équilibrée au préalable avec le tampon D (50 mM Tris-HCl pH 8,0, 1 M NaCl, 1 mM DTT). Après avoir chargé la fraction soluble, la résine est lavée avec une étape à 4 % du tampon E (Tampon D + 500 mM Imidazole) sur 5 CV afin d'éliminer une partie des contaminants protéiques ayant une faible affinité voire une affinité aspécifique pour les ions Nickel. La protéine fixée sur la colonne grâce à son étiquette His₆ va ensuite être éluée avec un gradient allant de 4 à 100 % de tampon E sur 120 mL à 2 mL/min.

La première étape de la purification des deux versions de la construction d'ADAR1long (ADAR1long-A et ADAR1long-B) s'effectue selon le même processus. La seule différence réside dans les tampons où le DTT est remplacé par du TCEP à la même concentration pour parer à la sensibilité à la dégradation de ces constructions.

Les fractions contenant la protéine la plus pure possible sont récupérées, rassemblées, et diluées pour diminuer les concentrations en NaCl et Imidazole (on passe de 1 M à 50 mM NaCl, et de 500 à 25 mM Imidazole) pour permettre une coupure optimale de l'étiquette et aussi réaliser l'étape suivante.

On procède ensuite à la coupure de l'étiquette GST-His₆ pour les constructions ADAR1court et ADAR1long-A. La coupure de la double étiquette se fait avec la protéase preScission HRV 3C (1 mg de protéases pour 200 mg de protéines), la coupure a lieu à température ambiante sur la nuit.

Pour la construction ADAR1long-B, la coupure de l'étiquette His₆ se fait par la protéase TEV (1 mg de protéases pour 100 mg de protéines), la coupure se fait à 20°C sur la nuit.

La deuxième étape de purification d'ADAR1long-B est également une chromatographie d'affinité mais cette fois-ci, inverse. En effet, une fois l'étiquette clivée, l'échantillon est à nouveau injecté sur la résine de Nickel, et comme elle n'a plus d'étiquette, elle ne se fixe pas sur la colonne, on peut ainsi séparer la protéine de l'étiquette. On récupère donc la fraction non fixée sur la résine.

1.2.2.3 Chromatographie d'échangeuses d'ions / d'interaction hydrophobe

En ce qui concerne ADAR1-court et ADAR1long-A, la deuxième étape de purification consiste à séparer la protéine de sa double étiquette GST-His₆ en utilisant une colonne échangeuse de cations (Resource S, 6 mL, GE Healthcare). Après avoir équilibré la colonne dans le tampon F (50 mM Tris-HCl pH 8,0, 50 mM NaCl, 1 mM DTT), elle est chargée avec l'échantillon obtenu à l'étape précédente. La construction ADAR1court a un pI de 9,6 et est chargée positivement à pH 8,0, elle s'accroche donc à la résine contrairement à la double étiquette. Après un lavage de 5 CV de tampon F, on effectue un gradient allant de 0 à 100 % de tampon D à 1 mL/min sur 20 CV pour l'élution de la protéine.

Malgré son pI de 8.8 (Tableau 4), la construction ADAR1long-A n'est pas séparée de son étiquette après coupure et réalisation de la chromatographie échangeuse de cations. La protéine et l'étiquette sont toutes les deux éluées en même temps.

Dans le cas où l'on souhaiterait garder l'étiquette GST-His₆, la deuxième étape de purification d'ADAR1long-A est une chromatographie d'interaction hydrophobe. Sur une colonne (HiPrep Phenyl, HP 16/10, 20 mL, GE Healthcare) préalablement équilibrée avec le tampon G (50 mM Tris-HCl pH 8,0, 50 mM NaCl, 1 M (NH₄)₂SO₄, 1 mM TCEP), on injecte la fraction contenant la protéine récupérée après la première étape de purification. On réalise ensuite un gradient allant de 0 à 100 % de tampon F sur 6 CV à un débit de 1 mL/min.

1.2.2.4 Chromatographie d'exclusion stérique

La dernière étape de purification est identique pour toutes les différentes constructions d'ADAR1. Elle consiste à faire passer l'échantillon récupéré après la deuxième étape de purification sur un tamis moléculaire (HiLoad 26/60, Superdex 75, Prep Grade, 320 mL, GE Healthcare) pour améliorer la pureté de l'échantillon. La colonne est préalablement équilibrée avec le tampon de transport, et un maximum de 10 mL d'échantillon sont injectés. Les échantillons sont concentrés par centrifugation sur une unité de filtration Amicon de 10 kDa pour les constructions ADAR1long et 3 kDa pour ADAR1court. Ils sont ensuite congelés dans la carboglace et conservés à -20°C.

1.2.2.5 Analyse par SDS-PAGE

Les étapes de purification sont suivies par électrophorèse. Entre chaque colonne de purification, les échantillons (8 µL) sont dénaturés à 95°C pendant 5 min en présence de 100 mM

β - mercaptoéthanol et de 0,25 % SDS. Les protéines sont séparées en fonction de leur masse sur un gel SDS-PAGE à 14 % (acrylamide).

1.3 Préparation du peptide ADAR1_pep

Pour augmenter les chances de former le complexe Trn1/ADAR1, nous avons aussi pensé à utiliser un peptide synthétique en plus des protéines recombinantes ADAR1 (ADAR1court et ADAR1long), ADAR1_pep. Il s'agit d'un peptide dont la séquence représente le NLS minimal d'ADAR1 (ADAR1pep :18 aa, séquence : H-**MMPNKVRGEHRQDRKAER**-OH (l'espaceur est en italique) ; PM = 2238.7 g/mol ; pI = 11.35). Le dsRBD3 d'ADAR1 n'est pas présent dans ce peptide, il est remplacé par un espaceur qui garantit la conformation fonctionnelle du NLS (174).

ADAR1_pep est synthétisé chimiquement par la société Eurogentec® d'une pureté >95 % (contrôle qualité par HPLC et MS). On prépare un stock à partir du peptide lyophilisé dans de l'eau pure dégazée. Ensuite, le stock est dilué puis conservé dans le tampon A à une concentration finale de 3 mM afin de faciliter la réalisation des mélanges avec Trn1 lors des différentes expériences et tests. Du stock dilué, sont préparés des aliquots de 20 μ L qui sont congelés et gardés à -20°C.

1.4 Construction et production des protéines chimériques

Toujours dans le but de favoriser la formation du complexe entre Trn1 et ADAR1, nous avons réalisé des protéines chimères en considérant qu'ADAR1 se lie au niveau de l'arche C- terminale de Trn1 comme le suppose le modèle d'inactivation du NLS en présence d'ARN (voir Introduction : paragraphe 3.2.3.1, page 53). Deux protéines chimères ont été élaborées pour la caractérisation du complexe Trn1/ADAR1.

La première chimère (Trn1 Δ H8-ADAR1_pep) est constituée de Trn1 (Trn1 Δ H8) et ADAR1 sous forme de peptide (ADAR1pep), séparés par un espaceur d'une quinzaine d'acides aminés flexibles riche en glycine et sérine. L'extrémité N-terminale du peptide est liée *via* l'espaceur à l'extrémité C-terminale de Trn1. La longueur de l'espaceur a été déterminée en se basant sur l'étude de la structure de Trn1 en complexe avec le PY-NLS de la protéine hnRNP A1 (62). Comme indiqué dans l'introduction, le PY-NLS de cette protéine occupe toute l'arche C- terminale de Trn1. Ainsi, le choix de la longueur de l'espaceur pour la réalisation des chimères est déterminé grâce à l'étude de la reconnaissance de ce PY-NLS (Figure 26). Le peptide d'ADAR1 comprend le NLS bimodulaire minimal composé des extensions N-terminal

et C-terminal reconnu par Trn1. Le dsRBD3 d'ADAR1 n'est pas présent dans ce peptide, il est remplacé par un espaceur qui garantit la conformation fonctionnelle du NLS (174).

La deuxième chimère (Trn1FL-ADAR1long) est constituée de Trn1 (Trn1FL) et d'ADAR1long. Sur le même principe, ADAR1long est greffé au niveau de l'extrémité C-terminale de Trn1FL avec un espaceur de 4 résidus (His-Met-Gly-Gly).

En plus des deux chimères présentées ci-dessus, nous avons aussi construit une chimère comprenant Trn1 (Trn1ΔH8) et un PY-NLS (hnRNP A1) de structure connue, pour évaluer si cette approche pouvait en effet aider à la cristallisation du complexe Trn1/peptide. Il s'agit de la chimère de Trn1 avec le PY-NLS de la protéine hnRNP A1 dont la longueur de la séquence a inspiré la construction de la chimère Trn1ΔH8-ADAR1_pep, on l'appelle Trn1ΔH8-A1_pep (Figure 26).

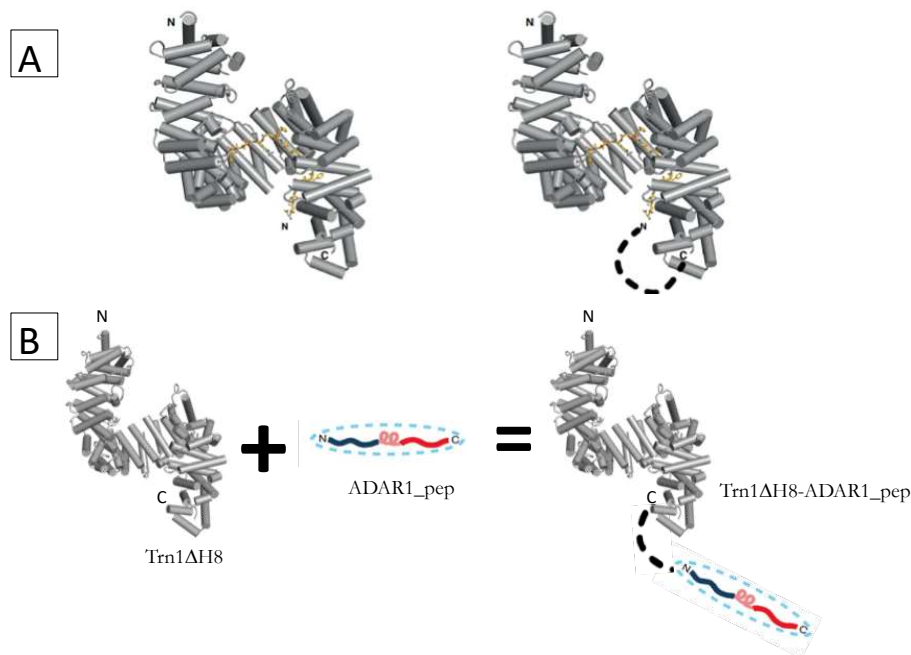


Figure 26 : Schéma de la conception des protéines chimères.

(A) Représentation en cartoon de la structure du complexe Trn1/hnRNP A1 (à gauche) et de la chimère Trn1-A1_{pep} (à droite). L'étude de la structure du complexe Trn1/hnRNP A1 a permis de déterminer la longueur de l'espaceur (en pointillé) utilisé pour la conception de la chimère Trn1ΔH8-ADAR1_{pep}. (B) Représentation en cartoon de la chimère Trn1ΔH8-ADAR1_{pep} composée de Trn1ΔH8 et ADAR1_{pep}. Le peptide ADAR1_{pep} est greffé au niveau de l'extrémité C-terminale de Trn1ΔH8 *via* un espaceur (en pointillé). En bleu l'extension N-terminale et en rouge l'extension C-terminales séparées par le petit espaceur en rose.

1.4.1 Souches bactériennes, vecteur plasmidique

Comme pour Trn1, le système pGEX a été utilisé pour la surexpression des différentes chimères. Pour Trn1 Δ H8-ADAR1_{pep}, deux étapes de mutations ont été nécessaires pour arriver à la formation de la chimère. Les différentes étapes de mutations ont été réalisées par mutagenèse dirigée par amplification PCR selon les protocoles décrits par Liu *et al* (177). Ainsi, une première étape de mutation consistant à supprimer le site EcoRI présent dans la séquence nucléotidique de Trn1 est réalisée pour pouvoir utiliser ce site de restriction lors du clonage de la chimère. Une deuxième étape de mutation a été réalisée pour supprimer le codon stop au niveau de l'extrémité C-terminale de Trn1 pour pouvoir y greffer la séquence de l'espaceur précédant celle du peptide. Une fois les deux mutations effectuées, j'ai ensuite pu procéder au clonage de la chimère. La séquence ADN du peptide ADAR1_{pep} précédée de celle de l'espaceur a été clonée par PCR à partir du plasmide pAM_{nonPY}-NLS (Figure 27) entre les sites de restrictions EcoRI et XhoI dans le vecteur d'expression pGEX-6p-1 (Trn1 Δ H8).

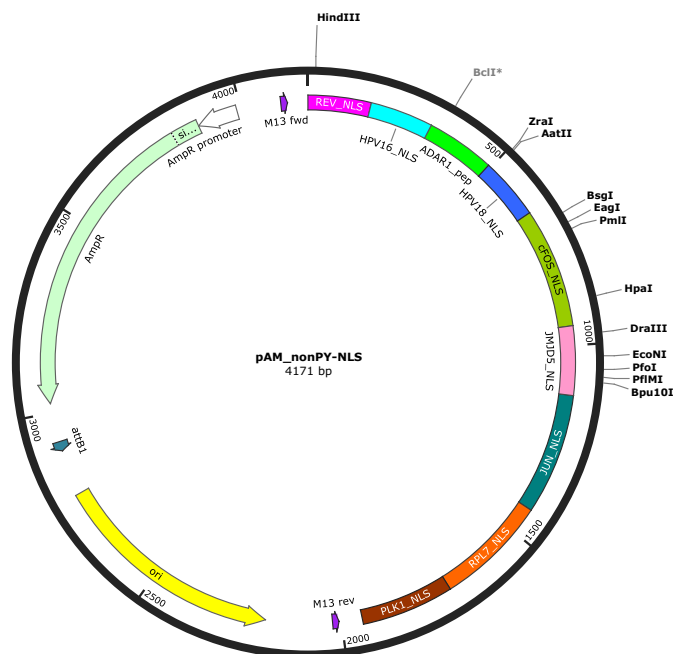


Figure 27 : Carte plasmidique pAM_{nonPY}-NLS.

Le plasmide pAM_{nonPY}-NLS est un plasmide synthétique regroupant les séquences ADN des différents non-PY-NLS étudiés au cours de la thèse. Il possède une cassette de résistance à l'ampicilline.

En ce qui concerne la chimère Trn1FL-ADAR1long, la séquence ADN d'ADAR1long a été clonée par PCR à partir du plasmide pET28 (ADAR1long_S687-R817) entre les sites de restrictions NdeI et XhoI dans le vecteur d'expression pGEX-6p-1 (Trn1FL). Pour cette

chimère une étape de suppression du codon stop remplacé par le site de restriction NdeI a été nécessaire pour la formation de la chimère.

Les constructions chimériques comportent une étiquette GST en N-terminal, un site de coupure à la preScission HRV 3C et leur expression est sous le contrôle du promoteur tac qui est induit par l'IPTG, comme il est le cas pour Trn1 Δ H8 ou Trn1FL. Les chimères sont exprimées dans les cellules chimiquement compétentes BL21 (DE3) codont + RIL d'*E. coli* transformées par choc thermique comme précédemment indiqué, en utilisant les mêmes antibiotiques (ampicilline et chloramphénicol), comme pour Trn1 Δ H8 ou Trn1FL.

Protéines	PM	pI	ϵ	Vecteur	Etiquette
Trn1FL-ADAR1long	116460	5.0	108750	pGEX	GST
Trn1 Δ H8-ADAR1_pep	102855	5.0	98780	pGEX	GST
Trn1 Δ H8-A1_pep	106230	5.0	106230	pGEX	GST

Tableau 5 : Tableau regroupant les différentes constructions chimériques.

Les chimères portent une étiquette GST (PM = masse moléculaire en Da, pI = point isoélectrique, ϵ = coefficient d'extinction molaire en L.mol⁻¹.cm⁻¹, obtenu avec l'outil ProtParam de ExPASy).

1.4.2 Surexpression et purification

Construites à partir des constructions Trn1, l'expression et la purification de ces chimères sont basées sur celle de ces dernières. Les protocoles d'expression et de purification de ces chimères sont ainsi identiques à ceux de Trn1 Δ H8 (pour Trn1 Δ H8-ADAR1_pep ou Trn1 Δ H8-A1_pep) et Trn1FL (pour Trn1FL-ADAR1long) vus plus haut. Les étapes de purification ainsi que les tampons utilisés sont identiques, et ne seront donc pas présentés plus en détail ici.

2 Caractérisation du complexe Trn1/ADAR1

Une fois le matériel obtenu, différentes méthodes sont employées pour la caractérisation du complexe Trn1/ADAR1, d'abord physico-chimiques puis structurales. Différentes techniques physico-chimiques ont été utilisées pour la mise en évidence de l'interaction entre les deux partenaires, Trn1 et ADAR1. Une combinaison de méthodes biophysiques a été utilisée pour la caractérisation structurale du complexe Trn1/ADAR1 en parallèle de la caractérisation par cristallographie aux rayons X.

2.1 Caractérisation physico-chimique des protéines

2.1.1 GST Pull-Down

La technique GST pull-down est une méthode qui permet d'étudier les interactions entre partenaire protéiques *in vitro*. Elle consiste à utiliser des protéines portant une étiquette GST. Tout d'abord, les protéines de fusion GST sont incubées avec de la résine glutathion (constituée de billes de glutathion agarose) pour permettre leur fixation sur les billes d'agarose grâce à l'interaction entre la GST et le glutathion. Ensuite, un lysat protéique, une protéine purifiée ou un peptide synthétique sont incubés avec les complexes formés par les protéines de fusion GST et les billes de glutathion agarose (Figure 28). Après plusieurs lavages successifs, les protéines ayant interagi avec la protéine de fusion GST sont éluées soit de manière compétitive à l'aide de glutathion, soit par des conditions dénaturantes à l'aide du tampon de charge. Une centrifugation permet de séparer le culot de billes du surnageant contenant la protéine de fusion GST en interaction avec le partenaire fixé. L'interaction entre la protéine de fusion GST et le partenaire protéique fixé est mise en évidence après une analyse SDS-PAGE.

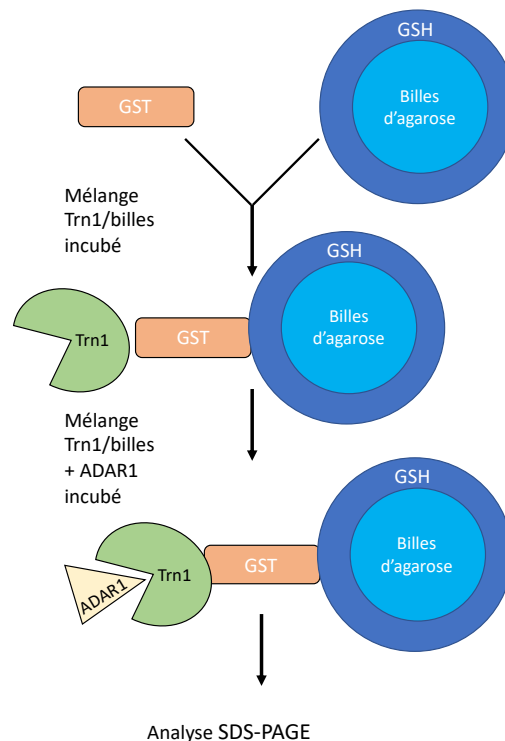


Figure 28 : Principe du GST-Pull Down.

Après expression de la protéine Trn1 Δ H8, celle-ci est purifiée en une seule étape, correspondant à la colonne d'affinité GST (protocole décrit plus haut). La protéine GST-Trn1 Δ H8 est alors fixée sur les billes de glutathion agarose (Glutathion Sepharose 4B, GE HealthCare), elle est

ainsi utilisée pour des expériences de GST pull-down. Le peptide synthétique ADAR1_pep est préparé comme décrit précédemment.

La protéine GST-Trn1 Δ H8 est incubée avec le peptide ADAR1_pep comme suit : 50 μ L de mélange contenant la protéine de fusion GST et le peptide est préparé à un ratio 1:8 (600 μ g de Trn1 soit 0,027 mM et 100 μ g d'ADAR1 soit 0,216 mM) dans le tampon A et incubé sur glace pendant 1 h sous agitation rotative. Ce mélange est ensuite ajouté aux billes de glutathion sépharose qui vont être incubées encore une heure sous agitation rotative à température ambiante cette fois-ci. Après trois lavages avec le tampon A, la protéine et le peptide sont élués dans du tampon de charge puis analysés sur gel SDS-PAGE.

2.1.2 Differential Scanning Fluorimetry (DSF)

Le DSF (Differential Scanning Fluorimetry) aussi connu sous le nom de TSA (fluorescence Thermal Shift Assay) ou ThermoFluor[®] est une méthode qui permet l'étude de la thermostabilité des protéines. C'est une technique basée sur la mesure de l'émission de fluorescence d'un fluorophore qui va interagir de manière non sélective avec la protéine d'intérêt à l'état dénaturé. Cette méthode permet ainsi de déterminer la température de fusion d'une protéine en caractérisant sa stabilité thermique. Le DSF est largement utilisé pour la caractérisation de la stabilité des protéines dans diverses solutions. Par ailleurs, elle permet aussi de caractériser l'interaction d'une protéine avec un ligand, de petites molécules comme des composés chimiques ou des peptides (178). En effet, lors de la fixation du ligand à la protéine, il est possible de mesurer la stabilité thermique de la protéine qui va être modifiée suivant la fixation de ce ligand. Ceci est possible dans le cas où la stabilité thermique de la protéine est modifiée suivant son interaction avec le ligand.

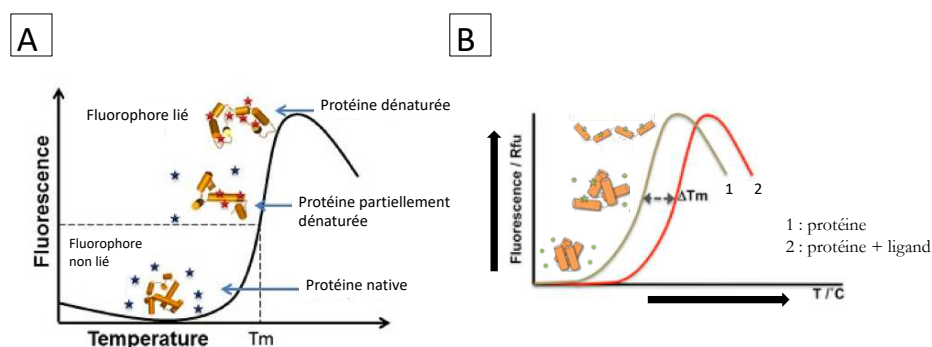


Figure 29 : Principe du DSF.

(A) Courbe de dénaturation thermique d'une protéine. (B) Schéma représentant (1) la courbe de dénaturation thermique d'une protéine libre et (2) la courbe de dénaturation thermique d'une protéine en interaction avec un ligand conduisant à une modification de la thermostabilité de la protéine.

L'utilisation d'agents fluorescents, tels que le Sypro Orange, dont l'émission de fluorescence est sensible à leur environnement permet de suivre l'émission de fluorescence en fonction de la température. Ajouté à une protéine à l'état replié en solution, l'agent fluorescent va présenter une émission de fluorescence faible. L'augmentation de la température au cours du temps entraîne un dépliement de la protéine qui permet l'exposition des régions hydrophobes normalement enfouies à l'état replié. L'agent fluorescent se lie alors aux régions hydrophobes entraînant une augmentation de l'intensité de fluorescence. L'analyse des variations de l'intensité de fluorescence en fonction de la température permet d'obtenir une sigmoïde dont le point d'inflexion permet de déterminer la température de fusion de la protéine ou T_m (melting temperature), c'est la température pour laquelle 50 % de la protéine est sous forme dépliée (Figure 29A).

La valeur de T_m de la protéine peut alors être calculée à l'aide de l'équation de Boltzmann (équation 1) qui décrit les variations de l'intensité de fluorescence en fonction de la température (179). On réalise l'ajustement de la courbe de dénaturation à cette équation grâce au logiciel Prism (GraphPad®).

$$I = A + \frac{(B-A)}{1 + e^{\left(\frac{T_m - T}{\text{pente}}\right)}} \quad (1)$$

où I est l'intensité de fluorescence à la température T ; A et B correspondent aux intensités de fluorescence avant et après la transition ; pente est le facteur de pente.

Lors de l'interaction spécifique d'une protéine avec un ligand, on observe généralement une meilleure stabilité de la protéine. Ceci se traduit par un retard de dénaturation thermique, et donc par une augmentation de la valeur du T_m reflétant un effet de stabilisation de la protéine. La variation du T_m observée (ΔT_m) permet ainsi de caractériser l'interaction moléculaire du ligand et de la protéine (178) (Figure 29B).

J'ai eu recours au DSF pour vérifier l'interaction entre la Transportine 1 et le peptide synthétique ADAR1_pep mais aussi pour vérifier la stabilité des différentes protéines chimériques formées.

Les expériences de DSF ont été réalisées avec un appareil CFX96 Real-Time PCR Detection (BioRad) disponible au laboratoire. Les expériences ont été effectuées dans le tampon A pour les protéines seules (Trn1 Δ H8, Trn1 Δ H8-ADAR1_pep et Trn1 Δ H8-A1_pep) ou en complexe

(Trn1 Δ H8 et ADAR1_{pep}) et dans le tampon de transport pour les protéines Trn1FL et Trn1FL-ADAR1long.

Les protéines testées sans partenaire sont concentrées à 2 μ M (environ 1 mg/mL). Et, dans le cas où l'expérience est réalisée pour vérifier la formation de complexe entre Trn1 et le peptide ADAR1_{pep}, Trn1 est concentrée à 2 μ M et une gamme de concentrations du peptide est testée : de 1 μ M – 40 μ M. Le fluorophore Sypro Orange (Thermo Fischer) est utilisé et est ajouté à une concentration finale de 1X. La température est augmentée par pas de 1°C / min entre 25°C et 90°C et la fluorescence est mesurée tous les 0,2°C.

2.1.3 Chromatographie d'exclusion stérique couplée à la diffusion de lumière multi-angles et réfractométrie (SEC-MALS)

Le SEC-MALS (Size Exclusion Chromatography – Multi Angles Light Scattering) est une méthode qui combine une chromatographie d'exclusion stérique (SEC) à un système de détection multiple comprenant notamment un détecteur d'indice de réfraction : un détecteur de diffusion de lumière à différents angles (22.5-147 °) (MALS) et un détecteur d'absorbance UV. On parle alors de chromatographie d'exclusion stérique couplée à la diffusion de lumière multi-angles, réfractométrie et spectrophotométrie UV-Vis. Cette méthode permet de déterminer la masse moléculaire de macromolécules en solution et par conséquent leur état oligomérique. Le MALS est basé sur deux principes de la diffusion de la lumière. D'abord, la quantité de lumière diffusée est directement proportionnelle au produit de la masse molaire et de la concentration. Puis, la variation d'intensité de la lumière diffusée en fonction de l'angle de mesure est proportionnelle à la taille des molécules. En mesurant donc la quantité de lumière diffusée par l'échantillon ainsi que l'indice de réfraction et en connaissant sa concentration, il est possible de déterminer la masse moléculaire d'une protéine en solution (180).

Cette méthode a l'avantage de pouvoir s'affranchir de la calibration de la colonne puisque la rétention d'une protéine dépend de sa masse moléculaire mais aussi de sa forme et de son affinité pour la matrice de la colonne. Grâce au détecteur de diffusion de la lumière, on peut déterminer le rayon de giration de la molécule en solution. Le spectrophotomètre permet de mesurer l'absorbance UV à 280 nm et le réfractomètre permet de mesurer l'indice de réfraction de la molécule.

Au cours de ma thèse, j'ai fait des expériences de SEC-MALS pour mettre en évidence l'interaction entre Trn1 et ADAR1 et aussi pour vérifier la stœchiométrie du complexe formé entre les deux partenaires.

Les expériences de SEC-MALS ont été réalisées avec un système SEC-MALS disponible au laboratoire. Le système est composé d'une colonne Bio SEC-3 (Agilent, 4 mL) combinée à un détecteur d'indice de réfraction Optilab T-rEX (Wyatt Technology), un détecteur de diffusion de lumière TREOS-miniDAWN (Wyatt Technology) à différents angles (45°, 90° et 135°) et un détecteur d'absorbance, spectrophotomètre UV-Vis SP-20AV (SHIMADZU). Les données sont analysées avec le logiciel ASTRA V (Wyatt Technology). Les échantillons sont concentrés à 50 μ M (environ 5 mg/mL). Pour vérifier la stœchiométrie du complexe formé entre Trn1 et ADAR1, Trn1 est concentrée à 40 μ M et différentes concentrations d'ADAR1 sont testées, allant de 200 μ M à 400 μ M. Pour chaque expérience, 20 μ L sont injectés sur la colonne Bio SEC-3 pré-équilibrée dans le tampon SEC (20 mM Hepes pH 7,3, 55 mM KOAc, 10 mM NaCl, 10 % (v/v) glycérol, 1 mM TCEP). En effet, pour les expériences de SEC-MALS, la purification des protéines est effectuée suivant les protocoles décrits plus haut, puis les protéines sont dialysées contre le tampon SEC.

2.1.4 Chromatographie d'exclusion stérique (SEC) analytique

Des tests d'interaction par chromatographie d'exclusion stérique ont été réalisés durant ma thèse. Toujours dans le but de caractériser l'interaction entre Trn1 et ADAR1, une « co- purification » des deux partenaires a été effectuée. Il s'agit de favoriser l'interaction entre les deux partenaires en les purifiant ensemble, pour ainsi avoir un complexe en sortie de colonne qui pourra ensuite être utilisé pour des tests de cristallisation.

Les expériences sont réalisées sur un tamis moléculaire en utilisant une colonne de type Superdex (Superdex 200 Increase, 10/300 GL, 24 mL, GE Healthcare). Elles ont été menées dans un premier temps dans le tampon A en utilisant la construction Trn1 Δ H8 pour Trn1 et les constructions ADAR1court et ADAR1long d'ADAR1. Différentes concentrations de Trn1 ont été testées ainsi que différentes gammes de concentrations d'ADAR1. Un travail d'optimisation a été effectué pour arriver à la formation d'un complexe 100 % entre les deux partenaires. Ceci sera développé dans la partie résultats.

2.1.5 Calorimétrie par titration isotherme (ITC)

La calorimétrie par titration isotherme ou ITC (Isothermal Titration Calorimetry) est une technique analytique qui permet de caractériser des interactions moléculaires. L'expérience est basée sur la mesure de variation de chaleur résultant de l'interaction entre une molécule et son ligand. Lors de l'expérience d'ITC, le ligand ou molécule titrante est ajouté progressivement à

l'aide d'une seringue à une solution contenant la molécule titrée dans la cellule de mesure. Après chaque ajout du ligand, la quantité de chaleur échangée (exothermique ou endothermique) est évaluée par rapport à une cellule de référence qui contient uniquement le tampon. La cellule de mesure et la cellule de référence sont toutes les deux placées dans une enceinte adiabatique qui permet de maintenir les deux cellules à la même température. S'il y a interaction entre la molécule et le ligand, il se produit alors une libération ou une absorption de chaleur qui crée une différence de température (ΔT) entre la cellule de mesure et la cellule de référence. La chaleur échangée est alors enregistrée sous forme de pics de chaleur lors de l'ajout du ligand. L'intégration de ces pics de chaleur permet d'obtenir une courbe ITC (sigmoïde) à partir de laquelle il est possible de déterminer les paramètres thermodynamiques : la constante de dissociation K_D , la stoechiométrie (N) et la variation d'enthalpie (ΔH).

Les expériences d'ITC ont été effectuées par nos collaborateurs dans le but de mesurer la constante de dissociation à l'équilibre K_D caractérisant le complexe Trn1/ADAR1. La protéine Trn1 est placée dans la cellule de mesure à une concentration de 20 μM , elle est alors titrée par 20 injections (4 μL toutes les 120 s) d'ADAR1 à 160 μM . Les mesures ont été réalisées à 25°C avec l'appareil MicroCal iTC200 (Malvern®) dans le tampon MACARA (20 mM Hepes pH 7,9, 110 mM KOAc, 2 mM MgOAc, 0.5 mM EGTA, 20 mM NaCl, 1 mM PMSF, 1 mM TCEP, 0,1 % (v/v) Tween). Pour les expériences d'ITC, les purifications de Trn1 et ADAR1 ont été effectuées suivant les protocoles présentés plus haut, puis les protéines ont été dialysées contre le tampon MACARA.

2.2 Caractérisation par des études structurales

2.2.1 Études structurales par cristallographie aux rayons X

La cristallographie aux rayons X est une technique classiquement utilisée pour la détermination de structure tridimensionnelle des macromolécules biologiques. La résolution d'une structure tridimensionnelle de protéine par cristallographie nécessite trois étapes. La première consiste à produire du matériel qui doit être homogène et très pur, en quantité suffisante pour faciliter l'interaction des protéines. Ensuite il est indispensable d'avoir un monocristal ayant un bon pouvoir de diffraction, c'est-à-dire la limite de résolution du cristal. Puis, vient la résolution de la structure qui consiste à résoudre le « problème de phases » et à réaliser l'affinement du modèle.

2.2.1.1 Cristallisation des protéines

Le principe de la cristallisation est présenté ici pour une protéine, mais ce principe reste le même pour toutes les macromolécules biologiques (protéines, ADN, ARN, et leurs complexes). La cristallisation d'une protéine repose sur une transition de phase. Il s'agit de faire passer la protéine d'un état en solution à un état cristallin, ces deux états étant en équilibre. Le phénomène de transition de phase dépend de plusieurs paramètres physico-chimiques qui affectent la solubilité de la protéine (la nature et la concentration du précipitant ; la température ; la pureté, l'homogénéité, la concentration de la protéine...) et peut se diviser en deux étapes qui sont la nucléation ou germination cristalline et la croissance cristalline.

La variation de la concentration du précipitant en fonction de la concentration de la protéine à cristalliser joue un rôle critique dans la formation de cristaux. Ceci peut être résumé par un diagramme de phase qui met en évidence l'existence de plusieurs zones : la zone de non saturation ou de solubilité et la zone de saturation (sursaturation) (Figure 30). La zone de solubilité est la zone dans laquelle la protéine est soluble et dans laquelle il n'y a pas de germination cristalline. Elle est délimitée par la courbe de solubilité qui représente la limite entre la phase soluble désordonnée (limpide) et la phase cristalline ordonnée (cristal ou précipité microcristallin). Par ailleurs, il existe aussi la zone de saturation qui délimite la zone de précipitation et la zone métastable (Figure 30).

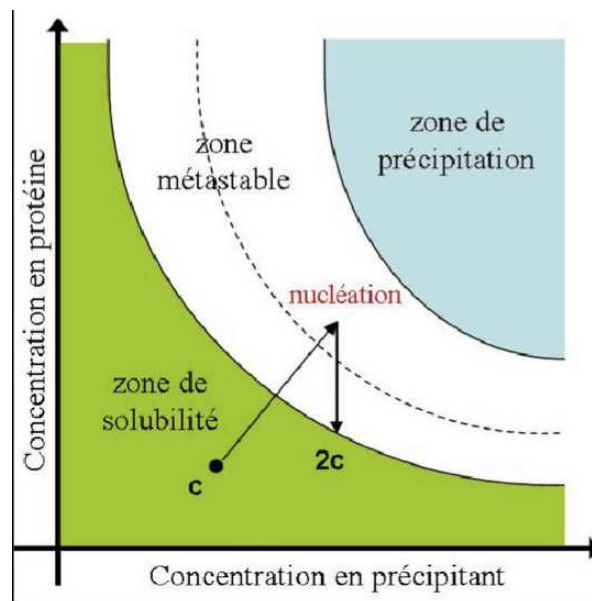


Figure 30 : Diagramme de phase pour la cristallisation des protéines.

Dans la zone métastable, la protéine n'est plus soluble mais plutôt dans un état de saturation qui permet la nucléation, c'est-à-dire l'initiation de la cristallisation. La nucléation est un phénomène qui permet la déstabilisation des interactions protéines-solvant au profit des interactions protéines-protéines. La formation des interactions protéines-protéines non spécifiques conduit à la précipitation (précipité amorphe) tandis que les interactions spécifiques conduisent à la formation de cristaux. Dans la zone métastable, la saturation reste limitée ce qui permet la croissance cristalline. Le cristal croît alors que la concentration en protéine baisse progressivement et tend vers la courbe de solubilité (zone d'équilibre entre l'incorporation de molécules dans le cristal et la dissolution du cristal).

2.2.1.1.1 Méthodes de cristallogenèse

Il existe différentes méthodes de cristallisation des protéines. Chacune permet d'atteindre et de traverser différentes zones de nucléation et de croissance cristalline. On compte la méthode par diffusion de vapeur, par dialyse, par « batch » ou encore par diffusion capillaire.

Au cours de cette thèse, la méthode de cristallisation utilisée est celle de la diffusion de vapeur (181) (en gouttes assises et en gouttes suspendues). Cette technique est basée sur les propriétés thermodynamiques qui imposent aux concentrations en agents précipitants de deux solutions (solution protéique et solution cristalline), enfermées hermétiquement, de tendre vers un équilibre par diffusion de vapeur d'eau ou de solvant.

Concernant la diffusion de vapeur en goutte suspendue, le but est de réaliser une goutte contenant à la fois la solution protéique et la solution de cristallisation (précipitant, additifs, sels...). Pour cela, on mélange à volume égal (1-5 μL) la solution protéique à la solution de cristallisation pour former une microgoutte sur une lamelle. Ensuite, la lamelle est retournée au-dessus d'un réservoir contenant une plus grande quantité de solution cristalline (500 μL – 1 mL). L'ensemble est scellé hermétiquement avec de la graisse ou un joint synthétique. Comme la solution de cristallisation est diluée dans la goutte par la solution protéique, sa concentration est donc plus faible que dans le réservoir. Le retour à l'équilibre des concentrations est atteint lentement par évaporation d'eau (diffusion de vapeur) ou de solvant de la goutte vers le réservoir. Le réservoir sert ainsi à maintenir une pression de vapeur constante. Le volume de la goutte va alors baisser et la concentration de la solution de cristallisation augmenter pour tendre vers celle du réservoir. La diminution du volume de la goutte entraîne en même temps l'augmentation des concentrations en précipitant et en protéines. Ceci va conduire à un état de saturation et par conséquent à la nucléation.

Dans le cas de la goutte assise, le principe reste le même excepté que la goutte est déposée sur une cupule au-dessus du réservoir, l'ensemble est fermé hermétiquement par un film. L'avantage de la goutte assise est que les gouttes peuvent être faites à l'aide d'un robot dans des boîtes 96 puits, ce qui permet de cribler un grand nombre de conditions en utilisant une faible quantité de protéine. On parle de criblage à haut débit (criblage large).

▪ **L'ensemencement avec des micro-cristaux : micro-seeding**

Dans certains cas, l'obtention de cristaux avec un bon pouvoir de diffraction est difficile. Pour y remédier on peut pratiquer l'ensemencement de cristaux ou seeding. Cette technique peut aussi être utilisée dans le cas de cristaux trop petits ou de cristal irrégulier. Le seeding est une méthode qui permet de séparer la phase de nucléation de la croissance cristalline. La technique consiste à utiliser les germes obtenus en forte sursaturation qu'on va ensemencer en faible saturation. Les germes peuvent être soit des monocristaux (macroseeding) ou un broyat de cristaux (microseeding). Puisqu'on maîtrise la quantité de germes ensemencés, cette technique permet ainsi de favoriser la croissance de cristaux de grande taille au profit de microcristaux. De plus, en séparant les phases de nucléation et de croissance cristalline, le seeding permet de faire pousser les cristaux dans des conditions plus variées.

En pratique, après l'obtention d'un échantillon protéique pur et homogène, une condition initiale de cristallisation est recherchée en effectuant un criblage large. Cette condition peut conduire à la formation de microcristaux ou d'un cristal non régulier. Les germes d'ensemencement constituent une matrice à partir de laquelle les molécules vont pouvoir s'assembler conduisant à la formation d'un cristal régulier. Étant donné que, la nucléation demande plus de sursaturation que la croissance cristalline, celle-ci est effectuée dans une solution moins concentrée en agent précipitant que celle dans laquelle le cristal est apparu, ou à des concentrations de protéines plus faibles.

Dans le cas du microseeding, le cristal est prélevé de la goutte, puis broyé dans une solution de force ionique plus faible. Des dilutions en cascades sont ensuite réalisées pour donner la solution d'ensemencement. Après centrifugation du broyat, on va former les gouttes en mélangeant la solution protéique au broyat, et ce mélange est ensuite mélangé avec la solution de cristallisation comme décrit plus haut.

▪ **Le trempage de cristaux : soaking**

Le trempage de cristaux ou soaking est une technique de cristallisation qui consiste dans un premier temps à initier la formation du cristal dans une condition initiale de cristallisation

(solution mère). Une fois le cristal formé, il est alors prélevé et trempé dans une nouvelle solution dans laquelle le cristal reste stable, c'est-à-dire qu'il ne se dissout pas. Cette technique peut servir pour la caractérisation de complexe entre une protéine et un peptide. On commence ainsi par former des cristaux de la protéine, puis le cristal formé est ensuite trempé dans des solutions contenant le peptide. Le peptide va pouvoir aller se fixer au niveau de son site de liaison en passant par des canaux de solvant se trouvant dans le cristal.

2.2.1.1.2 Recherche des conditions de cristallisation

Plusieurs paramètres peuvent influencer la cristallisation des macromolécules biologiques. La mise au point des conditions de cristallisation et leur amélioration sont, encore aujourd'hui pratiquées de manière très empirique et selon la méthode essai-erreur. Néanmoins il existe des kits basés sur des expériences de protéines déjà cristallisées qui permettent de tester un grand nombre de paramètres.

▪ Le criblage

Le criblage consiste à prendre en compte l'ensemble des conditions connues pour la cristallisation des macromolécules biologiques. Il s'agit de conditions ayant conduit à la cristallisation de différentes protéines. Ces conditions sont regroupées dans des kits commerciaux de 96 conditions (Hampton Research, Molecular Dimensions, Qiagen...).

Le criblage permet de voir rapidement les premières tendances à la cristallisation (HIT) des protéines en consommant de faible quantité de matériel tout en testant un grand nombre de paramètres. Dans les meilleurs cas, le criblage permet d'obtenir rapidement les premiers cristaux.

▪ L'optimisation

Dans certains cas, les premiers tests de cristallisation donnent des cristaux ayant une faible qualité de diffraction. Ainsi, une étape d'optimisation est nécessaire. Il peut s'agir d'ajuster finement les conditions dans le but de diminuer le taux de nucléation, d'augmenter la taille des cristaux ou d'améliorer leur aspect (cas des cristaux irréguliers) et d'augmenter l'ordre cristallin. Pour limiter le taux de nucléation ou ralentir la croissance cristalline, on peut réajuster plus finement les conditions « HIT » obtenues lors du criblage large, ou modifier le rapport volumique entre la solution protéique et le réservoir ce qui modifie la vitesse de concentration pour atteindre des niveaux de sursaturation différents. On peut aussi réaliser du seeding (macro ou microseeding), comme vu plus haut.

2.2.1.1.3 Cryocongélation des cristaux

L'une des difficultés à résoudre des structures par cristallographie est le pouvoir de diffraction qui va dépendre entre autres de la qualité du cristal. En effet, en cristallographie, le cristal doit être exposé aux rayons X plus ou moins longtemps pour l'enregistrement des données de diffraction. Or, sous l'effet d'une radiation intense, les molécules d'eau présents dans le cristal peuvent s'ioniser et se dissocier sous forme de radicaux libres (suite au phénomène de radiolyse) ce qui détériore le cristal (altération de l'organisation des molécules dans le cristal) et conduit à une diminution progressive de la qualité de diffraction. Pour y remédier, les cristaux sont congelés rapidement après avoir été prélevés. Effectivement, la congélation des cristaux à des températures inférieures à 140 K ralentit suffisamment le phénomène de radiolyse permettant la collecte des données. En pratique, les cristaux sont prélevés individuellement *via* une boucle en nylon et sont congelés dans l'azote liquide ou le propane liquide (éthane liquide). La congélation des cristaux permet aussi leur transport et leur conservation.

Pour limiter d'avantage le risque de formation de glace autour et au sein du cristal lors de la congélation des cristaux, il est nécessaire de « cryoprotéger » le cristal avec l'utilisation d'un cryoprotectant (PEG, alcool, sucre). L'utilisation d'un cryoprotectant permet d'éviter la formation de glace cristalline au profit d'un état vitreux qui est plus transparent aux rayons X. Généralement, le cryoprotectant est introduit directement dans la goutte avec le cristal mais cela à l'inconvénient de créer un déséquilibre dans la concentration des agents précipitant qui peut conduire à la dissolution partielle ou totale du cristal. Il est donc préférable de transférer le cristal dans une solution cryoprotectrice composée de la solution mère supplémentée de l'agent cryoprotecteur. Pour éviter un écart brutal de concentration, le cristal est brièvement trempé dans une goutte contenant une concentration en précipitant légèrement supérieur à celle du réservoir (environ 110 % de la concentration initiale) ainsi que le cryoprotectant.

2.2.1.2 Diffraction des rayons X : cristaux de molécules biologiques

La cristallographie est une technique physique qu'on peut appliquer aux macromolécules biologiques afin de déterminer leur organisation tridimensionnelle.

C'est l'équivalent d'une microscopie dans laquelle les lentilles, qui permettent de produire l'image d'un objet à partir de la lumière qu'il diffuse, seraient remplacées par une transformation mathématique, la transformée de Fourier (FT), qui permet de calculer cette image à partir de la mesure indirecte de l'intensité diffusée. En effet, il n'existe pas de systèmes

optiques permettant de former une image à partir des faisceaux de rayons X diffractés comme cela est le cas en microscopie électronique.

Pour avoir la structure tridimensionnelle d'une protéine, il faut pouvoir « voir » les atomes qui la constituent. Or, pour observer un objet, il faut utiliser une onde électromagnétique dont la longueur d'onde est de l'ordre de grandeur de cet objet. Ainsi, pour l'observation des atomes et donc de la structure des protéines, on utilise les rayons X dont la longueur d'onde est de l'ordre de l'Angström ($1 \text{ \AA} = 10^{-10} \text{ m}$). Par ailleurs, il est impossible d'obtenir la structure d'une protéine en n'utilisant qu'une seule molécule. Il convient donc d'utiliser un cristal de protéine à l'origine du phénomène de diffraction. Le cristal est soumis à un faisceau de rayons X. La diffusion restreinte à certaines directions de l'espace constitue le phénomène de diffraction. La somme des rayons diffractés se traduit par une tâche noire plus ou moins intense appelée tâche de diffraction. Ainsi, chaque image de diffraction contient un ensemble de tâches. Et, un jeu de données de diffraction est constitué d'une succession d'images qui grâce à différents programmes informatiques permettent en quelque sorte de reconstituer l'image de l'objet formant le cristal.

2.2.1.2.1 Caractéristiques d'un cristal

Un cristal est un solide qui résulte de l'arrangement périodique et ordonné d'un motif composé d'un ensemble de macromolécules se répétant à l'identique d'une maille à l'autre (Figure 31). La maille ou maille élémentaire, constitue la plus petite entité qui permet de retrouver le réseau cristallin en entier par des opérations de translation dans les trois directions de l'espace. Au sein de la maille, on définit l'unité asymétrique qui correspond au plus petit volume à partir duquel la maille peut être construite en appliquant les opérations de symétrie du groupe d'espace auquel appartient le cristal. L'unité asymétrique permet ainsi par les différents opérateurs du groupe d'espace de reconstruire la maille, puis l'ensemble du cristal par translation dans les trois directions de l'espace. L'arrangement des molécules dans le réseau cristallin régulier définit des plans réticulaires. De plus, la cohésion des macromolécules à l'intérieur du cristal est assurée par des interactions de type van der Waals, des ponts salins ou encore des liaisons hydrogènes. L'espace entre les molécules est occupé par le solvant (eau, tampon...) formant des canaux. En général, les canaux de solvant représentent 30 à 80 % du volume cristallin. La présence de ces canaux en grande quantité peut être défavorable puisque cela peut réduire le pouvoir de diffraction du cristal. Néanmoins, grâce à ces canaux, il est possible de pratiquer la

technique de trempage, vue plus haut, en faisant diffuser à l'intérieur du cristal de petites molécules comme des peptides.

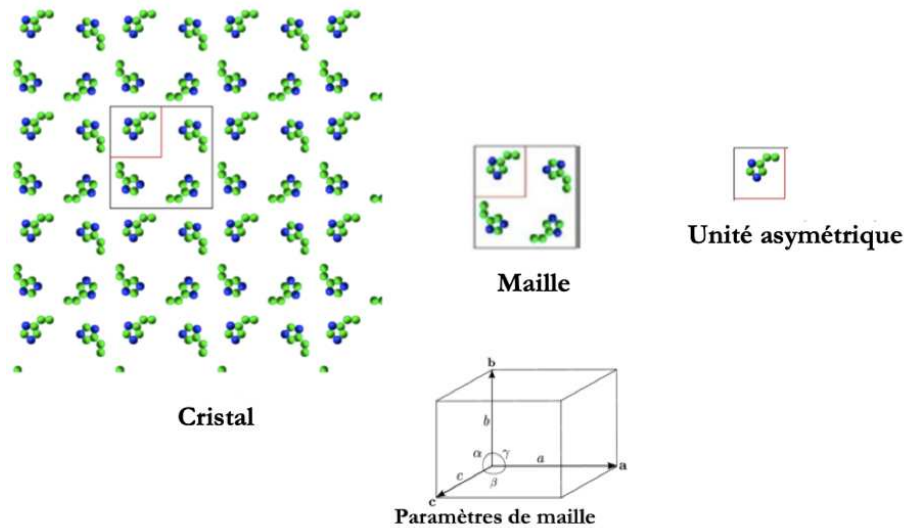


Figure 31 : Composition d'un cristal.

Schéma montrant la composition d'une unité asymétrique, d'une maille et d'un réseau cristallin.

2.2.1.2.2 Systèmes cristallins et groupes d'espace

La maille du cristal est caractérisée par trois vecteurs (**a**, **b** et **c**) et trois angles (α , β et γ) qui définissent les paramètres de maille (Figure 31). Suivant ces paramètres on peut distinguer sept systèmes cristallins différents dont l'arrangement ne peut se faire que selon quatre types de réseaux : Primitif (P), Centré (I), Base centrée (C) et Face centrée (F). Les systèmes cristallins peuvent alors être décrits sous la forme de quatorze réseaux de Bravais (Figure 32).

	Triclinique $a \neq b \neq c$ $\alpha \neq \beta \neq \gamma \neq 90^\circ$	Monoclinique $a \neq b \neq c$ $\alpha = \beta = 90^\circ$ et $\gamma \neq 90^\circ$	Hexagonal $a \neq b \neq c$ $\alpha = \beta = 90^\circ$ et $\gamma = 120^\circ$	Rhomboédrique $a = b = c$ $\alpha = \beta = \gamma \neq 90^\circ$	Orthorhombique $a \neq b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	Quadratique $a = b \neq c$ $\alpha = \beta = \gamma = 90^\circ$	Cubique $a = b = c$ $\alpha = \beta = \gamma = 90^\circ$
Primitif P							
Maille centrée I							
Deux faces centrées C							
Faces centrées F							

Figure 32 : Tableau regroupant les 14 réseaux de Bravais.

A partir des quatorze réseaux de Bravais et des différentes opérations de symétries possibles, il existe 230 possibilités d'empilement pour les objets dans l'espace, il s'agit de groupes d'espace. Certains groupes d'espace ne sont pas compatibles avec les macromolécules biologiques. Ce sont les groupes d'espace comportant des symétries miroir ou des centres d'inversions, qui sont donc incompatibles avec les molécules chirales. Dans le cas des macromolécules biologiques, seulement 65 groupes d'espace différents, sur les 230 groupes d'espace possibles, sont autorisés.

2.2.1.2.3 Phénomène de diffraction

La diffraction met en évidence le caractère ondulatoire de la lumière. La théorie ondulatoire montre que les rayons X peuvent être représentés par des ondes qui interagissent par des interférences constructives et destructives. Lors du phénomène de diffraction, chaque onde diffractée est générée par des interférences constructives entre les ondes diffusées par l'ensemble des plans réticulaires formés par le réseau cristallin. Effectivement, l'arrangement périodique d'un cristal permet d'avoir plusieurs faisceaux diffractés à partir d'un faisceau incident. La somme des faisceaux diffractés dans une direction se traduit par une tâche de diffraction aussi appelées réflexions qui ont chacune une direction et une intensité donnée.

L'intensité diffractée dans chaque direction dépend de la répartition des électrons à l'intérieur de la maille. La symétrie de l'arrangement cristallin impose des relations particulières sur l'intensité et le déphasage de certaines réflexions, on parle de réflexions équivalentes.

L'ensemble des directions de diffraction dépend intégralement de l'orientation des trois vecteurs **a**, **b** et **c** qui définissent la maille cristalline. Le phénomène de diffraction peut être décrit mathématiquement par les équations de Laue et de Bragg.

L'expérience de diffraction par un cristal est seulement observée dans des directions de l'espace qui vérifient les équations de Laue. Ces équations relient le vecteur de diffusion **S** avec le vecteur définissant la maille cristalline (**a**, **b** et **c**).

$$\mathbf{S} \cdot \mathbf{a} = h \quad \mathbf{S} \cdot \mathbf{b} = k \quad \mathbf{S} \cdot \mathbf{c} = l \quad (2)$$

où *h*, *k* et *l* sont des entiers.

Les équations de Laue constituent l'ensemble des directions de diffraction appelé espace réciproque. Cet espace est construit géométriquement à partir du réseau réel.

$$\mathbf{S} = h.\mathbf{a}^* + k.\mathbf{b}^* + l.\mathbf{c}^* \quad (3)$$

Chaque tâche correspond à un nœud du réseau réciproque et peut être indexée par trois indices h , k et l , appelés indices de Miller.

La loi de Bragg est une représentation mathématique du phénomène de diffraction qui définit le cristal comme un ensemble de plans réticulaires sur lesquels un faisceau de rayons X peut se réfléchir.

$$|\mathbf{S}| = \frac{2\sin \theta}{\lambda} = \frac{1}{d_{hkl}} \quad (4)$$

où λ est la longueur d'onde du faisceau qui irradie avec un angle 2θ (angle entre les faisceaux incidents et diffractés) les plans réticulaires parallèles (perpendiculaires à $\mathbf{S}$) d'indice h , k et l et de distance inter-réticulaire $d_{h,k,l}$.

A une tâche d'indice h , k et l dans l'espace réciproque correspond une résolution dans l'espace réel mesurée par d_{hkl} , la distance entre plans réticulaires de la famille de plans donnant lieu à cette réflexion.

2.2.1.2.4 Facteurs de structure

Les ondes diffractées correspondent à la transformée de Fourier de la densité électronique du motif, échantillonné aux nœuds du réseau réciproque. Chacune de ces réflexions est caractérisée par des indices h , k , l et peut être représentée par un nombre complexe $\mathbf{F}_{hkl}$ appelé facteur de structure.

$$\mathbf{F}_{hkl} = |\mathbf{F}_{hkl}| e^{i\varphi_{hkl}} \quad (5)$$

Aussi le cristal peut être représenté comme le produit de convolution d'une fonction motif (la densité électronique contenu dans la maille) et d'une fonction réseau (défini par les paramètres de maille) qui répète le motif dans tout le cristal. Les facteurs de structure de chacune des réflexions h , k , l , peuvent ainsi être reliés à la somme des densités électroniques ρ de coordonnées x , y , z , sur l'ensemble du volume de la maille (V).

$$\mathbf{F}_{hkl} = \int_V \rho_{xyz} e^{2\pi i (hx+ky+lz)} dV \quad (6)$$

Il est ensuite possible de déterminer l'image de la densité électronique par transformée de Fourier inverse.

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l \mathbf{F}_{hkl} e^{-2\pi i (hx+ky+lz)} \quad (7)$$

Ainsi la densité électronique d'une protéine peut être reconstruite à partir de l'ensemble des réflexions enregistrées comme suit :

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |\mathbf{F}_{hkl}| e^{i\varphi_{hkl}} e^{-2\pi i (hx+ky+lz)} \quad (8)$$

Il revient ensuite à mesurer le module $|\mathbf{F}_{hkl}|$ (l'amplitude) et la phase φ_{hkl} correspondante. L'amplitude et la phase doivent être connues pour parvenir à obtenir l'image de la densité électronique. Or, lors d'une expérience de diffraction aux rayons X, on mesure les intensités diffractées I_{hkl} . Il est alors possible d'accéder à l'amplitude des ondes puisque les intensités diffractées sont proportionnelles à leurs amplitudes au carré : $I_{hkl} \propto |\mathbf{F}_{hkl}|^2$. Cependant, l'information de phase est perdue au cours de l'expérience de diffraction. On ne peut donc pas reconstruire directement la carte de densité électronique, d'où ce que l'on appelle : « le problème des phases ». Plusieurs méthodes permettent néanmoins de retrouver les phases et seront abordées plus loin dans le manuscrit.

2.2.1.2.5 Collecte des données de diffraction

En cristallographie des macromolécules le résultat recherché est la structure tridimensionnelle de la macromolécule. Cela implique donc une collecte des données de diffraction dans les trois directions de l'espace. Les lignes de lumière permettent d'enregistrer les intensités diffractées sur des détecteurs bidimensionnels, au cours d'un ensemble de petits pas d'oscillations. Il est ainsi possible de mettre brièvement en condition de diffraction une grande partie des réflexions du réseau réciproque.

Le cristal préalablement congelé dans la boucle en nylon est placé sur un goniomètre qui permet de centrer et d'orienter le cristal sous le faisceau. Sur le goniomètre, le cristal est en permanence soumis à un jet d'azote gazeux refroidi à 100 K pour le maintenir à très basse température et ainsi éviter sa dégradation prématurée. Le cristal est ensuite soumis à un faisceau de rayons X au niveau du goniomètre qui permet un ensemble de petites rotations du cristal, par petit pas allant de 0.1 à 1° au cours de la collecte des données.

Le temps d'exposition et l'intensité du faisceau incident dépend du pouvoir de diffraction du cristal. A chaque pas d'oscillation, les tâches de diffraction sont enregistrées sur le détecteur bidimensionnel donnant le cliché de diffraction. Le jeu de données de diffraction, constitué d'une succession d'images reforme le réseau réciproque.

Les différentes collectes de données des cristaux des protéines étudiées ont été réalisées au synchrotron SOLEIL (Source Optimisée de Lumière d'Énergie Intermédiaire du LURE) de Saint Aubin et à l'ESRF (*European Synchrotron Radiation Facility*) de Grenoble.

2.2.1.3 Traitement et analyse des données de diffraction

Après l'enregistrement des clichés de diffraction, les images sont analysées. L'analyse des images enregistrées suit un protocole de traitement des images selon trois étapes fondamentales.

La première étape consiste à indexer les données, il s'agit de l'**indexation**. La position des tâches de diffraction et l'angle d'oscillation permettent de déterminer les paramètres de maille du cristal et son orientation. Les programmes d'indexation recherchent d'abord les trois vecteurs de base (**a**, **b**, et **c**) qui relient toutes les tâches, ensuite la matrice d'orientation puis la maille. Les indices de Miller (*h*, *k* et *l*) sont déterminés pour chaque tâche afin d'intégrer les intensités correspondantes. A partir de ces paramètres, le logiciel propose le système cristallin le plus probable.

La deuxième étape est l'**intégration** des intensités. Cette étape permet la mesure et la correction des intensités de la dispersion spatiale des rayons de diffraction mais aussi du groupe d'espace. Cette dispersion, principalement due à la divergence du faisceau incident ou à la mosaïcité du cristal entraîne l'étalement des tâches. Lors de l'intégration des intensités, les programmes vont évaluer les profils moyens des tâches de diffraction, et vont ensuite calculer l'intensité moyenne de chaque tâche par intégration numérique.

La troisième étape est la **mise à l'échelle** des intensités. En effet, l'intensité des réflexions théoriquement identiques par symétrie ne sera pas égale du fait de la détérioration du cristal lors de la collecte des données. Pour corriger ces effets, on effectue une mise à l'échelle des intensités. Les programmes de mise à l'échelle déterminent pour chaque cliché de diffraction, des facteurs à partir des réflexions supposées équivalentes par symétrie. Une fois corrigées, les intensités équivalentes sont moyennées et les modules des facteurs de structure sont calculés. Les données ainsi réduites vont permettre d'obtenir l'ensemble des réflexions uniques

nécessaires pour le calcul de l'image de la densité électronique par transformée de Fourier inverse.

Il existe différents programmes pour le traitement des images de diffraction permettant de réaliser ces trois étapes afin de calculer les facteurs de structure. Pour tous les jeux de données collectés durant cette thèse : l'indexation et l'intégration des tâches de diffraction ont été réalisées avec le programme XDS (X-ray Detector Software) (182) en utilisant les scripts XDSME (XDS Made Easy) associés au logiciel XDS et développés par Pierre Legrand (SOLEIL). Le programme AIMLESS de la suite de programmes CCP4 (183, 184) a été utilisé pour la réduction des données.

2.2.1.4 Paramètres du cristal

▪ Le facteur B moyen

Un cristal de macromolécules est composé de solvant. A l'intérieur du cristal, les régions des macromolécules exposées au solvant peuvent présenter une grande variabilité dans leur position. Cette dynamique se traduit par un affaiblissement global de l'intensité du signal diffracté qui peut être pris en compte par un facteur d'agitation thermique global, c'est le facteur B moyen exprimé en $\text{\AA}^2$. Ce facteur varie en général entre 10 et 70 $\text{\AA}^2$ pour des cristaux de protéines. Par ailleurs, ces régions exposées au solvant présentent un faible pouvoir de diffraction. C'est pourquoi il est difficile d'observer la densité électronique dans les régions moins ordonnées.

▪ L'analyse du contenu de l'unité asymétrique

L'unité asymétrique peut contenir une ou plusieurs molécules. Dans le cas où elle contient plusieurs molécules, ces dernières sont alors reliées entre elles par des symétries non cristallographiques (NCS). En utilisant la méthode de Matthews (185), il est possible d'estimer le nombre de molécules de protéines dans l'unité asymétrique. La quantité de solvant dans le cristal ($S\%$) peut aussi être déterminée à partir du poids moléculaire de la protéine cristallisée, des paramètres de maille du cristal et du groupe d'espace.

La méthode de Matthews montre que pour la majorité des cristaux, le rapport (exprimé en $\text{\AA}^3/\text{Da}$) entre le volume de la maille et le poids moléculaire contenu dans celle-ci se situe entre 2 et 4 $\text{\AA}^3/\text{Da}$.

2.2.1.5 Caractéristiques et validation des données

La qualité d'un jeu de données de diffraction est estimée par plusieurs critères statistiques calculés au cours du traitement des données.

La résolution détermine le degré de détail de la densité électronique obtenue par transformée de Fourier inverse. La résolution maximale des données peut être imposée par la distance entre le cristal et le détecteur, ou par la qualité de diffraction du cristal.

La complétude est la mesure du rapport entre le nombre de réflexions uniques mesurées et le nombre total de réflexions mesurables. Elle doit être la plus grande possible (plus de 90 %).

La multiplicité (ou redondance) correspond au nombre moyen de réflexions symétriques mesurées. Plus sa valeur est grande, plus l'intensité moyenne d'une réflexion est précise du fait que les erreurs de mesures sont minimisées.

Le rapport Signal/Bruit ($I/\sigma(I)$) est le rapport de l'intensité des réflexions uniques par rapport à leur écart type correspondant au bruit de fond. Plus le rapport signal/bruit est élevé, meilleure est l'estimation des facteurs de structure.

Le facteur R_{sym} (ou R_{merge}) indique l'écart d'intensité entre les réflexions équivalentes par symétrie. En bio-cristallographie, un facteur R_{sym} de bonne qualité se situe autour de 10 % pour l'ensemble des données.

Le $CC_{1/2}$ est le coefficient de corrélation calculé à partir du pourcentage de corrélation entre deux intensités provenant chacune d'une moitié du jeu des réflexions mesurées divisée aléatoirement. Le seuil de coupure de la résolution est admis pour un $CC_{1/2} \geq 0,3$.

2.2.1.6 Méthodes de phasage : le remplacement moléculaire

Le grand problème en cristallographie est l'incapacité à mesurer directement les phases. Lors d'une expérience de diffraction, seules les intensités diffractées sont enregistrées. Toutefois, pour la construction de l'image par transformée de Fourier inverse, il est nécessaire d'avoir des informations à la fois sur les phases et les intensités. D'autant plus que la qualité de la carte de densité électronique obtenue va essentiellement dépendre de la justesse des phases.

On compte différentes méthodes pour résoudre « le problème des phases ». Ces méthodes peuvent être classées en deux groupes, d'après l'utilisation comme source de phase d'un modèle atomique homologue ou du signal apporté par un atome lourd. Cette dernière est une méthode dite expérimentale qui requiert l'utilisation d'un atome lourd qui peut être ajouté lors de la

production de la protéine. L'ajout peut également se faire par diffusion dans le cristal au moment de la congélation.

La méthode de phasage dont la source est l'utilisation d'un modèle atomique homologue est le remplacement moléculaire. C'est la méthode la plus utilisée aujourd'hui pour la résolution du « problème des phases » puisque le nombre de structures à utiliser comme modèles est de plus en plus important dans la PDB. Cette méthode consiste à positionner un modèle structural préexistant dans la maille du cristal étudié dans le but d'obtenir une première estimation des phases pour les facteurs de structures. La difficulté réside dans le fait de trouver la matrice de rotation et le vecteur de translation qui vont permettre de positionner le modèle dans l'unité asymétrique. La détermination des fonctions de rotation et de translation est réalisée par le biais de fonctions de corrélation qui permettent de comparer les amplitudes mesurées (F_o) et les amplitudes calculées à partir du modèle (F_c).

La première étape du remplacement moléculaire consiste à trouver l'orientation du modèle. La procédure permet de déterminer les trois angles α , β , et γ pour lesquels la fonction de corrélation entre la fonction de Patterson du modèle orienté et celle des données de diffraction est maximale. Après la bonne orientation du modèle, la deuxième étape du remplacement moléculaire consiste à placer le modèle en translation dans la nouvelle maille obtenue.

Toutefois, le succès du remplacement moléculaire est conditionné par le modèle homologue dont le taux de similarité structurale avec la protéine connue doit être assez élevé.

Dans le cadre de ma thèse, les structures des différentes protéines étaient déjà disponibles dans la PDB, résolues soit par cristallographie soit par RMN. Ainsi, nous avons utilisé la méthode du remplacement moléculaire pour le phasage. Pour cela, nous avons utilisé le programme PHASER MR (186) de la suite de programmes CCP4. Ce programme est basé sur des statistiques comme le facteur d'accord Z-score entre la position du modèle et le jeu de données expérimentales, particulièrement le score TFZ. Ce facteur est calculé en comparant le logarithme de vraisemblance (LLG) de la fonction de translation déterminée à celui d'une fonction de translation aléatoire. Plus le score TFZ est élevé, plus la solution en sortie du remplacement moléculaire est probable.

2.2.1.7 Affinement de la structure

L'affinement consiste à minimiser l'écart entre les données observées (données de diffraction) et les données calculées à partir du modèle. L'objectif est que le modèle obtenu par remplacement moléculaire soit le meilleur reflet possible de la structure moléculaire cristallisée.

En sortie du remplacement moléculaire, il est possible d'avoir un modèle approximatif de la structure étudiée. Les facteurs de structure calculés à partir du modèle sont en général à peu près en accord avec les facteurs de structures observés lors de l'expérience de diffraction aux rayons X. Le modèle approximatif obtenu est ensuite progressivement amélioré par une procédure itérative de reconstruction manuelle et de reconstruction par des techniques d'affinements numériques. L'alternance de l'affinement numérique et de la reconstitution manuelle permet de compenser par la reconstruction manuelle les erreurs introduites par le modèle et de compenser par l'affinement numérique les erreurs introduites par la reconstruction manuelle. En corrigeant progressivement les positions atomiques de la protéine étudiée, on peut recalculer des phases plus correctes et ainsi mieux définir la carte de densité électronique.

Pendant le processus d'affinement, le modèle est validé avec les facteurs Rfacteur et Rlibre, qui mesurent la différence entre les facteurs de structure calculés à partir du modèle et les facteurs de structure calculés à partir des données de diffraction :

$$R_{\text{facteur}} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|} \quad (9)$$

Le Rlibre correspond à un Rfacteur, mais calculé sur environ 5 % des réflexions qui sont mises à part et qui n'interviennent pas dans le calcul de la carte de densité électronique utilisée lors de l'affinement. L'écart entre les facteurs Rfacteur et Rlibre est un indicateur de la qualité du modèle, c'est-à-dire de la justesse du modèle obtenu par remplacement moléculaire. Ils varient chacun entre 0 et 1 et doivent diminuer au fil de l'affinement. Plus ils sont faibles, meilleur est l'accord entre le modèle et les données. Ils doivent être d'autant plus bas que la résolution est bonne. Un faible écart (entre 5-7 %) entre ces deux facteurs indique que le modèle est correctement affiné.

Dans le cas d'une surinterprétation des données de diffraction, le nombre d'observations devient alors insuffisant par rapport au nombre de paramètres à affiner. Le facteur Rfacteur aura ainsi tendance à diminuer tandis que le facteur Rlibre aura tendance à augmenter.

2.2.2 Cristallisation et résolution de structure du complexe Trn1/ADAR1

2.2.2.1 Criblage robotisé des conditions de cristallisation

L'institut dispose d'une plateforme de cristallographie, tous les tests de cristallisation y ont été réalisés. Le criblage large des conditions de cristallisation a été réalisé en utilisant la méthode de diffusion de vapeur en gouttes assises avec l'aide d'un robot de cristallisation Mosquito (SPTlabtech). Pour cela, on utilise des plaques Swissci 96 puits, 3 cupules à fond rond facilitant le prélèvement des cristaux dans les gouttes. Les réservoirs des plaques Swissci sont pré-remplies (40 µL/puits) avec les solutions des kits commerciaux de cristallisation couvrant des conditions de cristallisation les plus rencontrées (les kits Index HT et crystal screen de Hampton Research et les kits JCSG plus, MIDAS plus et PACT premier de Molecular Dimension).

Chaque puits contient trois cupules dans lesquelles on peut déposer trois échantillons protéiques composés d'un mélange de différents ratios de protéine et de solution du réservoir. Ainsi, ces plaques permettent de tester jusqu'à 288 (3 x 96) conditions par plaque en faisant varier les différents paramètres qui affectent la solubilité de la protéine. Après la réalisation de la plaque avec le robot Mosquito qui dépose les gouttes de protéine et de réservoir à la suite, chacun des puits est scellé avec un film transparent qui permet de suivre l'évolution des gouttes. Les plaques sont ensuite placées dans le robot de visualisation (Explora Nova– Thermo Scientific – Leica) à 4°C ou 20°C pour favoriser la croissance cristalline. L'évolution des gouttes a été régulièrement suivie pendant plusieurs semaines par observation quotidienne des plaques au microscope à l'aide du robot de visualisation qui prend une photo de chaque goutte. Les photos sont alors analysées grâce au logiciel Xtal Focus (Explora Nova – Imascope), qui permet d'annoter manuellement chaque goutte. Des observations fréquentes permettent de surveiller l'évolution des gouttes et, en cas d'apparition de cristaux, les congeler rapidement avant leur dégradation.

2.2.2.2 Reproduction manuelle et optimisation des cristaux

Si aucune condition des kits commerciaux ne permet d'obtenir un cristal régulier unique suffisamment gros pour être testé au synchrotron, les conditions de cristallisation les plus prometteuses sont affinées en faisant varier les paramètres autour de la condition de cristallisation du kit commercial (concentrations en protéine et en agent précipitant, sels, pH, ajout d'additifs...). Ces optimisations peuvent être réalisées en nano-gouttes en utilisant le robot

pipeteur Dragonfly (SPTlabtech) pour préparer les réservoirs des plaques 96 puits à partir de solutions stock « maison » (agents précipitants, tampons, sels, additifs...) ou alors manuellement. Les conditions de cristallisation ayant donné des cristaux lors du criblage robotisé en nano-gouttes sont reproduites manuellement en microgouttes en utilisant la méthode de la goutte suspendue dans des boîtes VDX 24 puits (Hampton Research) avec des solutions faites à la main. Les optimisations manuelles entraînent l'utilisation de gros volumes ce qui permet d'avoir des cristaux plus gros. Par ailleurs, les plaques 24 puits ne sont pas prises en charge par le robot de visualisation, les plaques sont placées dans des placards à 20°C. L'évolution de la goutte est surveillée manuellement, et les gouttes sont observées régulièrement au microscope (Leica).

Quand même après optimisation, seuls des microcristaux ont pu être obtenus, des tests de microseeding ont été réalisés pour améliorer leur taille. Des gouttes fraîchement équilibrées contre une solution d'agent précipitant sont inoculées avec différentes dilutions (typiquement 1/1.000 et 1/1.000.000) de germes cristallins broyés puis resuspendus dans la solution de cristallisation.

De plus, pour l'amélioration du pouvoir de diffraction des cristaux, des tests de déshydratation ont été menés. Le but ici est de diminuer le taux de solvant des cristaux et par conséquent la dynamique des molécules à l'intérieur de ces derniers. Pour cela, les cristaux ont graduellement été transférés vers des solutions de plus en plus concentrées en précipitant dans lesquelles ils sont incubés de 6 h à 24 h (187). Une autre façon de faire, consiste à laisser sécher la goutte de cristallisation quelques minutes à l'air libre, juste avant de congeler les cristaux (188).

Pour leur cryoprotection, les cristaux sont transférés dans une solution cryoprotectante. En général, la solution de cristallisation est supplémentée de 20 à 30 % de glycérol, avant de congeler les cristaux dans l'azote liquide où ils seront conservés jusqu'à ce qu'ils puissent être amenés et testés au synchrotron.

2.2.2.3 Préparation de la solution protéique

Après obtention d'échantillon protéique suffisamment homogène, pur et concentré, des tests de cristallisation sont réalisés.

D'abord, les protéines seules de Trn1 (Trn1FL et Trn1ΔH8) ont été cristallisées par criblage large robotisé en testant différents kits (Index HT, JCSG plus, MIDAS plus, PACT premier et crystal screen), différentes concentrations de protéines (0,10 – 0,20 mM) et différents ratios de

gouttes (de 0,1 + 0,1 μ L à 1,5+ 1,5 μ L). Après l'obtention de cristaux reproductibles de la protéine seule, des tests de cristallisation du complexe Trn1/ADAR1 ont été menés.

Ensuite, des tests de co-cristallisation ont été réalisés en mélangeant les deux partenaires sur glace pendant 1 h avant la réalisation des plaques. Le mélange des deux partenaires est fait en ajoutant le partenaire ADAR1 à différents ratios (de 1:1 à 1:7). Les premiers tests de co-cristallisation ont été réalisés avec les protéines recombinantes ADAR1court et ADAR1long à des ratios allant de 1:1 à 1:5. Puis, les tests ont été réalisés avec le peptide ADAR1_pep à des ratios allant de 1:1 à 1:7. Le peptide ADAR1_pep a aussi été utilisé pour des expériences de trempage de cristaux décrit au paragraphe 2.2.1.1.1 (page 82). Pour cela, des cristaux de la protéine Trn1 seule ont été préparés, puis ces cristaux ont été trempés dans des solutions contenant le peptide à différentes concentrations (0,15 -1,05 mM).

Enfin, les tests de cristallisation de la protéine chimérique Trn1FL-ADAR1long ont été réalisés comme décrits pour les protéines de Trn1 seules à différentes concentrations (0.05 – 0,07 mM soit 5 – 7 mg/mL).

Après « co-purification » du complexe Trn1/ADAR1, des tests de cristallisation du complexe obtenu ont été menés en se basant sur ceux réalisés avec la chimère Trn1FL-ADAR1long.

2.2.2.4 Collecte des données et résolution de la structure cristallographique

Pour chaque cristal testé, 4 images sont enregistrées à 90° d'intervalle pour estimer la qualité de la diffraction. La distance cristal-détecteur est ajustée en fonction du pouvoir de diffraction du cristal. Lorsque le cristal présente un bon pouvoir de diffraction, les données sont analysées par le programme EDNA (ESRF) ou XDS (SOLEIL) qui détermine automatiquement les paramètres de maille et les symétries du cristal afin de proposer une stratégie de collecte (temps d'exposition, oscillation, nombre d'images et angle de départ). Ceci permet d'obtenir un jeu de données complet et de bonne qualité en un minimum d'images. Les jeux de données ont été collectés soit sur les lignes Proxima-1 et Proxima-2 (SOLEIL) soit sur les lignes ID30A-3, ID30B et ID23-1 (ESRF).

Les scripts XDSME associés au logiciel XDS ont été utilisés pour le traitement de tous les jeux de données. Dans un premier temps, le programme analyse les réflexions en discriminant les tâches de diffraction du bruit de fond. A partir des réflexions identifiées, le programme propose ainsi des systèmes cristallins compatibles avec l'organisation du cristal. Les solutions proposées sont classées en fonction d'une pénalité qui rend compte des écarts entre positions prédites et

observées. Le système cristallin choisi présente la plus haute symétrie et la plus faible pénalité. Après l'identification du groupe d'espace et des paramètres de maille, les indices de Miller sont attribués pour chacune des tâches de diffraction. Ensuite, les intensités I_{hkl} sont mesurées pour chaque tâche lors de l'intégration des données. Puis, après la mise à l'échelle des intensités, la moyenne des réflexions équivalentes par symétrie est calculée et le jeu de données est réduit à un jeu de réflexions uniques. C'est à cette étape que la résolution maximale des données est ajustée en fonction de la complétude qui doit être au moins à 90 %, le rapport $I/\sigma(I) \geq 1$ et le $CC_{1/2} \geq 0,3$.

La suite CCP4 est ensuite utilisée pour la détermination du nombre de molécules à placer dans l'unité asymétrique avec le programme MATTHEWS_COEF (méthode de Matthews). Les phases sont ensuite calculées par remplacement moléculaire en utilisant comme modèle une des structures résolues de la Transportine 1 (code PDB : 1QBK, 2H4M, 2OT8, 2Z5J, 2Z5K, 4FDD, 4FQ3 et 4JLQ). Le remplacement moléculaire a été réalisé avec le programme PHASER et les solutions ont été recherchées dans l'ensemble des groupes d'espace appartenant au groupe ponctuel de symétrie identifié lors de l'indexation. Le choix parmi les solutions proposées est basé sur la valeur le TFZ-score qui doit être supérieur à 8 pour que la solution soit considérée acceptable.

L'affinement a été réalisé à l'aide du programme REFMAC5 ou LOW RESOLUTION REFINEMENT pour des structures à basse résolution ($> 3 \text{ \AA}$) de la suite CCP4. L'ajustement manuel des positions atomiques dans la carte de densité électronique a été fait avec le programme de visualisation COOT (189) qui permet de superposer les deux cartes de densité ($F_o - F_c$) qui met en évidence les différences entre le modèle et les données expérimentales et ($2F_o - F_c$) qui fait apparaître une densité électronique même dans les régions où le modèle est en accord avec les données expérimentales. Dans cette dernière carte, les régions où des atomes ne devraient pas être à leurs positions actuelles n'auront peu ou pas de densité électronique, tandis que les régions dont la densité apparente est vide correspondent aux positions où des atomes doivent être ajoutés au modèle pour s'accorder avec les données expérimentales. Des cycles itératifs de correction-affinement sont menés jusqu'à convergence des facteurs d'accord Rfacteur et Rlibre. La validation du modèle s'est fait grâce aux outils de COOT (qui permettent de vérifier le diagramme de Ramachandran, d'optimiser les rotamères des chaînes latérales, la planarité, les clashes ...), et les serveurs Molprobit et PDB-Validation Server (190).

2.2.3 Étude structurale en solution

A défaut de la résolution de la structure cristallographique du complexe Trn1/ADAR1, une analyse structurale en solution du complexe a été entreprise.

2.2.3.1 Diffusion des rayons X aux petits angles (SAXS)

2.2.3.1.1 Principe du SAXS

La diffusion des rayons X au petits angles ou SAXS (Small angle X-ray scattering), comme la cristallographie aux rayons X est une technique basée sur l'interaction entre les rayons X et les électrons de la matière. Par contre, contrairement à la cristallographie, le SAXS est une technique d'étude structurale de macromolécules en solution à basse résolution. Dans certains cas (pour un échantillon pur et monodisperse), les données obtenues par SAXS permettent d'estimer le rayon de giration (R_g) donnant une première estimation de la taille, la distance maximale de l'objet (D_{max}), ainsi qu'une estimation de la forme globale de la protéine grâce à la densité de diffusion de la solution $P(r)$. On obtient ainsi une enveloppe basse résolution de la protéine ou du complexe étudié qui fait état de l'allure globale de la protéine ou du complexe.

Les données de diffusion obtenues par SAXS correspondent à des mesures d'intensités diffusées I_q qui sont fonction du module du vecteur de diffusion q ($q = 4\pi \sin\theta / \lambda$, 2θ est l'angle de diffusion et λ est la longueur d'onde du faisceau incident).

Comme les molécules sont orientées aléatoirement dans la solution, l'intensité mesurée est la moyenne de toutes les intensités diffusées pour une molécule dans toutes les orientations.

Le signal de diffusion de la protéine I_q est la différence entre l'intensité diffusée par le solvant et l'intensité diffusée par l'échantillon (la protéine). La moyenne des intensités peut être associée à la différence de densité électronique $\Delta\rho_r$ ($\Delta\rho_r = \rho_r - \rho_{solvant}$) entre la protéine et le solvant.

Dans le cas d'une solution monodisperse, on peut déterminer le rayon de giration de la protéine en utilisant la loi de Guinier.

$$I_q = I_0 e^{\frac{-q^2 R_g^2}{3}} \Leftrightarrow \ln(I_q) = \ln(I_0) - \frac{q^2 R_g^2}{3} \quad (10)$$

où I_0 est l'extrapolation de l'intensité à l'origine et R_g le rayon de giration.

Le rayon de giration correspond à la distance quadratique moyenne en tous points de la particule au centre de la gravité pondérée par la masse diffusante en chaque point. Ainsi, les objets

allongés ont un rayon de giration plus grand que les objets sphériques de même masse. Dans la zone de Guinier (pour $q.R_g < 1,3$), I_0 est obtenue par l'intersection de la droite obtenue par régression linéaire avec l'axe des ordonnées, et la pente de la droite ($-1/3 R_g^2$) permet d'obtenir le rayon de giration. La valeur du rayon de giration est une première indication sur la forme de la protéine, l'état d'oligomérisation est donné par I_0 qui est proportionnelle à la masse de la particule.

La fonction de distribution des distances $P(r)$ (ou densité de diffusion de la solution) est la représentation des données de diffusion dans l'espace réel. Elle correspond à la distribution de toutes les distances interatomiques de la protéine. Cette fonction permet d'avoir une représentation graphique des paramètres globaux de la protéine. Grâce à la fonction de distribution des distances, il est possible d'avoir une deuxième indication concernant la forme de la protéine. Elle aura une forme asymétrique pour une protéine de forme allongée ou non structurée. En revanche, elle aura une forme de gaussienne pour une protéine globulaire.

Elle représente la transformée de Fourier de l'intensité diffusée I_q .

$$P(r) = \text{TF}(I_q) = \frac{r^2}{2\pi^2} \int_0^{D_{\max}} q^2 I_q \frac{\sin qr}{qr} dq \quad (11)$$

La fonction $P(r)$ a une valeur maximale pour laquelle, elle s'annule. Cette valeur correspond à la distance maximale de la protéine ($D_{\max}$). En effet, le $D_{\max}$ d'une protéine peut être déterminée grâce à la fonction $P(r)$ qui est égale à zéro pour une valeur de q égale au $D_{\max}$.

2.2.3.1.2 Protocole expérimental

En attendant de pouvoir résoudre la structure du complexe Transportine 1/ADAR1 à haute résolution par cristallographie, j'ai réalisé des expériences de SAXS pour analyser la reconnaissance d'ADAR1 par la Transportine 1. Ces expériences de SAXS permettraient de voir si ADAR1 est également reconnue au niveau de l'arche C-terminale où sont reconnus les PY-NLS. Il sera ainsi possible de pouvoir statuer sur la première hypothèse du modèle d'inactivation du NLS en présence d'ARN (voir Introduction paragraphe 3.2.3.1.1, page 55).

Les données de SAXS de Trn1 (Trn1FL) seule ou en complexe avec ADAR1 (ADAR1long) ont été enregistrées sur la ligne de lumière SWING au synchrotron SOLEIL. Peu avant l'expérience, les protéines ont, si besoin, été mélangées pour la formation du complexe pendant environ 1h sur glace, placées dans des vials pour l'injection robotisée. Pour les expériences de SAXS, les différentes protéines ont été préparées et purifiées comme précédemment décrit puis dialysées contre le tampon SAXS (20 mM Hepes pH 7,3, 55 mM KOAc, 10 mM NaCl,

1 mM TCEP). Un volume de 50 μ L d'échantillons de Trn1FL, ADAR1long ou Trn1FL + ADAR1long (1:10) a été injecté automatiquement sur une colonne d'exclusion stérique Bio-SEC3 (Agilent) montée sur un système HPLC (High Performance Liquid Chromatography) et préalablement équilibrée dans le tampon SAXS. Les échantillons ont été élués à 300 μ L/min à travers une cellule UV qui est un capillaire (diamètre 1,5 mm) placé sous vide et maintenu à 15°C pour l'exposition de l'échantillon aux rayons X. Les courbes de diffusion des rayons X ont été collectées tout au long de l'élution, une image a été collectée toutes les secondes. Pour chaque injection, 180 images ont été collectées par tampon et 480 images par échantillon.

Les courbes mesurées avant le passage de la protéine correspondent à la diffusion du tampon. Elles sont moyennées et soustraites des courbes d'acquisition des échantillons, afin de n'obtenir que le signal de diffusion de la protéine ou du complexe en solution. Les données ont été traitées à l'aide du logiciel FOXTROT, développé par l'équipe de la ligne SWING, puis ont été analysées en utilisant le logiciel PRIMUS (191). La fonction de Guinier (approximation linéaire de la portion de courbe aux plus petits angles) a été appliquée aux courbes pour la détermination des valeurs des paramètres I_0 et R_g caractéristiques de l'échantillon.

Après comparaison des différentes courbes, elles ont été moyennées. La courbe moyenne I_q ainsi obtenue correspondant à la molécule d'intérêt a ensuite été analysée. Puis la fonction de distribution des distances $P(r)$ a été déterminée pour l'obtention de la dimension maximale de l'objet D_{max} . Il a alors été possible d'obtenir la masse moléculaire de l'échantillon.

Les données traitées grâce au logiciel FOXTROT ont été analysées avec le programme SHANUM et ont été sélectionnées avec le programme DATCROP (192–194).

2.2.3.2 Résonance magnétique nucléaire (RMN)

2.2.3.2.1 Principe de la RMN

La RMN (Résonance Magnétique Nucléaire) est une technique spectroscopique basée sur l'interaction entre un champ magnétique et les spins des noyaux atomiques. C'est une autre technique classique d'études des protéines en solution qui permet d'étudier des interactions entre une protéine et son partenaire. Néanmoins, la RMN est une technique qui demande de grande quantité de protéine et le marquage des protéines est souvent nécessaire. Toutefois, elle permet d'obtenir des détails sur la structure de la protéine, mais aussi des informations sur la thermodynamique et la cinétique de l'interaction d'un complexe.

En utilisant la RMN, l'interaction entre une protéine et son partenaire peut être détectée soit en analysant les changements survenus sur les signaux de la protéine suite à la présence du partenaire (observation de la protéine) soit en analysant les changements survenus sur le partenaire suite à la présence de la protéine (observation du partenaire). L'étude de l'interaction entre une protéine et un partenaire par RMN est plus simple et plus rapide qu'en cristallographie puisqu'elle permet de détecter même des interactions très faibles. Et contrairement à la cristallographie, en RMN, il est possible non seulement de pouvoir caractériser le site de fixation du partenaire en solution mais aussi, d'analyser la dynamique de la protéine suite à la fixation du partenaire.

L'étude en solution de protéines avec un poids moléculaire élevé (> 100 kDa) par RMN est limitée par l'augmentation du nombre de signaux de manière proportionnelle à la longueur de la séquence de la protéine. Le nombre de recouvrement possibles est alors augmenté complexifiant l'analyse des spectres RMN. Pour des protéines de haut poids moléculaire, l'élargissement des signaux entraîne également une forte diminution de la sensibilité des expériences RMN jusqu'à une disparition complète. Une solution permettant de compenser ces effets est l'utilisation de marquage spécifique des groupements méthyles ($-CH_3$) dans une protéine deutérée. En effet, les protéines sont composées de six acides aminés naturels (méthionine, thréonine, valine, leucine, isoleucine, alanine) ayant un groupement méthyle qui peut être marqué pour des expériences de RMN. Bien que ces acides aminés ne représentent pas la majorité des acides aminés contenus dans une protéine, il a tout de même été démontré par de nombreuses études qu'ils vont permettre d'obtenir des informations sur des interactions protéine-protéine (90, 195, 196). L'inconvénient de cette solution est la complexité du marquage des groupements méthyles qui vient du fait qu'il faut pouvoir obtenir une deutération supérieure à 98 % de la protéine. Il faut en parallèle pouvoir ré-introduire spécifiquement une protonation et un marquage ^{13}C des groupements méthyles sélectionnés à plus de 95-98 %.

La production des protéines marquées spécifiquement sur les groupements méthyles est réalisée en milieu minimum M9 (2H_2O), et le marquage est obtenu par l'ajout dans le milieu de culture d'un précurseur métabolique de l'acide aminé ou directement de l'acide aminé marqué afin de saturer sa voie de biosynthèse.

2.2.3.2.2 Protocole expérimental

Toutes les expériences de RMN ont été réalisées par nos collaborateurs dans le but d'apporter des détails moléculaires concernant la reconnaissance d'ADAR1 par la Transportine 1. Ces

détails devraient être confrontés aux expériences basses résolutions réalisées en SAXS dans l'objectif d'apporter une meilleure caractérisation de la reconnaissance entre les deux partenaires. Les différentes constructions d'ADAR1 (ADAR1long et ADAR1court) ont été produites comme décrit précédemment. Une fois purifiée, la protéine ADAR1 a ensuite été dialysée une nuit contre le tampon RMN (50 mM Tris-HCl pH 6,7 ; 150 mM NaCl ; 2 mM TCEP). Puis, les échantillons ont été concentrés jusqu'à 10 mg/mL (environ 0.7 mM). La protéine Trn1 marquée Trn1FL^{15N-13C} (I, M méthyle) a été produite dans le tampon RMN par nos collaborateurs. Les expériences de titrations (¹H, ¹⁵N HSQC ; ¹H ¹³C HMQC) ont été réalisées dans le tampon RMN (supplémenté de 0,04 % de NaN₃) à 25°C et enregistrées à l'aide d'un spectromètre 600 MHz Avance Neo ou 900 MHz Avance III (Bruker).

II. La reconnaissance d'un set de non-PY-NLS par Trn1

La deuxième partie de ce chapitre porte sur la reconnaissance d'un set de non-PY-NLS par Trn1. Pour étendre la question concernant la reconnaissance d'ADAR1 par Trn1, nous avons aussi essayé de caractériser l'interaction entre Trn1 et un set de NLS reconnu par Trn1 mais qui ne porte pas de motif PY, les non-PY-NLS. Le set comprend le NLS des protéines CD44, REV (VIH), L2_HP18, et L2_HP16. Le but étant de former différents complexes entre Trn1 et ces NLS, résoudre leurs structures cristallographiques et essayer d'établir des règles de reconnaissance pour les non-PY-NLS.

1 Obtention du matériel

1.1 Préparation des peptides synthétiques

Comme vu dans le paragraphe 1.3 (page 71) de la partie précédente, nous avons aussi utilisé des peptides synthétiques pour la caractérisation de ces différents complexes. Leur séquence représente le NLS minimal identifié comme interagissant avec Trn1. Il s'agit du non-PY-NLS de la protéine CD44 (H- NSR RRC GQK KKL VIN SGN GAV EDR KPS – OH), L2 (HPV16) (H- MRH KRS AKR TKR – OH), et REV (H-MRQ ARR NRR RRW RER – OH). Ces peptides ont été synthétisés chimiquement par la société Eurogentec ® et le protocole de la préparation des peptides est le même que celui décrit dans le paragraphe 1.3 (page 71).

Peptide NLS	aa	PM	ϵ	pI
CD44_pep	27	2998.6	120	12.07
HPV16_pep	12	1554.9	–	12.96
REV_pep	15	2183.6	5690	13.02

Tableau 6 : Tableau regroupant les différents peptides non-PY-NLS synthétiques.

Les peptides non-PY-NLS portent le NLS minimal identifié interagissant avec Trn1 (aa = acides aminés, PM = poids moléculaire en g/mol, ϵ = coefficient d'extinction molaire en L.mol⁻¹.cm⁻¹ fourni par le fournisseur, pI = point isoélectrique).

1.2 Production des peptides recombinants

Après les expériences avec les peptides synthétiques, nous avons opté pour des peptides recombinants dans le but d'avoir un meilleur contrôle sur la production et les conditions de préparation des différents partenaires. Il s'agit d'un set comprenant le NLS de la protéine REV (VIH) et de la protéine L2_HP16 (HPV16).

Pour mettre en évidence et caractériser ces interactions, nous avons travaillé avec la construction Trn1 Δ H8 dont le protocole de production a été vu dans la partie précédente (paragraphe 1.1, page 63). Concernant les NLS, nous avons travaillé avec des peptides exprimés de manière recombinante sous formes de protéines de fusion avec un tag SUMO en N-terminal. Les protocoles d'expression et de purification sont identiques pour les deux peptides recombinants REV-NLS et HPV16-NLS.

Protéines	PM	pI	ϵ	Vecteur	Etiquette
REV-NLS	16524	9.51	8480	pRSF	His ₆
HPV16-NLS	15613	8.19	1490	pRSF	His ₆

Tableau 7 : Tableau regroupant les différents peptides non-PY-NLS recombinants.

1.2.1 Surexpression

La production des deux peptides recombinants commence par une étape de surexpression dans des cellules bactériennes *E. coli* suivie d'une lyse cellulaire par sonication puis de différentes étapes de chromatographie pour pouvoir obtenir du matériel suffisamment pur.

1.2.1.1 Souches bactériennes, vecteur plasmidique

Le vecteur pRSF (Novagen) est utilisé pour la surproduction des deux peptides recombinants REV-NLS (résidus V₁₆-R₄₆) et HPV16-NLS (M₁-A₁₃). Les séquences d'ADN codant pour ces deux peptides (REV-NLS et HPV16-NLS) ont été clonées entre les sites de restrictions BamHI et XhoI dans le vecteur pRSF-Duet-Smt3, à partir de l'amplification des séquences codantes présentes sur un plasmide synthétique. Ce plasmide synthétique regroupe les séquences ADN des différents non-PY-NLS choisis pour ce projet (Figure 27). Ces peptides recombinants portent une étiquette His₆, un site de coupure à la protéase thrombine suivi du tag SUMO qui peut être clivé grâce à la protéase spécifique Ulp1. Les cellules chimiquement compétentes BL21 (DE3) codon+ RIL d'*E. coli* sont transformées par choc thermique avec le plasmide pRSF-Duet-Smt3 contenant le gène d'intérêt. Les cellules transformées sont étalées sur gélose LB-agar contenant de la kanamycine à 50 µg/mL (résistance portée par le plasmide pRSF-Duet-Smt3) et du chloramphénicol à 25 µg/mL (résistance intrinsèque des cellules BL21 (DE3) codon+ RIL) permettant la sélection des cellules transformées par le plasmide d'intérêt.

1.2.1.2 Culture cellulaire

En ce qui concerne la surexpression des peptides de fusion SUMO (REV-NLS et HPV16-NLS), on prélève un clone isolé qui va être mis en pré-culture dans 100 mL milieu LB en présence de kanamycine (50 µg/mL) et de chloramphénicol (25 µg/mL). La pré-culture est placée pendant une nuit à 37°C. Elle sert à ensemercer un litre de milieu LB additionné des antibiotiques appropriés en partant d'une DO_{600nm} initiale de 0,05. Les cultures sont incubées à 37°C sous agitation jusqu'à atteindre une densité DO_{600nm} de 0,5. A cette densité, les cultures sont refroidies à 18°C et l'expression des peptides est induite par ajout d'IPTG (0,4 mM final) lorsque la mesure de la DO_{600nm} se situe entre 0,6 et 0,7. La culture est alors poursuivie à 18°C pendant 20 h. L'arrêt de la culture se fait par centrifugation à 5000 g pendant 20 min. Les culots sont repris dans 40 mL du milieu de culture puis à nouveau centrifugés à 5000 g pendant 15 min. Enfin, les culots sont congelés et conservés à -80°C.

1.2.2 Purification

La purification des deux peptides recombinants (REV-NLS et HPV16-NLS) est réalisée sans coupure de l'étiquette SUMO. Le tag SUMO est conservé pour favoriser la solubilité du peptide. La purification est effectuée en trois étapes de purification après la lyse bactérienne.

1.2.2.1 Lyse bactérienne

Pour les deux peptides recombinants, le culot bactérien est repris dans 20 mL de tampon de lyse (50 mM Tris-Hepes pH 8,2, 50 mM NaCl, 1 mM EDTA, 0,2 % (v/v) Triton X-100, 2 mM β Me) additionné d'une pastille de cocktail EDTA-free d'inhibiteurs de protéases. Ensuite, les cellules sont lysées à froid par sonication : 40 % d'amplitude, 30 cycles de 10 sec et pause de 10 sec entre chaque cycle. Le lysat est alors centrifugé à 30000 g pendant 30 min à 4°C. Le surnageant récupéré est ensuite filtré sur une membrane de 0,22 μ m.

1.2.2.2 Chromatographie d'affinité

La première étape de purification est une colonne d'affinité Nickel (HisTrap, HP, 5 mL, GE Healthcare). La colonne est préalablement équilibrée avec le tampon 1 (50 mM Tris-HCl pH 8,0, 1 M NaCl, 10 % (v/v) glycérol, 1 mM EDTA, 0,5 mM PMSF, 10 mM β Me). Le lysat est alors injecté sur la résine. Ensuite, après un lavage par 5 CV de tampon 1, on effectue un gradient allant de 0 à 100 % de tampon 2 sur 120 mL à 2 mL/min (tampon 1 + 500 mM Imidazole) qui permet l'élution du peptide.

1.2.2.3 Chromatographie d'interaction hydrophobe

La deuxième étape de purification consiste en une chromatographie d'interaction hydrophobe, la colonne (Hiprep Phenyl, HP 16/10, 20 mL, GE Healthcare) est préalablement équilibrée dans du tampon 3 (50 mM Tris-HCl pH 8,0, 100 mM NaCl, 1 M $(\text{NH}_4)_2\text{SO}_4$, 10 % (v/v) glycérol, 1 mM EDTA, 0,5 mM PMSF, 10 mM β Me) avant l'injection de la fraction contenant la protéine récupérée précédemment. On réalise ensuite un gradient de 6 CV entre 0 -100 % de tampon 4 (50 mM Tris-HCl pH 8,0, 100 mM NaCl, 10 % (v/v) glycérol, 1 mM EDTA, 0,5 mM PMSF, 10 mM β Me) à un débit de 1 mL/min qui permet l'élution du peptide.

1.2.2.4 Chromatographie d'exclusion stérique

La dernière étape consiste à améliorer la pureté de l'échantillon. Pour cela, on fait passer l'échantillon récupéré à l'étape précédente sur un tamis moléculaire (HiLoad 26/60, Superdex 75, Prep Grade, 320 mL GE Healthcare). La colonne est préalablement équilibrée avec du tampon de transport, et un maximum de 10 mL d'échantillon est injecté. Les échantillons sont concentrés par centrifugation sur une unité de filtration Amicon de 10 kDa. Enfin, ils sont congelés dans la carboglace et conservés à -20°C.

1.2.2.5 Analyse par SDS-PAGE

Les étapes de purification sont suivies par électrophorèse. Entre chaque colonne de purification, les échantillons (10 μ L) sont dénaturés à 95°C pendant 5 min en présence de 100 mM β -Me et de 0,25 % SDS. Les protéines sont séparées en fonction de leur masse sur un gel SDS-PAGE à 16 % (acrylamide) pour les deux peptides (REV-NLS et HPV16-NLS).

1.3 Construction et production des protéines chimériques

Comme présenté dans la partie précédente (paragraphe 1.4, page 71), nous avons aussi construit des protéines chimériques pour favoriser la formation des différents complexes entre Trn1 et le set de non-PY-NLS. Au set présenté plus haut, constitué du NLS de la protéine REV (VIH) et L2_HP16 (HPV16), nous avons ajouté un autre NLS, celui de la protéine L2_HP18 (HPV18).

Ainsi, nous avons construits trois chimères de Trn1 avec les non-PY-NLS étudiés : la chimère de Trn1 avec le non-PY-NLS de la protéine REV (Trn1 Δ H8-REV_pep), la chimère de Trn1 et le non-PY-NLS de la protéine L2_HP16 (Trn1 Δ H8-HPV16_pep) et la chimère de Trn1 avec le non-PY-NLS de la protéine L2_HP18 (Trn1 Δ H8-HPV18_pep).

Protéines	PM	pI	ϵ	Vecteur	Etiquette
Trn1 Δ H8-REV_pep	102858	5.23	105770	pGEX	GST
Trn1 Δ H8-HPV16_pep	101946	5.16	98780	pGEX	GST
Trn1 Δ H8-HPV18_pep	101846	5.10	98780	pGEX	GST

Tableau 8 : Tableau regroupant les différentes constructions chimériques (Trn1 et les différents non-PY-NLS).

1.3.1 Souches bactériennes, vecteur plasmidique

Comme pour la chimère Trn1 Δ H8-ADAR1_pep, le système pGEX a été utilisé pour la surexpression des différentes chimères ici. La construction de ces chimères est basée sur le même principe que précédemment, et le protocole de clonage est également le même. Pour chaque chimère, la séquence ADN du peptide précédée de celle de l'espaceur a été clonée par PCR entre les sites de restrictions EcoRI et XhoI dans le vecteur d'expression pGEX-6p-1 (Trn1 Δ H8).

Les constructions chimériques comportent une étiquette GST en N-terminal, un site de coupure à la PreScission HRV 3C et leur expression est sous le contrôle du promoteur tac, induit par l'IPTG. Les chimères sont exprimées dans les cellules chimiquement compétentes BL21 (DE3) codon + RIL d'*E. coli* transformées par choc thermique comme précédemment indiqué.

1.3.2 Surproduction et purification

Les protocoles d'expression et de purification de ces chimères sont identiques à ceux des chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep, présentés dans la partie précédente (paragraphe 1.4, page 71) et ne seront donc pas présentés plus en détail ici.

2 Caractérisation des complexes Trn1/non-PY-NLS

Après l'obtention du matériel compatible avec des expériences d'études structurales, une caractérisation d'abord physico-chimique des différents complexes Trn1/non-PY-NLS est réalisée par trois méthodes différentes.

2.1 Caractérisation physico-chimique des protéines

2.1.1 GST Pull-Down

Après expression de la protéine Trn1ΔH8, purifiée en une étape, elle a ensuite été fixée sur les billes de glutathion agarose (Glutathion Sepharose 4B, GE HealthCare). Elle a ainsi été utilisée pour des expériences de GST pull-down. Les peptides synthétiques REV_pep, HPV16-pep et CD44_pep ont été préparés comme décrit précédemment (paragraphe 1.1, page 104).

La protéine GST-Trn1ΔH8 est incubée avec chacun des différents peptides comme suit : 50 µL de mélange contenant la protéine de fusion GST et le peptide sont préparés à un ratio 1 : 8 (600 µg de Trn1 soit 0,027 mM, 70 µg de HPV16-pep, 100 µg de REV_pep et 132 µg de CD44) et incubés à 4°C pendant 1h sous agitation rotative. Ce mélange est ensuite ajouté aux billes de glutathion sépharose qui vont être incubées pendant encore une heure sous agitation rotative à température ambiante. Après trois lavages avec le tampon A, les protéines sont éluées dans du tampon de charge puis séparées sur gel SDS-PAGE.

2.1.2 Differential Scanning Fluorimetry (DSF)

J'ai effectué des expériences de DSF pour vérifier la stabilité des protéines chimériques construites (Trn1ΔH8-REV_pep, Trn1ΔH8-HPV16_pep et Trn1ΔH8-HPV18_pep) et aussi pour vérifier l'interaction entre Trn1 et les différents peptides non-PY-NLS étudiés.

Les expériences de DSF ont été réalisées comme décrit dans la partie précédente (paragraphe 2.1.2, page 109). Les protéines testées sans partenaire ont été concentrées à 2 μM (environ 1 mg/mL). Et, pour vérifier l'interaction entre Trn1 et les différents peptides Trn1 a été concentrée à 2 μM et une gamme de concentrations des peptides a été testée : de 1 μM à 40 μM . Les expériences ont été réalisées dans le tampon de transport. Le Sypro Orange a été ajouté à une concentration finale de 1X. La température est augmentée par pas de 1°C / min entre 25 et 90°C et la fluorescence est mesurée toutes les 0,2°C.

2.1.3 Interférométrie de bio-couche (BLI)

2.1.3.1 Principe du BLI

L'interférométrie de bio-couches ou Bio-Layer Interferometry (BLI) est une technique basée sur le principe de l'interférométrie en lumière blanche. C'est une technique optique qui mesure les variations d'épaisseur optique à la surface d'un biocapteur grâce aux interférences créées lors de la réflexion d'un faisceau lumineux à la surface du biocapteur et à la surface de la couche biomoléculaire. Elle permet ainsi la caractérisation des interactions de macromolécules.

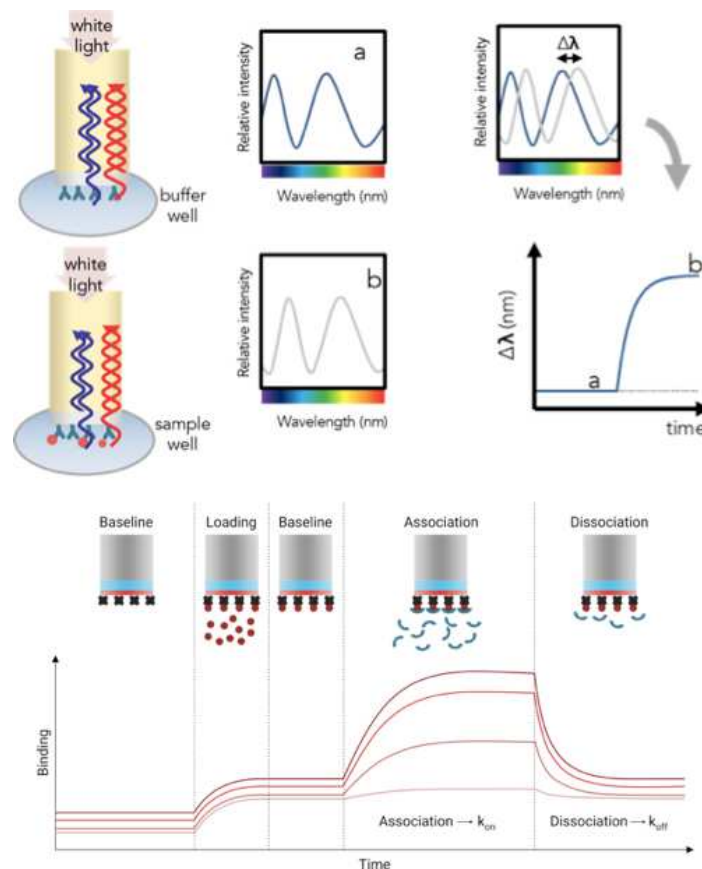


Figure 33 : Principe de l'interférométrie de bio-couches (BLI).

Le système interférométrique se compose d'un biocapteur et d'une fibre optique, il est recouvert à son extrémité d'une matrice biocompatible qui va permettre l'immobilisation de molécules d'intérêt (Figure 33). Cette construction optique permet la surveillance en temps réel des interactions biomoléculaires qui se produisent sur la surface du biocapteur. Plus les molécules se fixent à la surface (augmentation de l'épaisseur), plus le profil interférométrique est déplacé vers la droite (déplacement des longueurs d'onde). A l'inverse, si les molécules se dissocient de la surface le profil interférométrique est décalé vers la gauche par rapport à sa position initiale. Lorsque ce décalage est mesuré sur une durée donnée, une courbe d'association/dissociation est alors obtenue (Figure 33).

2.1.3.2 Protocole expérimental

Les expériences de BLI ont été réalisées au cours d'un atelier de biologie structurale intégrative. Le but de ces expériences était de caractériser les interactions entre la Transportine 1 et le set de non-PY-NLS étudié. La protéine Trn1 Δ H8 a été produite selon le protocole décrit dans la partie précédente (paragraphe 1.1 , page 63). Les peptides ont également été préparés comme indiqué plus haut (paragraphe 1.1 , page 104). Pour la réalisation des expériences BLI, la protéine ainsi que les peptides ont été dialysés contre le tampon BLI (tampon PBS pH 7,4) pour permettre la réaction de biotinylation. Pour les différentes expériences effectuées, le système streptavidine-biotine a été utilisé. Ces deux molécules présentent une forte affinité l'une pour l'autre (l'une des plus fortes interactions non covalentes connues dans le monde du vivant). La streptavidine est fixée sur la matrice du biocapteur, la Transportine 1 (Trn1 Δ H8 concentrée à 20 ou 10 μ M) biotinylée crée alors une liaison extrêmement stable à la surface du biocapteur. Des mesures d'interaction sont alors effectuées à la suite de l'ajout du peptide. Pour chaque peptide différentes concentrations ont été testées (de 40 – 5000 nM).

2.2 Cristallisation des complexes Trn1/non-PY-NLS

Comme pour la cristallisation du complexe Trn1/ADAR1, la cristallisation des complexes de Trn1 (Trn1 Δ H8) avec le set de non-PY-NLS (REV, L2_HP16 et L2_HP18) étudié a été réalisée au sein du laboratoire.

Dans un premier temps, des tests de co-cristallisation ont été réalisés avec les peptides synthétiques par diffusion de vapeur en gouttes assises. Le peptide synthétique est mélangé à la protéine à différents ratio (de 1:1 à 1:7). Les peptides ont aussi été utilisés pour des expériences de trempage permettant de favoriser la cristallisation des différents complexes. Les

cristaux de la protéine Trn1ΔH8 seule ont été trempés dans des solutions contenant les différents peptides à différentes concentrations (0.15-1.05 mM). Ensuite, les tests de cristallisation des différentes chimères ont été effectués. Puis, des tests de co-cristallisation ont été réalisés avec les peptides produits de manière recombinante à des ratios allant de 1:1 à 1:5. Plusieurs étapes d'optimisation ont été menées, des tests de microseeding ont été réalisés afin d'améliorer la qualité de certains cristaux.

III. La reconnaissance d'un ARN double brin par ADAR1

1 Obtention du matériel

Cette dernière partie porte sur la reconnaissance de l'ARN double brin (ARNdb) par le dsRBD3 de la protéine ADAR1. Pour mettre en évidence et caractériser l'interaction entre le dsRBD3 et l'ARNdb, nous avons entrepris l'expression et la purification d'une construction qui comprend uniquement le dsRBD3 (sans les extensions N- et C-terminales qui constituent le NLS d'ADAR1), cette construction est appelée ADAR1ΔNLS

1.1 Production de la protéine ADAR1ΔNLS

1.1.1 Surexpression

Comme pour les différentes constructions précédentes d'ADAR1, la production de cette construction commence également par une étape de surexpression dans des cellules bactériennes *E. coli* suivie d'une lyse cellulaire par sonication puis de différentes étapes de chromatographie ainsi qu'une étape de coupure de l'étiquette.

1.1.1.1 Souches bactériennes, vecteur plasmidique

Le système pET (Novagen) a été utilisé pour la surexpression de la construction ADAR1ΔNLS. La séquence d'ADN codant pour la protéine ADAR1 humaine (ADAR1ΔNLS - résidus I₇₁₆-E₇₉₇) a été auparavant clonée dans le vecteur pET28 permettant de travailler avec une étiquette His₆ en N-terminal ainsi qu'un site de coupure à la protéase TEV. Son expression est sous le contrôle du promoteur T7 qui est induit par l'IPTG.

La protéine ADAR1 Δ NLS est exprimée dans les cellules chimiquement compétentes BL21 (DE3) codon+ RIL d'*E. coli* transformées par choc thermique comme précédemment indiqué, en utilisant les antibiotiques kanamycine et chloramphénicol. Les cellules transformées sont donc étalées sur gélose LB-agar contenant de la kanamycine (50 μ g/mL) et du chloramphénicol (25 μ g/mL). Enfin, les cellules sont incubées à 37°C pour la nuit.

1.1.1.2 Culture cellulaire

L'expression d'ADAR1 Δ NLS se fait comme suit : une colonie isolée obtenue de l'étape précédente est prélevée et mise en pré-culture dans 100 mL milieu LB en présence des antibiotiques appropriés. La culture est incubée pendant une nuit à 37°C sous agitation. Cette pré-culture sert ensuite à ensemer 1 litre de milieu LB en partant d'une DO_{600nm} de 0,05. La culture est refroidie à 30°C quand elle atteint une DO_{600nm} de 0,5. L'induction se fait par l'ajout d'IPTG (1 mM final) lorsque la DO_{600nm} mesurée se situe entre 0,6 et 0,7. L'expression se poursuit sur la nuit à 30°C et l'arrêt de la culture se fait par centrifugation à 5000 g pendant 20 min. Les culots sont repris dans 40 mL du milieu de culture puis à nouveau centrifugés à 5000 g pendant 15 min. Enfin, les culots sont congelés et conservés à -80°C.

1.1.2 Purification

1.1.2.1 Lyse bactérienne

Le culot bactérien est repris dans 20 mL de tampon de lyse (50 mM Tris-HCl pH 8,0, 1 M NaCl, 1 mM DTT, 1 mM EDTA) supplémenté par une pastille de cocktail EDTA-free d'inhibiteurs de protéases (ROCHE). Les cellules bactériennes sont ensuite lysées à froid par sonication : 40 % d'amplitude, 30 cycles de 10 sec et pause de 10 sec entre chaque cycle. Le lysat est alors centrifugé à 30000 g pendant 45 min à 4°C. Le surnageant est récupéré, puis filtré sur une membrane de 0,22 μ m avant purification.

1.1.2.2 Chromatographie d'affinité

Les échantillons d'ADAR1 Δ NLS sont purifiés par une première étape de chromatographie d'affinité en utilisant une colonne d'affinité Nickel (HisTrap, HP, 5 mL GE Healthcare). La colonne est préalablement équilibrée avec le tampon G (50 mM Tris-HCl pH 8,0, 1 M NaCl). Le lysat est alors injecté sur la résine et grâce à son étiquette His₆, la protéine se fixe sur la colonne. Après un lavage de 5 CV avec un step de 4 % de tampon E (50 mM Tris-HCl pH 8,0,

1 M NaCl, 1 mM DTT, 500 mM Imidazole), on effectue un gradient allant de 4 à 100 % de tampon E permettant l'élution de la protéine. Les fractions contenant la protéine d'intérêt sont récupérées, et les concentrations en NaCl et Imidazole sont diminuées (on baisse les concentrations de 1 M à 300 mM NaCl, et de 500 à 35 mM Imidazole) pour permettre une coupure optimale de l'étiquette et pour aussi réaliser l'étape suivante. On procède ensuite à la coupure de l'étiquette His₆ comme pour la construction ADAR1long-B. La coupure de la double étiquette se fait avec la protéase TEV (1 mg de protéases pour 100 mg de protéines), la coupure est réalisée à 20°C sur la nuit.

1.1.2.3 Chromatographie d'affinité inverse

La deuxième étape de purification d'ADAR1ΔNLS est également une chromatographie d'affinité mais cette fois-ci, inverse. En effet, une fois l'étiquette clivée, l'échantillon est à nouveau injecté sur la colonne de Nickel préalablement équilibrée dans le tampon H (50 mM Tris-HCl pH 8,0, 300 mM NaCl, 1 mM DTT). Comme elle n'a plus d'étiquette, la protéine ne se fixe pas sur la colonne, on peut ainsi séparer la protéine de l'étiquette qui reste fixée sur la colonne. On récupère donc la fraction non fixée sur la résine.

1.1.2.4 Chromatographie d'exclusion stérique

Pour une pureté optimale de la protéine, la dernière étape de purification consiste à faire passer l'échantillon récupéré après la deuxième étape de purification sur un tamis moléculaire (HiLoad 26/60, Superdex 75, Prep Grade, 320 mL, GE Healthcare). La colonne est préalablement équilibrée avec du tampon I (50 mM Tris-HCl pH 8,0, 100 mM KCl, 1mM TCEP), et un maximum de 10 mL d'échantillon est injecté. Les fractions contenant la protéine pure sont rassemblées puis l'échantillon est concentré par centrifugation sur une unité de filtration Amicon de 3 kDa. Enfin, l'échantillon est congelé dans la carboglace et conservé à -20°C.

1.1.2.5 Analyse par SDS-PAGE

Les différentes étapes de purification sont suivies par électrophorèse. Entre chaque colonne, les échantillons (5 µL) sont dénaturés à 95°C pendant 5 min en présence de 100 mM β-Me et de 0,25 % SDS. Les protéines sont séparées en fonction de leur masse sur un gel SDS-PAGE à 18 % (acrylamide).

1.2 Préparation de l'ARN double brin

Pour la résolution de la structure du complexe entre le dsRBD3 d'ADAR1 et l'ARN double brin, nous avons choisi un ARN dont la séquence est « palindromique » en se basant sur l'étude de *Bou-Nader et al* (197). Il s'agit d'une petite séquence (11-13 nucléotides) « palindromique » qui, en solution, est capable de s'apparier avec elle-même et de s'auto-assembler en double hélice grâce à des extrémités cohésives stables (GC) pour former une longue hélice d'ARN double brin (Figure 34). L'ARN double brin ainsi formé est susceptible de se lier au dsRBD3 d'ADAR1 (ADAR1 Δ NLS). Par ailleurs, la séquence de l'ARN présente une zone centrale AU (Tableau 9) qui permet de la discriminer de la zone GC favorisant ainsi un assemblage unique.

Les ARN utilisés sont chimiquement synthétisés à l'échelle de 1 μ mol par la société Dharmacon®. Les oligonucléotides sont préparés selon le protocole de déprotection obtenu du fournisseur. Il s'agit d'une déprotection de l'ARN en le chauffant 30 min à 60°C. Après lyophilisation, ils sont resolubilisés dans le tampon I à une concentration finale d'environ 2,5 mM afin de faciliter la réalisation des mélanges avec la protéine ADAR1 Δ NLS. Les échantillons sont congelés et gardés à -20°C.

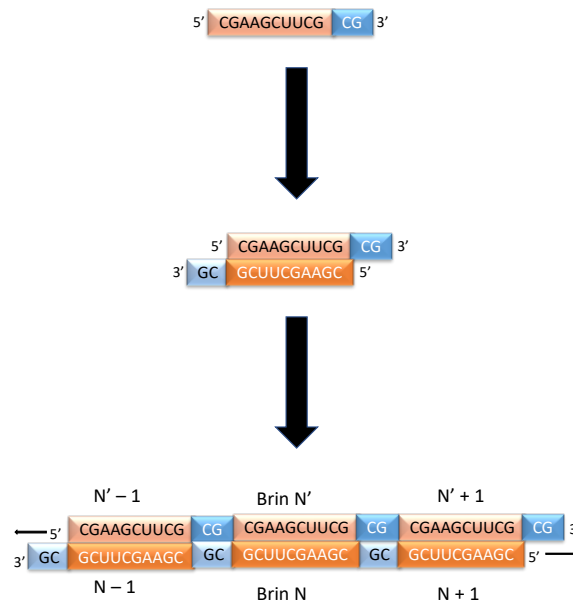


Figure 34 : Schéma représentant la formation d'un ARNdb à partir de la séquence "palindromique".

La séquence « palindromique » est capable de s'auto-assembler de manière unique pour former d'abord une séquence d'ARN double brin. Cette séquence d'ARN double brin peut ensuite s'agrandir, grâce aux extrémités cohésives GC, pour former un ARN double brin susceptible d'interagir avec un dsRBD.

ARN	nt	PM	sequence	ϵ
ARN-I	12	3810	CGAAGCUUCGCG	112,1
ARN-II	11	3465	CGAACUUCGCG	102,2
ARN-III	13	4115	CGAAGCCUUCGCG	119,1

Tableau 9 : Tableau regroupant les différents ARN synthétiques étudiés.

Trois séquences « palindromiques » sont étudiées pour cette partie (nt = nombre de nucléotides, PM = poids moléculaire en g/mol, ϵ = coefficient d'extinction molaire en $\text{L.mol}^{-1}.\text{cm}^{-1}$ (fournisseur)).

2 Caractérisation du complexe ADAR1/ARNdb

Après obtention d'échantillon protéique suffisamment homogène, pur et concentré, des tests de cristallisation ont été réalisés.

Dans un premier temps, la protéine seule ADAR1 Δ NLS a été cristallisée par criblage large robotisé en testant différents kits (Index HT, JCSG plus, PACT premier et crystal screen), et différentes concentrations de protéines (0,5 – 1,6 mM). Ces expériences de cristallisation ont été réalisées sur le même principe que celles avec la protéine seule Trn11 Δ H8.

Après l'obtention de cristaux reproductibles de la protéine ADAR1 Δ NLS seule, des tests de cristallisation du complexe ADAR1 Δ NLS/ARNdb ont été réalisés. Pour cela trois ARN (ARN- I, ARN-II et ARN-III) de séquences différentes ainsi que différents kits commerciaux de cristallogénèse ont été testés. Les tests de co-cristallisation ont été faits en mélangeant la protéine et l'ARN à différents ratios (1:1 et 1:2) sur glace pendant 30 min avant la réalisation des boîtes de cristallisation. Plusieurs étapes d'optimisation ont été effectuées afin d'obtenir des cristaux de taille convenable et aussi, afin d'améliorer la qualité de certains cristaux.

La résolution de la structure de la protéine seule a été réalisée par remplacement moléculaire en utilisant comme modèle la structure du dsRBD3 d'ADAR1 résolue par RMN. La résolution de la structure de la protéine en complexe avec l'ARN a également été faite par remplacement moléculaire mais en utilisant comme modèle la structure cristallographique de la protéine seule ADAR1 Δ NLS.

Résultats

Résultats

I. Reconnaissance d'ADAR1 par Trn1

Pour caractériser l'interaction entre la Transportine 1 et ADAR1, nous avons entrepris de résoudre la structure du complexe Trn1/ADAR1 par cristallographie aux rayons X et avons analysé ce complexe par différentes approches.

1 Caractérisation du complexe Trn1 Δ H8 / ADAR1court

Afin d'essayer de faciliter la cristallisation du complexe Trn1/ADAR1, nous avons travaillé avec une construction de la protéine Trn1 dans laquelle la longue boucle H8 est délétée (Trn1 Δ H8). Cette construction a largement été utilisée dans les études structurales de Trn1. En effet, sur les 18 structures cristallographiques de Trn1 disponibles sur la PDB, seules 5 ont été résolues avec la protéine Trn1 entière (Trn1FL). De plus, il a été montré par différentes études notamment celle d'Imasaki *et al* (77) que la boucle H8 n'est pas impliquée dans la reconnaissance des cargos de Trn1 mais seulement dans leur libération. Pour ADAR1, la construction ADAR1court (ADAR1_708-801) a d'abord été utilisée pour les tests de cristallisation du complexe Trn1/ADAR1. Cette construction est uniquement composée du dsRBD3 qui porte le NLS minimal d'ADAR1 reconnu par Trn1 (102, 174).

1.1 Purification de Trn1 Δ H8

Le protocole de purification de Trn1 Δ H8 consiste en trois étapes de purification : une colonne d'affinité glutathion sepharose, puis une colonne échangeuse d'anions et enfin une colonne d'exclusion stérique. Historiquement, les tampons utilisés dans ce protocole contenaient du DTT 10 mM et du glycérol 5 % (v/v). Avec ce protocole de purification, nous n'avons pas pu obtenir suffisamment de matériel pur pour entreprendre les tests de cristallisation. J'ai donc optimisé le protocole de purification, notamment en modifiant les tampons. D'une part la concentration de glycérol a été augmentée (20 % (v/v)), ce qui permet de rendre la protéine plus soluble. D'autre part, le DTT a été remplacé par le TCEP (1 mM), ce qui renforce la stabilité de la protéine. Les résultats de la purification présentés ci-dessous ont été obtenus grâce à cette optimisation.

Après la lyse bactérienne par sonication, centrifugation et filtration, le surnageant est injecté sur la colonne glutathion d'affinité sepharose. Cette première chromatographie d'affinité permet d'éliminer une grande partie des contaminants cellulaires. La protéine est éluée par compétition en injectant le tampon d'élution sur la colonne, le tampon d'élution contenant 20 mM de glutathion. Lors de l'élution, un pic unique d'absorbance à 280 nm représentant Trn1ΔH8 est observé (Figure 35A). Les fractions contenant Trn1ΔH8 ont ensuite été regroupées et l'étiquette GST a été coupée grâce à la protéase HRV 3C. Après coupure de l'étiquette, une chromatographie échangeuse d'anions a été réalisée. Cette deuxième étape de purification permet de séparer l'étiquette GST de Trn1ΔH8 grâce à un gradient de NaCl. Trn1ΔH8 est éluée à 75 mM de NaCl (Figure 35B). Puis, les fractions contenant Trn1ΔH8 ont été regroupées et injectées sur une colonne d'exclusion stérique équilibrée préalablement dans le tampon A (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 20 % (v/v) glycérol, 1 mM TCEP). La protéine est éluée après passage de 185 mL de tampon A, on observe un pic unique d'absorbance à 280 nm de Trn1ΔH8 (Figure 35C). L'analyse de la purification par SDS-PAGE indique qu'après la dernière étape de purification, Trn1ΔH8 est suffisamment pure pour les tests de cristallisation (Figure 35D).

En suivant ce protocole, environ 3 mg de protéine pure par litre de culture sont obtenus. Pour optimiser le rendement et limiter les pertes dues à la précipitation de la protéine, les conditions de purification et de concentration ont été améliorées. La fraction non retenue par la colonne glutathion a été injectée une deuxième fois sur la colonne d'affinité pour permettre une plus grande fixation de Trn1ΔH8 (Figure 35A). De plus, entre chaque étape de purification, une fois les fractions rassemblées, la concentration en TCEP dans l'échantillon est ramenée à 2 mM de TCEP afin de limiter l'oxydation de la protéine qui entraîne sa précipitation. Avec toutes ces mesures, j'ai obtenu jusqu'à 5 mg de protéine Trn1ΔH8 pure par litre de culture.

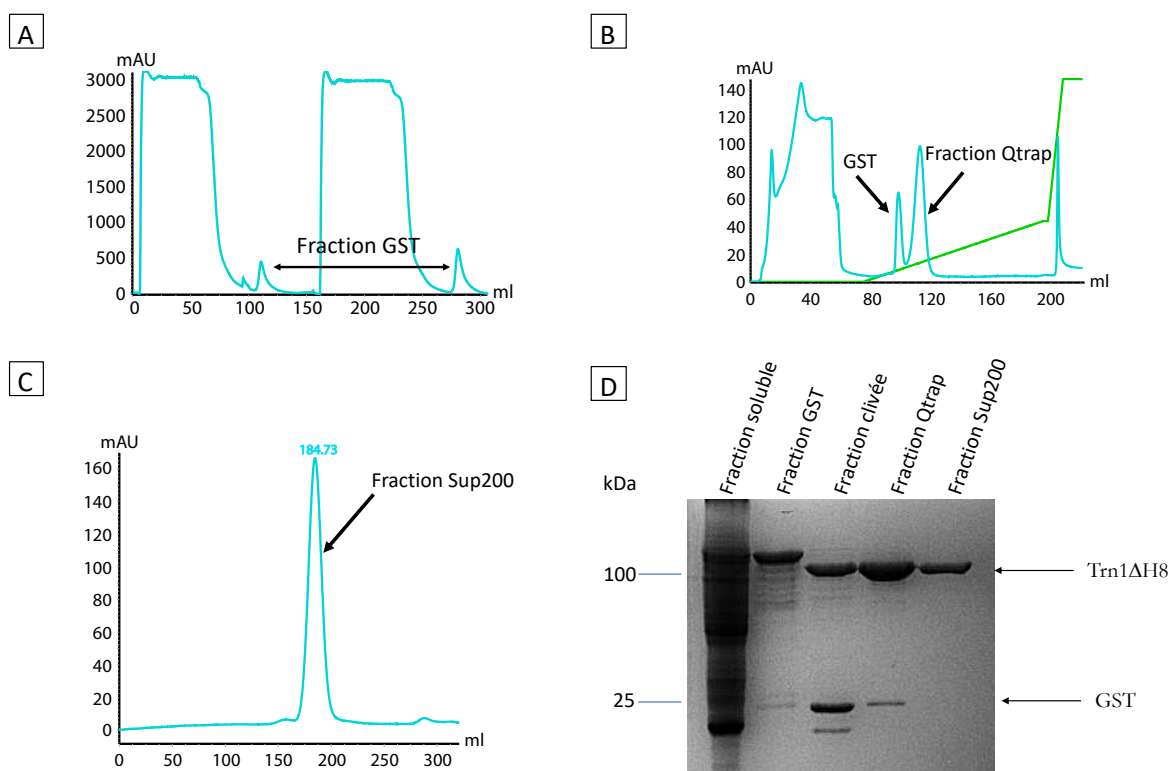


Figure 35 : Purification de Trn1ΔH8.

La purification de Trn1ΔH8 se fait en trois étapes chromatographiques avec une coupure de l'étiquette GST (les chromatogrammes représentent l'absorbance à 280 nm (mAU) en fonction du volume (mL) (chromatogramme d'absorbance (en cyan) et profil du gradient d'élution (en vert)). (A) Chromatographie d'affinité glutathion sepharose (GST). (B) Chromatographie échangeuses d'anions Q sepharose (Qtrap). (C) Chromatographie d'exclusion stérique (Sup200). (D) Analyse des fractions par SDS-PAGE 12 % (fraction soluble = surnageant récupéré après lyse, centrifugation et filtration, fraction GST = fraction regroupant les fractions collectées après la glutathion sepharose, fraction clivée = fraction dont l'étiquette GST est coupée, fraction Qtrap = fraction regroupant les fractions collectées après la Q sepharose, fraction Sup200 = fraction regroupant les fractions collectées après la Superdex S200).

1.2 Purification d'ADAR1court

La protéine ADAR1court (ADAR1_708-801) est purifiée en trois étapes : une chromatographie d'affinité, une chromatographie échangeuses de cations et une chromatographie d'exclusion stérique.

La fraction soluble a été injectée sur une colonne d'affinité de nickel sepharose. Une partie des contaminants protéiques a été éliminée grâce à l'ajout de 10 mM d'Imidazole dans le tampon de lyse, ainsi les protéines ayant une faible affinité pour la résine ne se fixeront pas sur la résine de nickel. ADAR1court est éluée à 140 mM d'Imidazole. Les fractions contenant la protéine ont ensuite été regroupées puis l'étiquette GST-His₆ a été clivée par la protéase HRV 3C. Après coupure, une chromatographie échangeuse de cations a été réalisée. Cette deuxième colonne

permet de séparer l'étiquette coupée de la protéine ADAR1court grâce à un gradient de NaCl. ADAR1court est éluée à 175 mM de NaCl. Les fractions contenant la protéine ont été rassemblées, et l'ensemble a été injecté sur une colonne d'exclusion stérique équilibrée préalablement dans le tampon A. Le tamis moléculaire permet de séparer ADAR1court des derniers contaminants encore présents dans l'échantillon, mais aussi de la mettre dans le même tampon que son partenaire de cristallisation, Trn1ΔH8. La protéine ADAR1court est éluée après passage de 176 mL de tampon A.

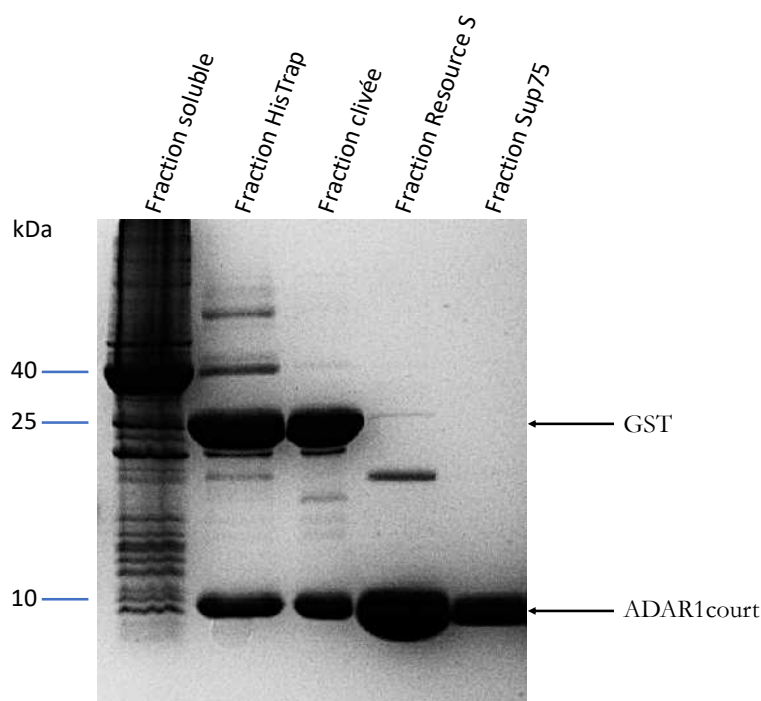


Figure 36 : Analyse SDS-PAGE des différentes étapes de purification d'ADAR1court.

L'analyse de la purification par SDS-PAGE (14 %) montre une pureté satisfaisante (fraction soluble = surnageant récupéré après lyse, HisTrap = fraction regroupant les fractions collectées après la première étape, fraction clivée = fraction dont l'étiquette est coupée, fraction Resource S = fraction regroupant les fractions collectées après la deuxième étape, fraction Sup75 = fraction regroupant les fractions collectées après la dernière étape).

L'analyse d'ADAR1court par SDS-PAGE au cours de la purification révèle que l'étiquette GST-His₆ commence peu à peu à se dissocier de la protéine ADAR1court avant même l'étape de coupure par la protéase HRV 3C (Figure 36). Ceci indique une possible auto-coupure de la protéine puisqu'on peut observer sur le gel (fraction HisTrap), deux bandes : une bande à 25 kDa qui représente l'étiquette GST-His₆ et une deuxième bande autour de 10 kDa qui représente la protéine ADAR1court. Néanmoins, en fin de tamis moléculaire, on arrive tout de même à obtenir une protéine suffisamment pure pour initier les tests de co-cristallisation.

La purification d'ADAR1court est satisfaisante puisqu'en suivant ce protocole, il est possible d'obtenir jusqu'à 10 mg de protéine pure par litre de culture.

1.3 Caractérisation du complexe Trn1ΔH8/ADAR1court

Une fois la protéine Trn1ΔH8 purifiée en quantité suffisante, j'ai mené des tests de cristallisation, dans un premier temps, de Trn1ΔH8 seule. Après l'obtention de cristaux de Trn1ΔH8 seule, j'ai entrepris la co-cristallisation du complexe Trn1ΔH8/ADAR1court.

1.3.1 Cristallisation de Trn1ΔH8

Les premiers tests de cristallisation ont été effectués avec Trn1ΔH8 seule concentrée entre 10 et 20 mg/mL. Un premier criblage large a été réalisé à deux températures différentes : 4 et 20 °C. Des cristaux irréguliers ont été obtenus dans plusieurs conditions de cristallisation notamment celles-ci (0,14 M K₂HPO₄, 1,26 M NaH₂PO₄), (0,91 M K₂HPO₄, 0,49 M NaH₂PO₄), (35 % (v/v) TACSIMATE) et (0,10 M Hepes pH 7,00, 0,50 % (v/v) Jeff ED-2001, 0,10 M Na₂ Malonate) à 4 ou 20 °C (Figure 37). Dans les autres conditions, il s'agissait de cristaux plus irréguliers avec un plus faible pouvoir diffractant (au-delà de 8 Å). Ainsi, nous avons choisi d'optimiser uniquement les conditions les plus prometteuses, c'est-à-dire les conditions pour lesquelles des cristaux irréguliers étaient apparus avec un pouvoir de diffraction encourageant (~6-7 Å). De plus, les expériences d'optimisation ont été effectuées uniquement à 20°C car aucun des cristaux obtenus à 4°C ne présentaient une diffraction satisfaisante (Figure 37 A-B). J'ai particulièrement optimisé les cristaux obtenus dans la condition (35 % (v/v) TACSIMATE) (Figure 37C) qui avaient un pouvoir de diffraction autour de 6 Å. Cette condition est composée de TACSIMATE, un mélange d'acides organiques (100 % de TACSIMATE contient : 1,36 M acide malonique, 0,25 M citrate de tri-ammonium, 0,12 M acide succinique, 0,3 M acide DL-malique, 0,4 M acétate de sodium, 0,5 M formate de sodium et 0,16 M tartrate de di-ammonium). Les cristaux obtenus dans cette condition présentaient des formes avec plus de régularité.

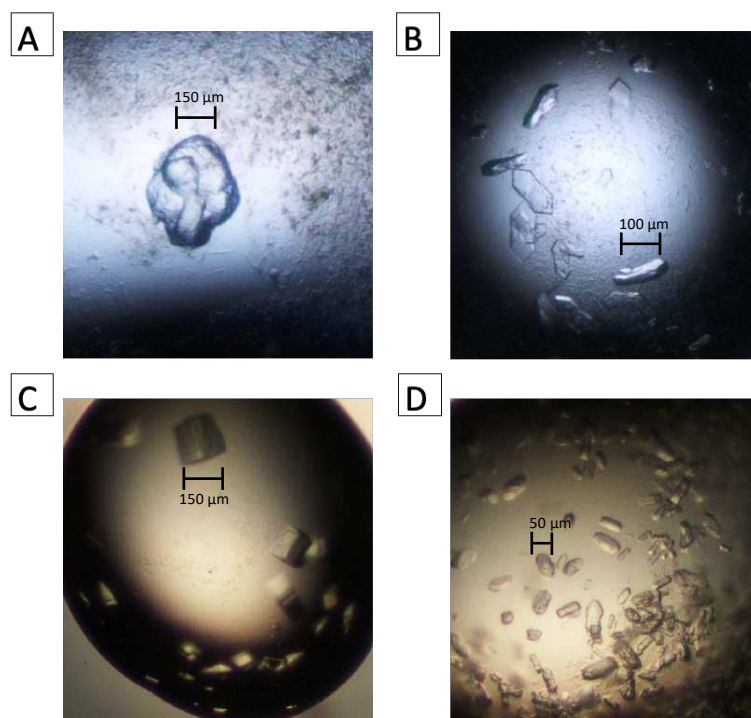


Figure 37 : Cristallisation de Trn1ΔH8.

Les premiers cristaux de Trn1ΔH8 obtenus par criblage large : les cristaux irréguliers de Trn1ΔH8 obtenus à 4°C dans les conditions (A) (0,10 M Hepes pH 7,00, 0,50 % (v/v) Jeff ED-2001, 0,10 M Na₂ Malonate) et (B) (0,14 M K₂HPO₄, 1,26 M NaH₂PO₄). Cristaux irréguliers de Trn1ΔH8 obtenus à 20°C dans les conditions (C) (35 % (v/v) TACSIMATE) et (D) (0,10 M Hepes pH 7,00, 0,50 % (v/v) Jeff ED-2001, 0,10 M Na₂ Malonate).

J'ai ensuite procédé à l'optimisation de ces cristaux dans des boîtes 24 puits en utilisant la méthode de cristallisation par diffusion de vapeur en gouttes suspendues. Les gouttes ont été faites en mélangeant 1,5 μL de protéine (12–17 mg/mL) et 1,5 μL de solution de cristallisation (20–40 % (v/v) TACSIMATE). Grâce à ces optimisations, deux nouvelles formes cristallines ont été obtenues. La première forme est apparue dans des solutions contenant 30 à 37 % (v/v) de TACSIMATE (la protéine concentrée à 15 mg/mL). Les cristaux adoptaient une forme quadratique allongée et formaient un cluster (Figure 38A). Ces derniers étaient fragiles mais leur pouvoir de diffraction était meilleur, autour de 4 Å. Afin d'obtenir des cristaux avec une meilleure qualité de diffraction, différentes concentrations de protéines et différents ratios de goutte ont été testés manuellement. La deuxième forme cristalline est ainsi apparue en augmentant la concentration de la protéine cristallisée (la protéine concentrée à 17 mg/mL). Ces cristaux adoptaient une forme plus fine, en baguette (Figure 38B). Les conditions dans lesquelles les cristaux sont apparus correspondent à 20-25 % (v/v) TACSIMATE.

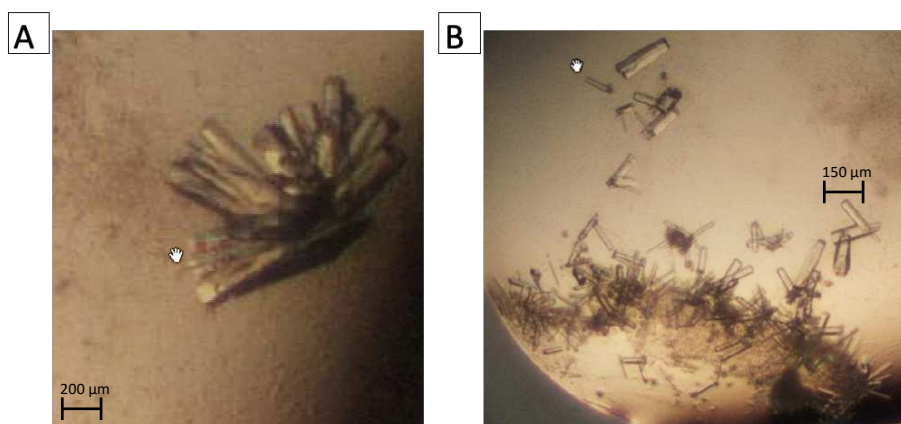


Figure 38 : Optimisation de la cristallisation de Trn1ΔH8.

(A) Cristaux en cluster (quadratique allongée) de Trn1ΔH8 obtenus après optimisation des cristaux obtenus dans 35 % (v/v) TACSIMATE. (B) Cristaux de Trn1ΔH8 en forme de baguette obtenus 20- 25 % (v/v) TACSIMATE, après optimisation des cristaux obtenus en (A).

1.3.2 Résolution de la structure cristallographique de Trn1ΔH8

Deux méthodes de cryoprotection ont été utilisées lors du prélèvement des cristaux de Trn1ΔH8. La première consistait à incuber rapidement les cristaux dans une goutte de solution de cristallisation contenant 25 à 30 % (v/v) de glycérol. La seconde méthode utilisée consistait à laisser sécher la goutte quelques minutes avant de prélever les cristaux afin d'éviter au maximum la manipulation des cristaux. Ainsi, l'évaporation de la goutte permettait de cryoprotéger les cristaux avant de les congeler dans l'azote liquide. Il est important de noter que cette approche fonctionne bien dans notre cas car la solution initiale de Trn1ΔH8 contient 20 % (v/v) de glycérol. Le meilleur jeu de données a été obtenu en laissant sécher la goutte avant le prélèvement du cristal (technique de la « goutte sèche »). Ce cristal a été obtenu avec 20-25 % (v/v) TACSIMATE.

1.3.2.1 Collecte et traitement des données de diffraction

La collecte du jeu de données a été effectuée à l'ESRF sur la ligne de lumière ID23-1. Grâce à la forme en baguette du cristal, il a été possible de réaliser une collecte hélicoïdale permettant d'obtenir un jeu de meilleure qualité. Le jeu de données a été indexé et intégré avec les scripts XDSME associés au logiciel XDS. Ce logiciel a déterminé le groupe d'espace $P2_12_12_1$ comme étant le groupe d'espace le plus probable. C'est un groupe d'espace retrouvé parmi certaines structures de Trn1 déjà résolues (66, 77). Le jeu de données était assez fortement anisotrope, avec une diffraction allant jusqu'à $\sim 1,8$ Å dans une dimension et allant jusqu'à $\sim 2,1-2,3$ Å dans

les deux autres dimensions. Une coupure « classique » aurait amené à couper ce jeu à une résolution $\sim 2,0$ - $2,1$ Å, mais avec des statistiques médiocres à haute résolution. La mise à l'échelle des données a de ce fait été réalisée avec le serveur STARANISO afin de corriger l'anisotropie (<http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi>). Les réflexions ont ainsi été réduites, mises à l'échelle et coupées de manière anisotrope. Les statistiques obtenues sont satisfaisantes et acceptables jusqu'à une résolution de $1,81$ Å, le coefficient $CC_{1/2}$ dans la dernière tranche de résolution étant autour de $0,7$, supérieur à la limite de $0,3$ (198) (Tableau 10). C'est la meilleure résolution obtenue pour une structure cristalline de Trn1FL ou Trn1ΔH8. En effet, les structures de Trn1 disponibles dans la PDB (avec ou sans partenaire) ont une limite de résolution allant de $4,05$ à $2,30$ Å, avec une résolution moyenne de $3,05$ Å.

Collecte des données	Trn1ΔH8
Beamline	ID23-1 (ESRF)
Longueur d'onde (Å)	0,7293
Limite de résolution (Å)	44,7 – 1,81 (2,02 – 1,81)
Groupe d'espace	P 2 ₁ 2 ₁ 2 ₁
Paramètres de maille	
a b c (Å)	a = 71,07 b = 135,28 c = 109,03
α β γ (°)	α = β = γ = 90
Multiplicité	3,5 (3,2)
Réflexions mesurées	210926 (9504)
Réflexions uniques	59803 (2990)
R _{merge} (%)	0,053 (0,602)
R _{pim} (%)	0,034 (0,039)
Complétude (%)	
sphérique	61,7 (11,1)
ellipsoïdale	92,3 (74,7)
I/s(I)	13,6 (1,90)
CC _{1/2}	0,99 (0,69)

Tableau 10 : Les statistiques du jeu de données de Trn1ΔH8.

Le tableau regroupe les statistiques (XDS et STARANISO) du jeu de données de Trn1ΔH8. Les valeurs entre parenthèses correspondent aux statistiques dans la dernière tranche de résolution ($2,02$ – $1,81$ Å).

1.3.2.2 Phasage par remplacement moléculaire

Une fois traité par XDS et STARANISO, la qualité du jeu de données a été évaluée avec le programme X-triage (PHENIX). Nous avons ensuite procédé au phasage par remplacement moléculaire afin d'obtenir un modèle en accord avec le jeu de données.

La première étape a été la détermination du coefficient de Matthews qui indique le nombre de copies du modèle à chercher dans l'unité asymétrique lors de la procédure du remplacement moléculaire. Le calcul du coefficient de Matthews nous a permis d'estimer à un, le nombre de monomère dans l'unité asymétrique (Tableau 11).

Nombre de molécules / unité asymétrique	Coefficient de Matthews	Pourcentage de solvant	probabilité
1	2,66	53,79	1,00
2	1,33	7,59	0,00

Tableau 11 : Détermination du coefficient de Matthews pour la structure cristalline de Trn1ΔH8. Le tableau présente la probabilité du nombre de molécules par unité asymétrique (déterminée par la méthode Matthews (185)).

Du fait de la flexibilité structurale intrinsèque des protéines à répétition HEAT formant des superhélices, le remplacement moléculaire à partir d'une structure de Trn1 de la PDB ne donne pas de bonne solution. En effet, la structure de la super-hélice peut être plus ou moins étirée en fonction des structures et il faut tenir compte de cette flexibilité pour obtenir un bon modèle de remplacement moléculaire. Nous avons donc commencé par construire un modèle initial par déformation par modes normaux à partir d'une structure connue de Trn1 (code PDB 2Z5K) (199). Cette construction d'un modèle initial par mode normaux a été réalisée à l'aide du serveur elNémo (<http://www.sciences.univ-nantes.fr/elnemo/nmode.cgi>), permettant ainsi d'obtenir un ensemble de structures. Cet ensemble a ensuite été utilisé comme modèle pour le phasage par remplacement moléculaire à l'aide du programme PHASER MR (CCP4). La meilleure solution en sortie de remplacement moléculaire présente un score TFZ de 16,5 indiquant une solution hautement probable.

1.3.2.3 Construction du modèle

Un premier affinement de type « jelly-body » avec le programme REFMAC5 (CCP4) a été réalisé à la suite du remplacement moléculaire (200). A ce stade, le modèle obtenu semblait être correct au centre de la protéine, au niveau des HEAT H4-H17. Par contre, les HEAT en N- et C-terminal, respectivement H1-H3 et H18-H20, présentaient des densités électroniques de très mauvaise qualité, les hélices étaient mal placées. Le modèle a ensuite été reconstruit avec le programme ARP/WARP (CCP4) (201). Le modèle en sortie de ce programme était un modèle incomplet, mais certaines hélices des HEAT en N- et C-terminal étaient mieux

modélisées et présentaient une meilleure densité électronique. Un nouveau modèle a alors été construit par tranches, en réalisant une superposition de différentes structures. Ce modèle a été construit d'une part à partir du modèle de référence (PDB 2Z5K) pour les parties manquantes du modèle ARP/WARP, et d'autre part à partir du modèle ARP/WARP pour les parties bien reconstruites. Ce modèle composite a ensuite été utilisé pour un deuxième remplacement moléculaire qui a permis d'obtenir un score TFZ de 87,3. A noter qu'en sortie du deuxième remplacement moléculaire et avant le premier affinement, le set Rlibre a été transféré du précédent remplacement moléculaire à ce nouveau modèle (pour ne pas introduire de biais dans l'affinement et le calcul du Rlibre).

1.3.2.4 Affinement et validation du modèle

Le modèle obtenu a ensuite été affiné, il présente globalement une densité électronique bien définie sur l'ensemble de la structure, du N- au C-terminal. Il a été corrigé et reconstruit manuellement dans le logiciel COOT avec des cycles itératifs de correction/affinement dans REFMAC5 jusqu'à des facteurs d'accord Rfacteur = 20,5 % et Rlibre = 25,4 % (189) (Tableau 12). Le modèle a été validé avec le programme MOLPROBITY ainsi que le serveur PDB-Validation Server qui indiquent des statistiques correctes (Tableau 12).

Affinement de la structure	Trn1ΔH8
Limite de résolution (Å)	44,7 – 1,81 (2,02 – 1,81)
No. réflexions	59787 (3052)
Rfacteur	20,5 (24,9)
Rlibre	25,4 (27,6)
No. atomes	6739
RMSD	
Longueur de liaison(Å)	0,007
Angle de liaison (°)	1,43
Ramachandran	
Favorisé (%)	98,28
Autorisé (%)	1,72
Non autorisé (%)	0
Score Molprobitiy	
Score	1,44
Clashscore, tous atomes	8,03

Tableau 12 : Les statistiques d'affinement de la structure de Trn1ΔH8.

Tableau regroupant les statistiques (REFMAC et MOLPROBITY) d'affinement de Trn1ΔH8. Les valeurs entre parenthèses correspondent aux statistiques dans la dernière tranche de résolution (2,02 - 1,81 Å).

1.3.2.5 Analyse structurale du modèle

Seules deux structures de la protéine Trn1 sans partenaires ont été résolues par cristallographie. Il s'agit des structures de Trn1FL, codes PDB 2Z5J et 2QMR (77, 81), dont les limites de résolution sont respectivement 3,4 Å et 3,0 Å. La résolution de notre structure permet une meilleure observation des chaînes latérales dont les cartes de densités sont plus facile d'interprétation, notamment au niveau de l'extrémité C-terminale (Figure 39) qui présente une importante flexibilité intrinsèque.

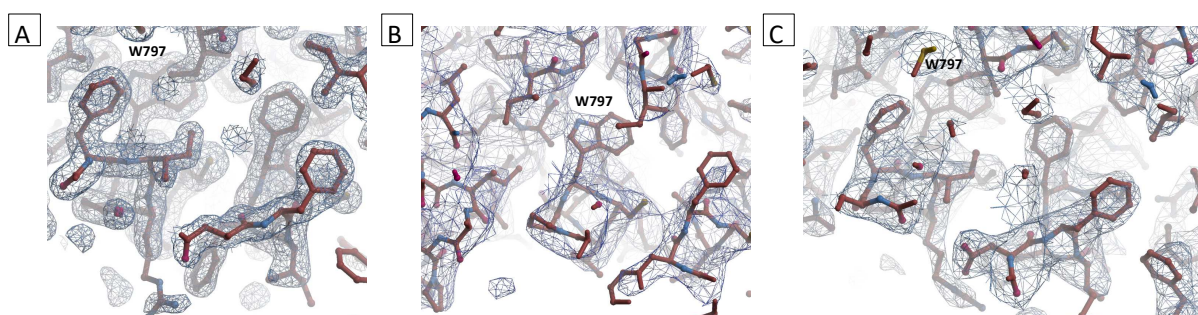


Figure 39 : Comparaison des cartes de densité électronique des structures cristallines de Trn1.

Les différentes cartes de densité électronique (carte 2Fo-Fc, contours à 1σ) des structures cristallographiques de Trn1 seule sont comparées au niveau de l'extrémité C-terminale. (A) Structure de Trn1ΔH8 résolue à 1,81 Å. (B) Structure 2Z5J à 3,40 Å. (C) Structure de 2QMR à 3,00 Å.

Les trois structures de Trn1 sont résolues dans des groupes d'espace différents (Tableau 13). La structure Trn1ΔH8 appartient au système orthorhombique avec un groupe d'espace $P2_12_12_1$ tandis que la structure 2Z5J avec un groupe d'espace $P4_12_12$ appartient au système tétragonal. Quant à la structure 2QMR, elle appartient au système monoclinique avec un groupe d'espace $P2_1$. Par ailleurs, le pourcentage de solvant dans les différentes structures pourrait expliquer la différence de résolution de Trn1ΔH8 avec les deux autres structures. En effet, la structure de Trn1ΔH8 présente un plus faible pourcentage de solvant que les deux autres structures, bien que la différence avec la structure 2Z5J ne soit pas très importante (Tableau 13).

Protéine	Code PDB	Groupe d'espace	Paramètres de maille (Å , °)	Résolution (Å)	Pourcentage de solvant (%)
Trn1ΔH8	-	P2 ₁ 2 ₁ 2 ₁	a = 71,07 b= 135,28 c=109,03 $\alpha = \beta = \gamma = 90^\circ$	1,81	53,79
Trn1FL	2Z5J	P4 ₁ 2 ₁ 2	a = b= 107,47 c=194,772 $\alpha = \beta = \gamma = 90^\circ$	3,40	55,64
Trn1FL	2QMR	P12 ₁ 1	a = 129,34 b= 168,89 c=140,96 $\alpha = \gamma = 90^\circ \quad \beta = 93,18^\circ$	3,00	67,54

Tableau 13 : Tableau comparatif des trois structures cristallographiques de Trn1.

Les paramètres de maille ainsi que les groupes d'espace sont différents pour les trois structures. La différence de pourcentage de solvant au sein des différents cristaux pourrait expliquer la différence de résolution entre les différentes structures.

Il faut noter que les différentes structures de Trn1 ne présentent pas toutes le même arrangement au sein du cristal (Figure 40). La structure de Trn1ΔH8 (Figure 40A) est plus compacte que les deux autres structures, 2Z5J (Figure 40B) et 2QMR (Figure 40C). Les canaux de solvant sont plus larges pour les structures 2Z5J et 2QMR alors que ceux de Trn1ΔH8 sont plus étroits. Au-delà de l'empilement cristallin, la méthode de cryoprotection de la « goutte sèche » (voir paragraphe 1.3.2 page 125) que j'ai utilisée lors de la congélation du cristal de Trn1ΔH8 a vraisemblablement eu un impact positif sur le pouvoir de diffraction du cristal.

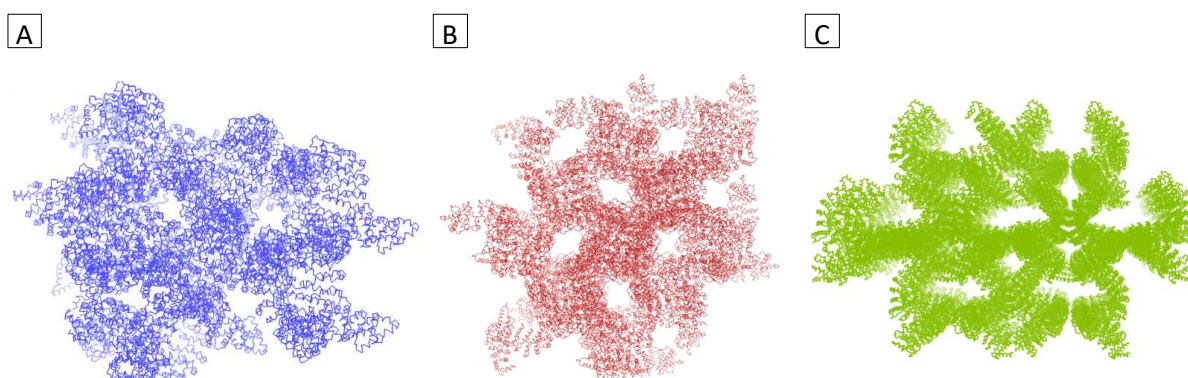


Figure 40 : Comparaison de l'empilement cristallin des structures de Trn1.

Les structures cristallographiques de (A) Trn1ΔH8 (en bleu), (B-C) des Trn1FL 2Z5J (en rouge) et 2QMR (en vert) sont représentées en mode ruban. La structure de Trn1ΔH8 présente un empilement cristallin plus compact et des canaux de solvant moins grands par rapport aux deux autres structures.

La structure cristallographique de Trn1ΔH8 révèle bien la structure en « S-like » de la Transportine 1 comprenant la totalité des 20 HEAT qui la constituent. Seule la boucle reliant les HEAT H4 et H5 dont la densité n'est pas observée, est manquante dans notre structure (Figure 41A).

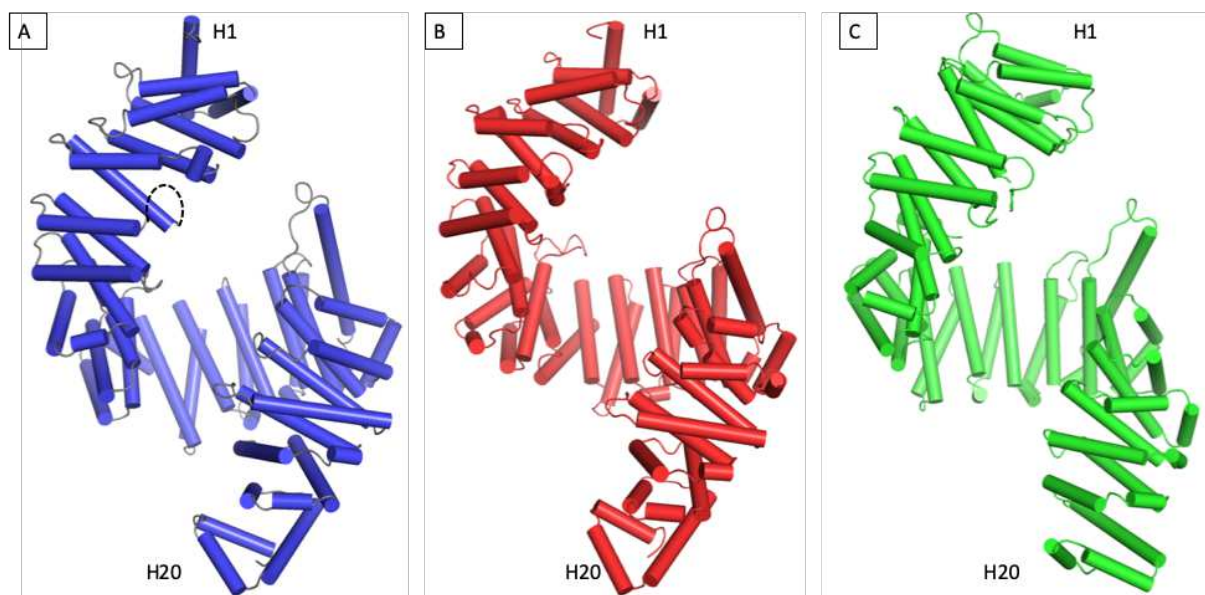


Figure 41 : Comparaison des structures cristallographiques de Trn1.

Les différentes structures cristallographiques de Trn1 sont représentées en mode cartoon. (A) La structure de Trn1ΔH8 est en bleu, en pointillée la boucle reliant les HEAT H4 et H5 dont la densité n'est pas observée. (B) Structure 2Z5J en rouge et (C) Structure 2QMR en vert.

La comparaison des différentes structures de Trn1 permet de souligner l'effet « ressort » que possède cette protéine. Trn1ΔH8 est légèrement plus allongée que les deux autres structures. Selon l'organisation de la protéine dans le cristal, elle s'adapte et grâce à sa flexibilité intrinsèque, elle se retrouve plus ou moins étirée.

Globalement, la structure de Trn1ΔH8 (Figure 41A) est légèrement plus proche de la structure 2Z5J (Figure 41B) que de la structure 2QMR (Figure 41C), avec un RMSD calculé sur les Cα d'environ 6,5 Å (Tableau 14). Les différences résident principalement au niveau des extrémités N- et C-terminales de la protéine qui présentent la plus grande flexibilité structurale dans les différentes structures de Trn1. Il faut noter que le HEAT H1 n'est pas présent dans la structure 2QMR.

Structure	Chain	atoms	residues	Mean B-Iso	RMSD (Å)
Trn1DH8	A	6566	822	37,02	-
2Z5J	A	6652	835	127,23	6,487
2QMR	B	6308	801	73,73	6,889

Tableau 14 : Tableau comparatif des facteurs B des structures de Trn1.

La comparaison des facteurs B des différentes structures montre bien les différences de flexibilité intrinsèque de Trn1.

La flexibilité intrinsèque précédemment décrite (81) au niveau des extrémités N- (H1-H4), et C- (H17-H20) terminales est aussi observée au niveau des valeurs des facteurs B des trois structures de Trn1 (Figure 42) qui adopte des valeurs très différentes d'une structure à l'autre. La flexibilité intrinsèque de ces régions permet à Trn1 d'adapter sa structure en fonction du partenaire à reconnaître. La région N-terminale reconnaît la petite protéine Ran et la région C- terminale reconnaît les différents cargos à transporter. Par ailleurs, le centre de la protéine reste assez homogène.

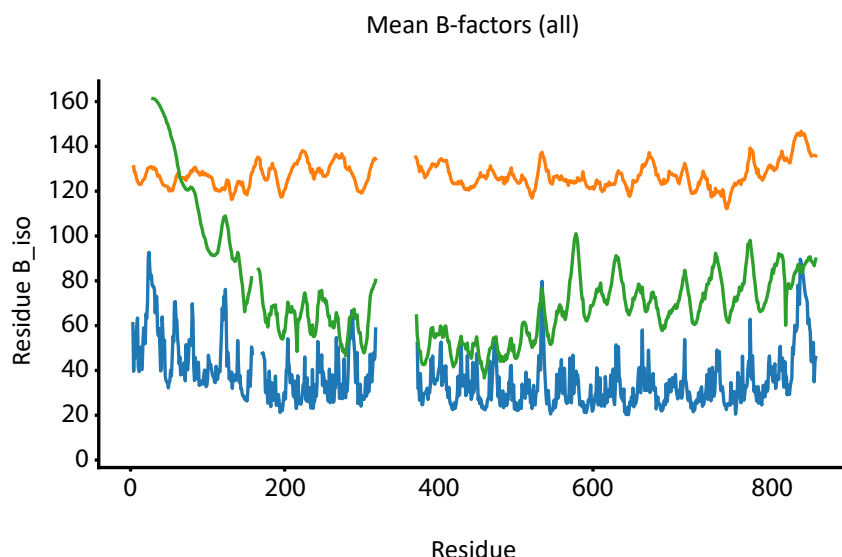


Figure 42 : Représentation graphique comparative des facteurs B des structures cristallographiques de Trn1.

Le graphique représente les facteurs B (chaines principales et chaines latérales) des différentes structures de Trn1 (Trn1ΔH8 en bleu, 2Z5J en orange, 2QMR en vert) en fonction des résidus.

Les différences observées au niveau de l'extrémité C-terminale (HEAT H18-H20) viennent essentiellement de l'état étiré-allongé des trois structures. Ces différences concernent principalement les chaines latérales sauf au niveau du HEAT H20 pour lequel la superposition des hélices des trois structures n'est pas possible (RMSD calculé sur les Cα de H20 est d'environ 1,2 Å et 0,98 Å entre Trn1ΔH8 et respectivement 2Z5J et 2QMR) (Figure 43A). La différence entre le H20 des trois structures s'explique par un empilement cristallin différent. Le H20 de la structure Trn1ΔH8 interagit avec les HEAT H5-H7 au niveau de l'arche N-terminale du symétrique (Figure 43B). Le H20 de la structure 2Z5J est plus contraint et interagit avec les HEAT H8-H11 au niveau de la partie centrale du symétrique (Figure 43C). Le H20 de la structure 2QMR est plus libre et interagit avec la boucle reliant les deux hélices A et B du HEAT H1 à l'extrémité N-terminale du symétrique (Figure 43D). Les différences de contacts

du H20 de chaque structure expliquent les diff rences observ es qui sont dues   un empilement cristallin diff rent entre les trois structures.

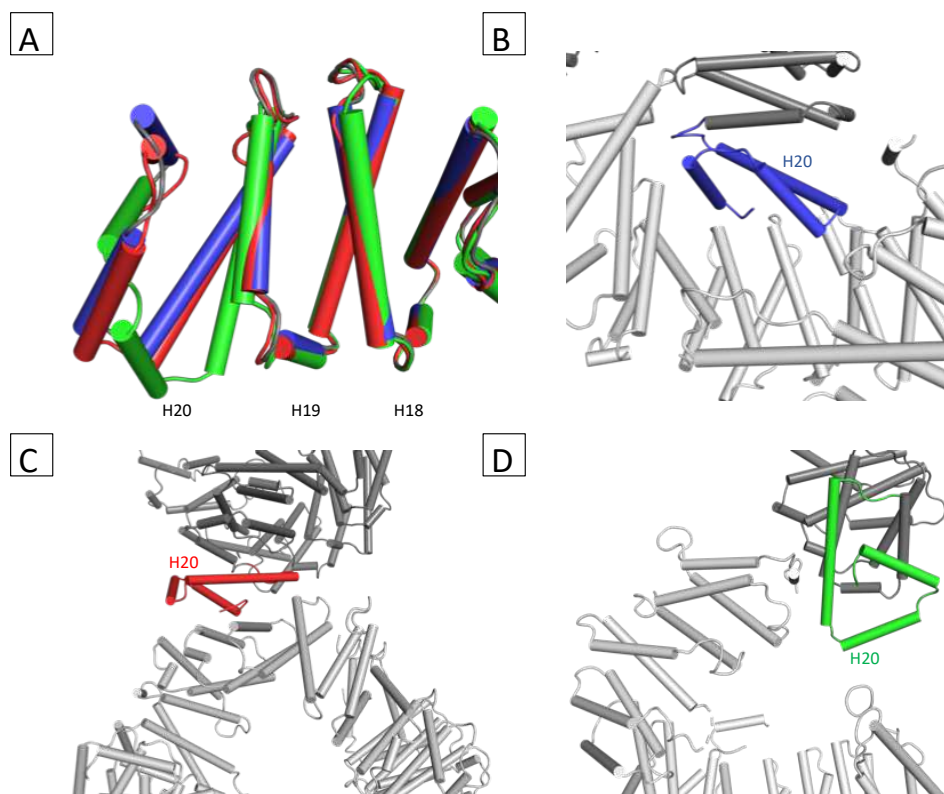


Figure 43 : Comparaison des HEAT H18-H20 des structures cristallographiques de Trn1.

(A) Superposition des HEAT H18-H20 des trois structures : Trn1ΔH8 (en bleu), 2Z5J (en rouge) 2QMR (en vert). (B-D) Empilement cristallin des HEAT H20 des trois structures : Trn1ΔH8 (B), 2Z5J (C), 2QMR (D) qui montre des interactions diff rentes du HEAT H20. Les sym triques sont en gris clairs, le reste de la structure (HEAT H1-H19) est en gris.

Peu de diff rences sont observ es au niveau de l'extr mit  N-terminale (HEAT H1-H4), les h lices (A et B) des diff rents HEAT sont superposables dans l'ensemble (RMSD calcul  sur les C  de cette r gion est d'environ 0,97   et 0,72   entre Trn1ΔH8 et respectivement 2Z5J et 2QMR) sauf au niveau du HEAT H1 (Figure 44A). Les diff rences concernent essentiellement les connexions entre les diff rents HEAT. L'h lice B du HEAT H3 est reli e   l'h lice A du H4 par une petite h lice (H₃₋₄) dans la structure Trn1ΔH8, cette connexion est r alis e par une boucle dans les deux autres structures (Figure 44B). Par ailleurs, la boucle manquante qui relie l'h lice B du H4 et l'h lice A du H5 dans la structure Trn1ΔH8 est incompl te dans la structure 2QMR. Ces deux h lices sont connect es par une petite h lice (H₄₋₅) dans la structure 2Z5J (Figure 44B). La diff rence observ e au niveau de l'h lice (H₃₋₄) peut s'expliquer par un empilement cristallin diff rent. Dans la structure Trn1ΔH8, cette r gion est contrainte par la

partie centrale du sym trique (HEAT H12-H14) alors qu'elle est plus libre dans les deux autres structures (Figure 44C-D).

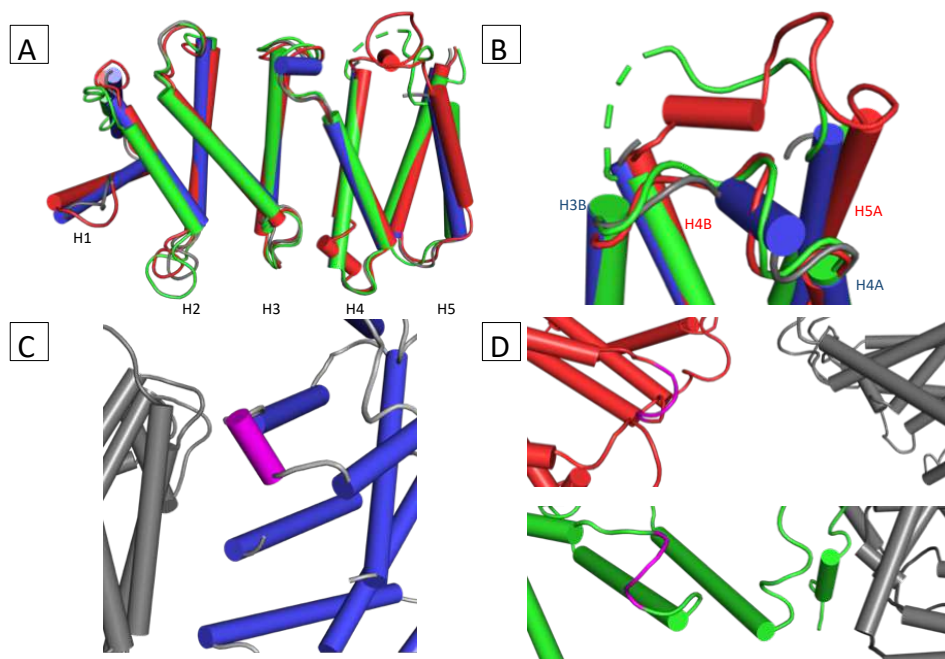


Figure 44 : Comparaison des HEAT H1-H4 des structures cristallographiques de Trn1.

De petites diff rences sont observ es pour les trois structures : Trn1ΔH8 (en bleu), 2Z5J (en rouge) 2QMR (en vert). (A) La superposition des HEAT H1-H4. (B) La diff rence de connexion entre les h lices des HEAT H3 et H4, et H4 et H5. (C-D) Empilement cristallin des trois structures.

Par ailleurs, les interactions et les associations entre les diff rents HEAT restent inchang es dans le reste de la structure. Les m mes associations (les h lices A et B sont connect es par les m mes  l ments, petites h lices ou boucles) sont retrouv es dans les diff rentes structures et m me au niveau de l'extr mit  C-terminale qui pr sente la plus grande variabilit .

1.3.3 Cristallisation du complexe Trn1ΔH8/ADAR1court

Apr s avoir r solu la structure cristallographique de Trn1ΔH8, nous nous sommes ensuite concentr s sur la co-cristallisation du complexe Trn1ΔH8/ADAR1court (ADAR1_708-801). La co-cristallisation des deux partenaires est r alis e par la mise en pr sence de Trn1ΔH8 (0,15 mM) et ADAR1court (0,15-0,30 mM) avant la r alisation des bo tes de cristallisation.

Ces exp riences de co-cristallisation ont permis l'obtention de cristaux qui sont apparus dans les trois conditions suivantes : (0,80 M Na₂ Succinate), (35 % (v/v) TACSIMATE) et (0,10 M Hepes pH 7,00, 0,50 % (v/v) Jeff ED-2001, 1,10 M Na₂ Malonate pH 4,0). Diff rentes formes de cristaux ont pu  tre observ es (des baguettes, des prismes, des plaques...). Apr s cong lation

et tests des cristaux au synchrotron, plusieurs jeux de données ont été collectés et traités. Cependant, aucun des jeux de données traités n'a pu révéler la présence de la densité électronique du partenaire ADAR1court. Ils contenaient seulement la Transportine 1. Nous n'avons pas pu résoudre la structure du complexe Trn1 Δ H8/ADAR1court par cristallographie aux rayons X.

2 Caractérisation du complexe Trn1 Δ H8/ADAR1_pep

Puisque nous ne sommes pas parvenus à produire des cristaux du complexe Trn1 Δ H8/ADAR1court, nous avons ensuite utilisé un peptide mimant le NLS minimal d'ADAR1 pour apporter des informations structurales sur le complexe Trn1/ADAR1. Une précédente étude a montré qu'un peptide qui porte uniquement les deux modules (séparés par un petit espaceur) constituant le NLS d'ADAR1 (ADAR1_pep) pouvait être importé dans le noyau (174).

2.1 Étude de l'interaction Trn1 Δ H8/ADAR1_pep par DSF et GST Pull-down

Avant de mener des études cristallographiques du complexe Trn1 Δ H8/ADAR1_pep, nous avons entrepris des études biochimiques pour caractériser l'interaction entre les deux partenaires *in vitro*.

Dans un premier temps, nous avons mené des expériences de DSF pour analyser l'effet de l'ajout du peptide ADAR1_pep sur Trn1 Δ H8. Pour cela, nous avons mesuré la température de fusion (T_m) de Trn1 Δ H8 qui caractérise sa stabilité thermique, seule et suite à l'ajout du peptide ADAR1_pep à différentes concentrations. La comparaison des T_m mesurés permet d'analyser s'il y a interaction entre ces deux partenaires.

En absence du peptide ADAR1_pep, Trn1 Δ H8 a un T_m de $59,9 \pm 0,2^\circ\text{C}$ (Figure 45A). En présence du peptide ADAR1_pep, le T_m de Trn1 Δ H8 ne subit aucun changement, on observe un T_m de $59,5 \pm 0,7^\circ\text{C}$ lorsque les deux partenaires sont ajoutés à équimolarité. Cela signifie que la présence d'ADAR1_pep ne stabilise pas Trn1 Δ H8, ce qui suggère qu'ADAR1_pep n'interagit pas avec Trn1 Δ H8, du moins dans les conditions dans lesquelles les expériences sont réalisées. L'augmentation de la quantité du peptide ADAR1_pep par rapport à la protéine Trn1 Δ H8, allant jusqu'à un ratio 1:20 ne permet toujours pas l'observation d'une augmentation de la valeur du T_m , puisque la valeur de T_m de Trn1 Δ H8 pour un ratio 1:20 est de $59,8 \pm 0,2^\circ\text{C}$

(Figure 45A). Les différentes expériences de DSF menées n'ont pas permis de mettre en évidence une interaction entre les deux partenaires.

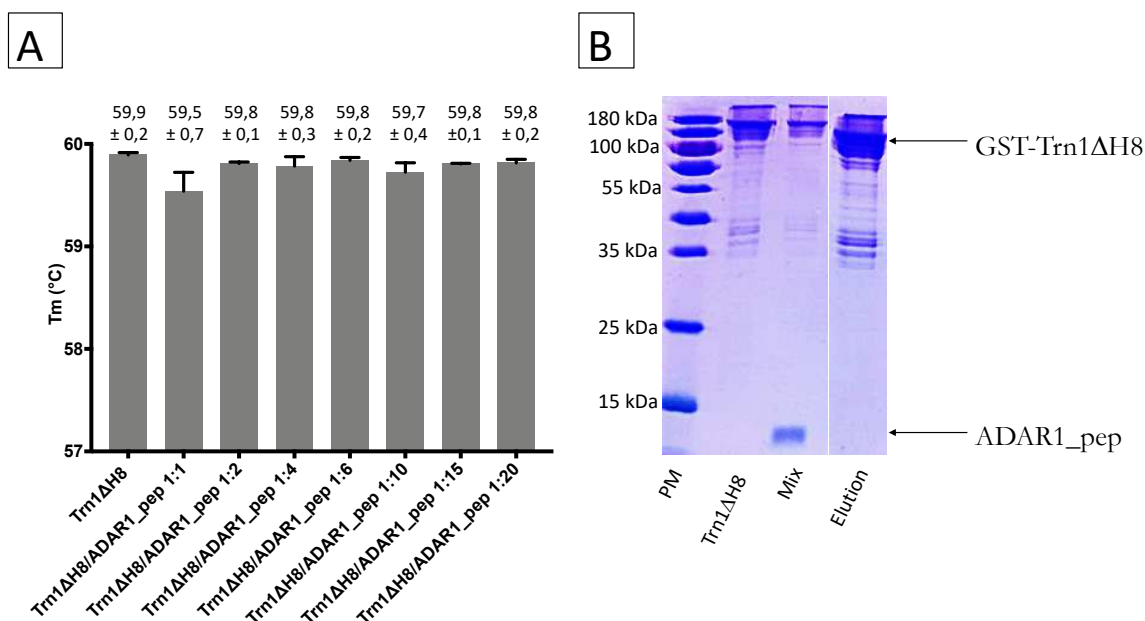


Figure 45 : Étude biochimique de l'interaction Trn1ΔH8/ADAR1_pep.

(A) Effet de la fixation du peptide ADAR1_pep sur la stabilité de Trn1ΔH8. Les températures de fusions (T_m) sont mesurées à une concentration finale de 2 μ M pour Trn1ΔH8 et une gamme de concentration du peptide allant de 2 à 40 μ M. **(B) Analyse SDS-PAGE des expériences de GST Pull-down.** Les expériences sont effectuées avec Trn1ΔH8 concentrée à 27 μ M en présence du peptide ADAR1_pep concentrée à 216 μ M (ratio 1:8). L'analyse est faite par SDS-PAGE 16 % (PM = poids moléculaire en kDa, mix = mélange binaire des deux partenaires, Éluat = éluat de l'expérience GST Pull-down).

Dans un second temps, d'autres expériences pour la mise en évidence de l'interaction entre Trn1ΔH8 et ADAR1_pep ont été menées en utilisant la méthode du GST pull-down. Ces expériences ont été réalisées avec Trn1ΔH8 fusionnée à la GST (GST-Trn1ΔH8). Le peptide a été ajouté à la protéine GST-Trn1ΔH8 complexée aux billes d'agarose, le mélange a ensuite été incubé. Puis, l'éluat du matériel fixé a été effectuée dans le tampon de charge. Enfin, l'analyse de ces expériences a été réalisée sur un gel SDS-PAGE pour déterminer si le peptide a pu interagir avec la protéine GST-Trn1ΔH8 retenue sur les billes agaroses. Le contenu de l'éluat analysé a permis de montrer que le peptide ADAR1_pep ne s'est pas accroché à la protéine GST-Trn1ΔH8. En effet, sur le gel (Figure 45B), seule la bande représentant GST-Trn1ΔH8 (ainsi que les contaminants bactériens) est observée.

Ces résultats montrent donc qu'il n'y a pas d'interaction entre Trn1ΔH8 et ADAR1_pep dans les conditions testées. Il est possible que l'interaction des deux partenaires ne soit pas observable car il s'agit d'une interaction de faible affinité. Une autre raison serait que le NLS minimal porté par le peptide ADAR1_pep, bien que fonctionnel *in vivo*, ne soit pas suffisant pour observer une interaction *in vitro* entre les deux partenaires.

2.2 Utilisation d'une protéine chimère pour la caractérisation de l'interaction Trn1ΔH8/ADAR1_pep

A la suite des résultats obtenus précédemment, nous avons entrepris de construire une protéine chimère qui permet de rapprocher les deux partenaires pour ainsi favoriser la formation du complexe. Sur le principe expliqué au paragraphe 1.4, page 71 (Matériels et Méthodes), nous avons accroché à l'extrémité C-terminale de Trn1ΔH8 le peptide de séquence (*DNQPEGMISESLDNLES***MMPNKVR***gehrqdrKAERM***MGFTEVTPVTGASLRRT**) contenant la séquence minimale du NLS d'ADAR1 en gras, l'espaceur connectant les deux modules du NLS en lettres minuscules et les résidus flanquants supplémentaires en bleu et italique. Nous avons construit la protéine chimère Trn1ΔH8-ADAR1_pep en considérant que le peptide ADAR1_pep se lie au niveau de l'arche C-terminale de Trn1ΔH8 (Figure 26). La séquence des parties flanquantes (en amont ou en aval du NLS) est celle des résidus présents dans la protéine ADAR1 de part et d'autre du dsRBD3. Ces parties correspondent à des fragments non structurés et apportent de la flexibilité au NLS qui pourrait ainsi en principe se fixer librement au niveau de son site d'interaction. De même une deuxième chimère « contrôle » a également été construite, la chimère Trn1ΔH8-A1_pep. Cette dernière est composée de Trn1ΔH8 à laquelle la séquence M9 (PY-NLS (78)) de la protéine hnRNP A1 est accrochée à l'extrémité C-terminale de la protéine (Figure 26A). Cette dernière a servi en quelque sorte de contrôle pour l'évaluation de cette approche car l'interaction de ce PY-NLS avec la Transportine 1 est bien caractérisée structuralement.

2.2.1 Préparation et purification des protéines chimères

Pour les deux chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep, le protocole suivi pour l'expression est celui de la protéine Trn1ΔH8 avec de légères modifications, notamment l'augmentation de la concentration d'IPTG (paragraphe 1.4.2, page 74). Ce protocole d'expression a permis une surexpression suffisante des chimères (Figure 46) pour la purification de celles-ci.

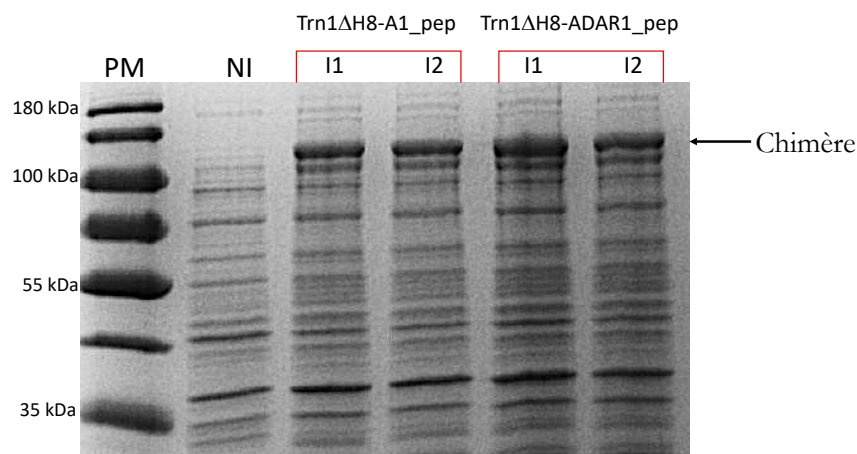


Figure 46 : Expression des chimères Trn1ΔH8-A1_pep et Trn1ΔH8-ADAR1_pep.

Les résultats obtenus pour les expressions de ces chimères sont révélés par analyse SDS-PAGE 16 % (PM = marqueur de poids moléculaire, NI = l'extrait bactérien total avant l'induction, I1=I2 = l'extrait bactérien total après 20 h d'induction à 30 °C (les quantités de matériel déposées sont identiques pour I1 et I2)).

2.2.1.1 Purification des chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep

Le protocole de purification des deux chimères, Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep est le même que celui de Trn1ΔH8, et consiste en trois étapes de purification et une étape de coupure de l'étiquette GST.

Concernant la première colonne d'affinité glutathion sepharose, les différentes chimères ont été éluées en injectant directement le tampon d'éluion (tampon B) sur la colonne. Après l'éluion, les fractions contenant la chimère d'intérêt ont été rassemblées, et une coupure de l'étiquette GST a été effectuée par la protéase HRV 3C. Lors de la deuxième colonne d'échangeuse d'ions, l'éluion de Trn1ΔH8-ADAR1_pep se fait distinctement de l'étiquette GST à 120 mM de NaCl et celle de Trn1ΔH8-A1_pep se fait conjointement avec la GST à 65 mM de NaCl. Les derniers contaminants persistants ont été éliminés par chromatographie d'exclusion stérique (Sup200). Les deux chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep sont respectivement éluées après passage de 175 mL et 150 mL de tampon A. Les différentes étapes de purification ont été analysées par SDS-PAGE qui montre qu'en fin de purification, les différentes chimères Trn1ΔH8-ADAR1_pep (Figure 47A) et Trn1ΔH8-A1_pep (Figure 47B) sont suffisamment pures pour les différents tests de cristallisation. Les rendements des purifications pour les chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep sont respectivement d'environ 4 mg et 3 mg de protéine pure par litre de culture bactérienne.

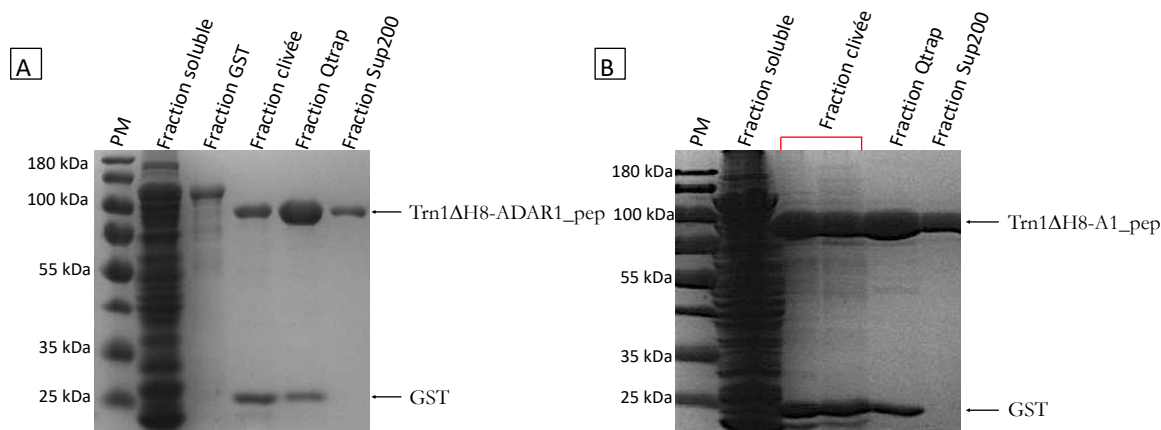


Figure 47 : Purification des chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep.

L'analyse de la purification par SDS-PAGE 10 % (PM = marqueur de poids moléculaire) montre une pureté satisfaisante pour les deux chimères (A) Trn1ΔH8-ADAR1_pep et (B) Trn1ΔH8-A1_pep.

Contrairement à ce qui a pu être observé pour Trn1ΔH8 (Figure 35), lors de la deuxième étape de chromatographie, la colonne échangeuse d'ions ne permet pas de parfaitement séparer les deux chimères de l'étiquette GST. Effectivement, bien qu'élues distinctement, une partie de l'étiquette GST est également retrouvée dans la fraction de Trn1ΔH8-ADAR1_pep. La chimère est moins retenue sur la colonne alors que les valeurs de pI pour Trn1ΔH8 et Trn1ΔH8-ADAR1_pep sont identiques et égales à 5,0. Trn1ΔH8-A1_pep est quant à elle, éluee en même temps que la GST alors que son pI est aussi égal à 5,0. Toutefois, la dernière étape de purification permet quand même d'obtenir des chimères suffisamment pures pour les tests de cristallisation (Figure 47).

2.2.2 Cristallisation des protéines chimères

Concernant la cristallisation de la chimère Trn1ΔH8-A1_pep, à partir d'un criblage large, une première piste avec des cristaux « amorphes » a été obtenue dans la condition 35 % (v/v) TACSIMATE (Figure 48A) avec la chimère Trn1ΔH8-A1_pep concentrée à 15 mg/mL. Nous avons ensuite entrepris l'optimisation de ces cristaux, tout d'abord en affinant la condition dans laquelle ils sont apparus. Cette première étape d'optimisation a permis de produire des microcristaux dans la condition 33 % (v/v) TACSIMATE (Figure 48B). Puis, à partir de cette dernière condition, nous avons réalisé une deuxième étape d'optimisation avec l'utilisation d'un kit d'additifs (ANGSTROM, Molecular Dimensions). Pour cela, chacun des puits contient la solution de réservoir préparée avec 33 % (v/v) TACSIMATE supplémentée de 10 % du kit d'additifs. Cette deuxième étape d'optimisation a permis l'obtention de cristaux en forme de bâtonnets dans la condition 33 % (v/v) TACSIMATE, 4 % (w/v) Triméthylolpropane

(Figure 48 C). Enfin, une derni re  tape d'optimisation a permis la production de cristaux de diff rentes formes dont des  batonnets plus gros (Figure 48D) dans la condition 33 % (v/v) TACSIMATE, 2,50-3,00 % (w/v) Trimethylolpropane.

Les cristaux optimis s (Figure 48D) ont  t  pr lev s et cryoprot g s par incubation dans une goutte de solution de cristallisation contenant 20   25 % (v/v) de glyc rol avant d' tre congel s dans l'azote liquide. Ces cristaux n'ont pas permis de r soudre la structure cristallographique de la chim re Trn1 H8-A1_p p, puisqu'ils diffractaient   tr s faible r solution (7 – 9  ).

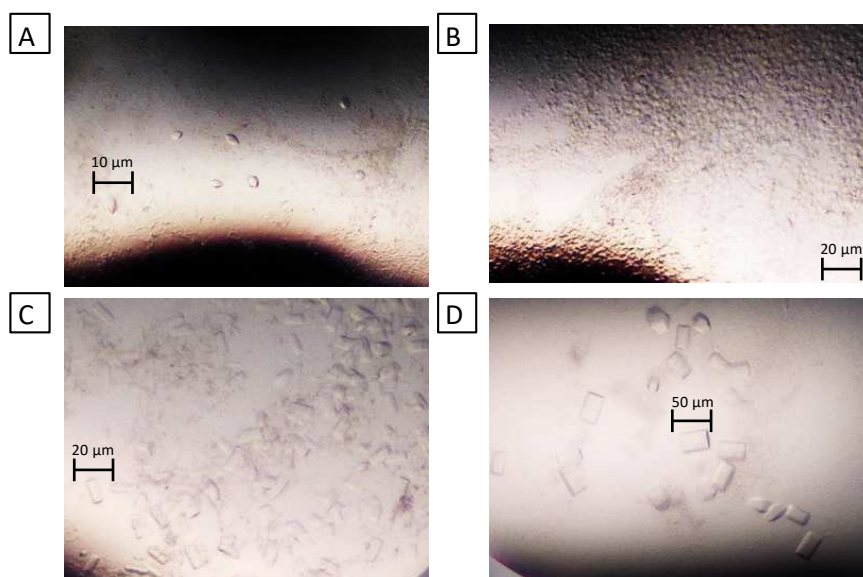


Figure 48 : Cristallisation de la chim re Trn1 H8-A1_p p.

(A) Les cristaux « amorphes » obtenus dans la condition 35 % (v/v) TACSIMATE. (B) Les microcristaux obtenus dans la condition 33 % (v/v) TACSIMATE. (C) Les cristaux en forme de  batonnets fins obtenus dans la condition 33 % (v/v) TACSIMATE, 4 % (w/v) Trimethylolpropane. (D) Les cristaux en forme de  batonnets plus gros obtenus lors de la derni re  tape d'optimisation des cristaux obtenus en (C).

Concernant la cristallisation de la chim re Trn1 H8-ADAR1_p p, malgr  les diff rentes  tapes d'optimisation, nous avons uniquement obtenus des microcristaux dans la condition 0,10 M Na K Phosphate pH 6,20, 25 % (v/v) 1,2 propandiol, 10 % (v/v) glyc rol. Ces microcristaux n'ont pas pu  tre reproduits et cette piste n'a pas  t  poursuivie. La cristallisation de la chim re Trn1 H8-ADAR1_p p n'a pas permis la production de cristaux pouvant  tre pr lev s et test s au synchrotron.

3 Caractérisation du complexe Trn1 Δ H8/ADAR1long

Les approches consistant à utiliser les constructions Trn1 Δ H8 et ADAR1court, ou à utiliser la protéine chimérique Trn1 Δ H8-ADAR1_{pep} n'ont pas permis de résoudre la structure cristallographique du complexe Trn1/ADAR1. Nous avons alors pensé à travailler avec une construction plus longue d'ADAR1. Il s'agissait de favoriser la formation du complexe Trn1/ADAR1 en améliorant l'affinité entre les deux partenaires par l'augmentation de la surface d'interaction. Pour cela, nous avons construit la protéine ADAR1long (ADAR1₆₈₇₋₈₁₇) qui présente le dsRBD3 portant le NLS minimal d'ADAR1 allongé d'une vingtaine de résidus en amont du NLS et une quinzaine de résidus en aval du NLS.

3.1 Purification d'ADAR1long

Les premiers essais de purification de la protéine ADAR1long (ADAR1₆₈₇₋₈₁₇) ont été effectués suivant le protocole de purification d'ADAR1court (ADAR1₇₀₈₋₈₀₁). Cependant, ce protocole n'a pas été optimal pour ADAR1long (ADAR1long-A) puisqu'il n'a pas permis la production de matériel pur. Il a donc été nécessaire de construire une nouvelle version d'ADAR1long (ADAR1long-B) pour l'obtenir suffisamment pure afin de réaliser des tests de cristallisation. Il est par contre possible de purifier la construction ADAR1long-A, en gardant la double étiquette GST-His₆ (GST-ADAR1long), alors que la construction ADAR1long-B, qui ne comprend qu'une étiquette His₆, peut être clivée et purifiée en l'absence de l'étiquette.

3.1.1.1 Purification de GST-ADAR1long.

La purification de GST-ADAR1long s'effectue en trois étapes sans coupure de l'étiquette à partir de la construction ADAR1long-A : une chromatographie d'affinité, une chromatographie d'interaction hydrophobe et une chromatographie d'exclusion stérique.

La fraction soluble a été injectée sur une colonne de nickel sepharose. La protéine GST-ADAR1long interagit sur la résine grâce à sa double étiquette GST-His₆, une partie des contaminants cellulaires a été éliminée grâce à un lavage avec 4 % de tampon D (50 mM Tris- HCl pH 8,0, 1 M NaCl, 0,5 M Imidazole, 1 mM DTT). La protéine est éluée à 175 mM d'Imidazole. A la suite de cette étape, les fractions contenant la protéine ont été regroupées et saturées en (NH₄)₂SO₄ (1 M), l'ensemble a été utilisé pour réaliser une chromatographie d'interaction hydrophobe sur une colonne phényl sépharose.

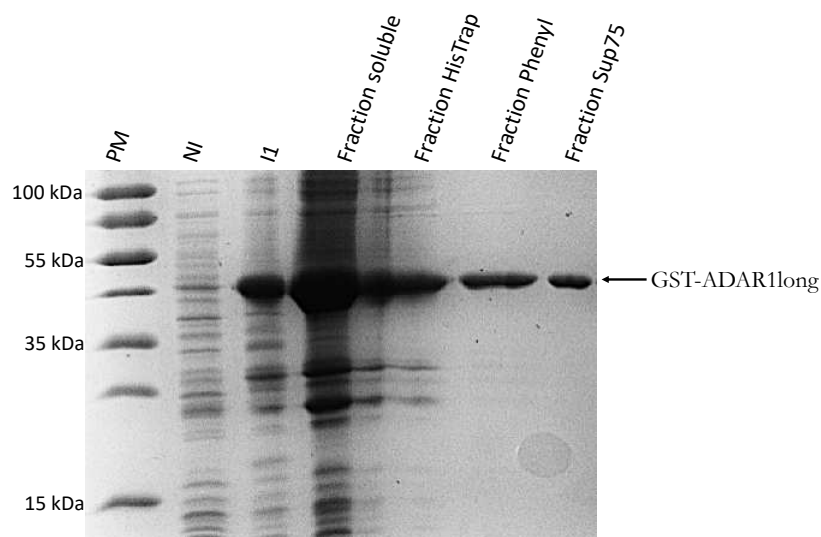


Figure 49 : Purification de GST-ADAR1long.

La purification de GST-ADAR1long est réalisée en trois étapes chromatographiques sans coupure de l'étiquette. L'analyse de la purification par SDS-PAGE 14 % (PM = marqueur de poids moléculaire, NI = extrait bactérien total avant induction, I1 = extrait bactérien total après 3 h d'induction à 37 °C) montre une pureté satisfaisante.

Cette étape permet d'éliminer des contaminants qui ne pourraient pas être éliminés par chromatographie d'exclusion stérique. Les fractions contenant la protéine ont été regroupées, et l'ensemble a été injecté sur une colonne d'exclusion de taille (Superdex S75) équilibrée dans le tampon A. La protéine est éluée après passage de 150 mL de tampon A. L'analyse par SDS- PAGE montre une pureté suffisante de la protéine en fin de purification (Figure 49). La purification de GST-ADAR1long est très satisfaisante puisqu'en suivant ce protocole, il est possible d'obtenir jusqu'à 10 mg de protéine pure par litre de culture.

3.1.1.2 Purification d'ADAR1long

La purification d'ADAR1long s'effectue également en trois étapes chromatographiques à partir de la construction ADAR1long-B : une chromatographie d'affinité suivie d'une coupure de l'étiquette His₆, une chromatographie d'affinité inverse et une chromatographie d'exclusion stérique.

La première étape de purification est identique à celle de GST-ADAR1long. La protéine ADAR1long est éluée à 225 mM d'Imidazole. Les fractions contenant ADAR1long ont été regroupées, puis l'étiquette a été coupée par la protéase TEV. Après la coupure de l'étiquette, l'échantillon a été réinjecté sur la résine de Nickel. Comme elle n'a plus d'étiquette, la protéine ne se fixe pas sur la résine, elle est ainsi séparée de l'étiquette. On récupère la fraction non fixée

sur la résine. Cette fraction est ensuite concentrée jusqu'à un volume d'environ 10 mL pour pouvoir être injectée sur la colonne d'exclusion stérique équilibrée (Superdex S75) dans le tampon A.

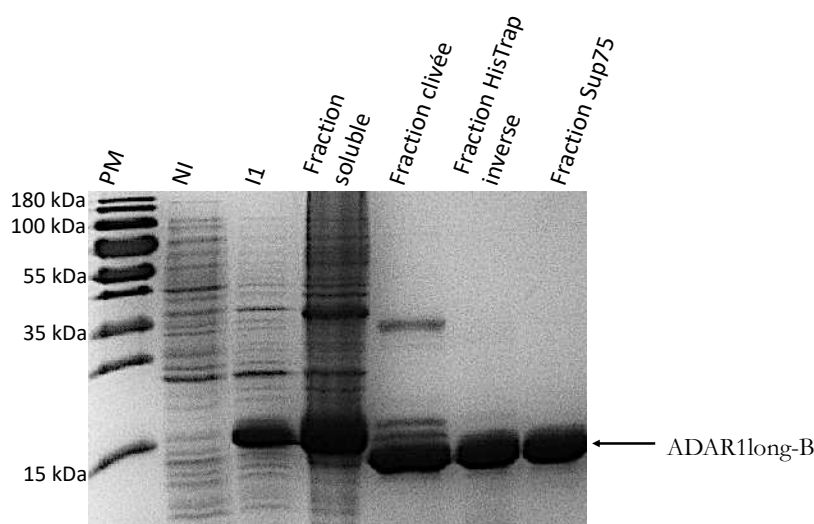


Figure 50 : Purification d'ADAR1long.

La purification d'ADAR1long est réalisée en trois étapes chromatographiques avec coupure de l'étiquette His₆. L'analyse de la purification par SDS-PAGE 16 % (PM = marqueur de poids moléculaire, NI = extrait bactérien total avant induction, I1 = extrait bactérien total après 20 h d'induction à 30°C) montre une pureté satisfaisante.

La protéine ADAR1long est éluée après passage de 145 mL de tampon A. Après la dernière étape de purification, la protéine présente une pureté suffisante pour les tests de cristallisation (Figure 50). En suivant ce protocole, il est possible d'obtenir environ 5 mg de protéine pure par litre de culture.

3.2 Étude de l'interaction Trn1ΔH8/ADAR1long par SEC

Cette partie décrit l'étude de la formation du complexe Trn1ΔH8/ADAR1long par chromatographie d'exclusion stérique avec une colonne Superdex (Increase S200). Les expériences ont été réalisées dans un premier temps dans le tampon A (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 20 % (v/v) glycérol, 1 mM TCEP), mais n'ont pas pu mettre en évidence l'interaction entre les deux partenaires. Après différentes optimisations, nous avons réalisés les expériences dans le tampon SEC (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 20 mM NaCl, 20 % (v/v) glycérol, 1 mM TCEP). Les protéines ont été purifiées comme décrit précédemment, puis ont été dialysées contre le tampon SEC. Trn1ΔH8 et ADAR1long sont respectivement concentrées à 50 μM et 750 μM, soit un ratio de 1:15.

En pr sence d'ADAR1long, nous observons un pic l g rement d cal  vers la gauche par rapport   celui de Trn1 H8. Ce pic pr sente un volume d' lution (13,18 mL) inf rieur   celui de Trn1 H8 seule (13,23 mL) ce qui signifie qu'un complexe s'est form  (Figure 51A). De plus, le complexe est observ  sur le gel SDS-PAGE (Figure 51B). Nous avons ainsi pu mettre en  vidence, pour la premi re fois une interaction entre la Transportine 1 et ADAR1 en solution.

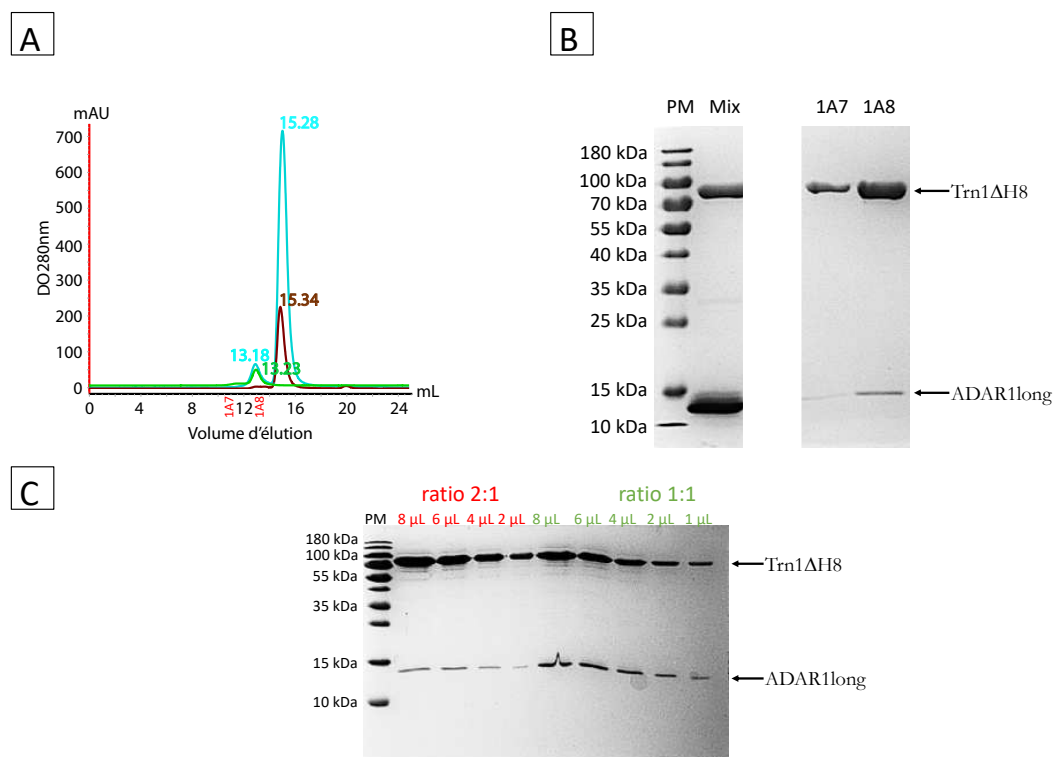


Figure 51 : Tests d'interaction par SEC des prot ines Trn1 H8 et ADAR1long.

(A) Profil d' lution de la chromatographie d'exclusion st rique (Trn1 H8 seule en vert, ADAR1long seule en marron, m lange Trn1 H8 + ADAR1long en cyan). Le chromatogramme repr sente l'absorbance UV   280 nm (mAU) en fonction du volume d' lution en mL. Les exp riences ont  t  men es avec ADAR1long en exc s   un ratio 1:15. (B) Analyse SDS-PAGE(14 %) de la r tention sur tamis mol culaire de l'exp rience r alis e en (A). (C) Estimation de la formation du complexe Trn1 H8/ADAR1long par analyse SDS-PAGE (14 %). Deux ratios Trn1 H8/ADAR1long (2:1 et 1:1) et diff rents volumes ont  t  d pos s sur gel afin d'estimer le complexe form  en (B).

Cependant, bien que nous ayons pu mettre en  vidence l'interaction entre les deux partenaires, nous n'avons pas pu former un complexe minimal compatible avec une  tude cristallographique. L' chantillon obtenu est suffisamment pur pour des tests de cristallog nese, mais nous ne sommes pas en pr sence d'un complexe complet (ratio 1:1). Toutes les prot ines Trn1 H8 ne sont en effet pas en interaction avec ADAR1long. Cette observation a  t  d montr e par une analyse SDS-PAGE qui a permis de montrer qu'au mieux, 50 % des prot ines Trn1 H8 interagissent avec ADAR1long (Figure 51C) dans les conditions o  ADAR1long  tait pourtant en large exc s (1:15).

3.3 Étude de l'interaction Trn1ΔH8/ADAR1long par SEC-MALS

Afin de mieux caractériser les conditions de formation du complexe Trn1ΔH8/ADAR1long en solution et sa stœchiométrie, nous avons réalisé des expériences de SEC-MALS. Les premières expériences portent d'abord sur les protéines seules (Trn1ΔH8 et ADAR1long), puis sur le complexe mis en évidence lors des expériences de SEC.

J'ai d'abord optimisé le tampon utilisé lors des expériences de SEC-MALS afin de favoriser la formation du complexe. Les interactions principalement attendues entre les deux partenaires étant des interactions électrostatiques, le tampon SEC qui contient des concentrations relativement importantes en sel ne favorise pas l'interaction. Les expériences ont donc été réalisées dans le tampon SEC-MALS (20 mM Hepes pH 7,3, 55 mM KOAc, 10 mM NaCl, 10 % (v/v) glycérol, 1 mM TCEP) dans lequel la quantité en sel a été diminuée. Les protéines ont été purifiées comme décrit précédemment, puis dialysées contre le tampon SEC-MALS. Les protéines concentrées à 40 μM ont été passées sur une colonne d'exclusion stérique Bio SEC-3 couplée au système MALS.

La protéine Trn1ΔH8 seule présente un pic majoritaire correspondant à une masse moléculaire mesurée de $74,9 \pm 0,9$ kDa qui est 20 % en dessous de la masse théorique attendue de 98,5 kDa. Cette différence est sûrement due aux conditions dans lesquelles sont réalisées les mesures, notamment la présence de glycérol dans le tampon. La protéine ADAR1long présente un pic correspondant à une masse moléculaire mesurée de $15,0 \pm 0,1$ kDa qui correspond à la masse théorique attendue de 14,5 kDa (Figure 52).

Après l'analyse des protéines seules, l'analyse du complexe Trn1ΔH8/ADAR1long a été réalisée avec un excès d'ADAR1long. L'analyse de Trn1ΔH8 en présence d'ADAR1long à un ratio 1:5, donne un profil similaire à celui observé avec Trn1ΔH8 seule. Une masse moléculaire mesurée de $84,5 \pm 0,9$ kDa est observée pour la forme majoritaire (Figure 53). Une légère augmentation de la masse moléculaire d'environ 9,6 kDa qui ne correspond pas à la masse totale d'ADAR1long est observée. Comme cela a été observé avec les tests d'interaction par SEC, le complexe se forme en solution mais uniquement à ~ 65 %.

Après les expériences réalisées avec les protéines Trn1ΔH8 et ADAR1long, nous avons ensuite voulu comparer ces résultats avec ceux du complexe Trn1ΔH8/ADAR1court. En effet, alors que nous avons pu observer la formation d'un complexe sous-stœchiométrique par SEC avec

ADAR1long, un tel complexe n'avait pas pu être observé avec ADAR1court (résultats non montrés).

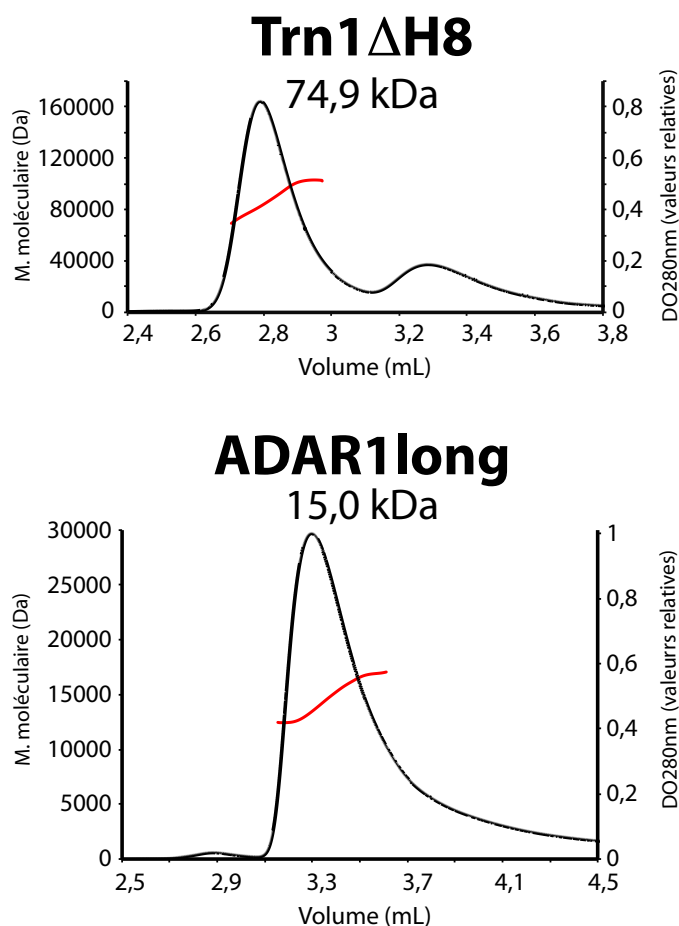


Figure 52 : Analyse par SEC-MALS de Trn1 Δ H8 et ADAR1long seules.

20 μ L d'échantillon contenant Trn1 Δ H8 et ADAR1long (40 μ M) ont été injectées sur la colonne équilibrée dans le tampon SEC-MALS. Le chromatogramme représente l'absorbance UV à 280 nm en noir (axe des ordonnées de droite, valeurs relatives) et la masse moléculaire en rouge (axes des ordonnées de gauche, Da) en fonction du volume d'élution en mL.

L'analyse par SEC-MALS de Trn1 Δ H8 en présence d'ADAR1court donne un profil très proche de celui observé avec Trn1 Δ H8 seule. Une masse moléculaire mesurée de $76,8 \pm 0,5$ kDa est observée (Figure 53). Nous n'observons aucune augmentation significative de la masse moléculaire (environ 2 kDa de différence). Les expériences de SEC-MALS montrent donc qu'il n'y a pas de formation du complexe Trn1 Δ H8/ADAR1court en solution dans les conditions utilisées. Ces résultats viennent confirmer ce qui avait été observé avec les expériences précédentes.

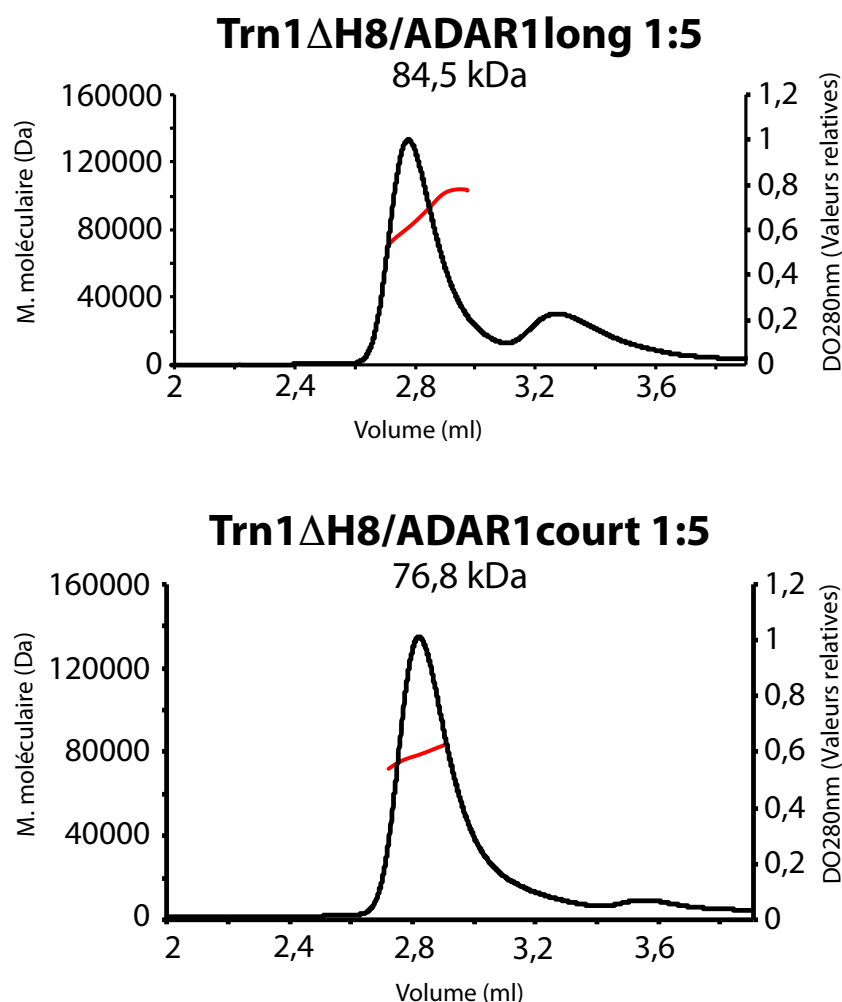


Figure 53 : Analyse par SEC-MALS de Trn1 Δ H8 en présence d'ADAR1long / ADAR1court.

20 μ L d'échantillon contenant Trn1 Δ H8 (40 μ M) et ADAR1long /ADAR1court (200 μ M) ont été injectés sur la colonne. Le chromatogramme représente l'absorbance UV à 280 nm en noir (axe des ordonnées de droite, valeurs relatives) et la masse moléculaire en rouge (axes des ordonnées de gauche, Da) en fonction du volume d'élution en mL. Ces expériences suggèrent une interaction entre Trn1 Δ H8 et ADAR1long, et ne montrent aucune interaction entre Trn1 Δ H8 et ADAR1court.

A la suite de ces derniers résultats, nous avons une fois de plus optimisé les conditions de mélange des protéines. Pour cela, nous avons amélioré le tampon dans lequel les expériences ont été réalisées. Le tampon SEC-MALS a alors été remplacé par le tampon SAXS (20 mM Hepes pH 7,3, 55 mM KOAc, 10 mM NaCl, 1 mM TCEP), les expériences ont donc été réalisées en absence de glycérol. Les protéines ont été purifiées comme précédemment puis dialysées contre le tampon SAXS.

Contrairement à ce qui a pu être observé en présence de glycérol, Trn1 Δ H8 présente un pic correspondant à une masse moléculaire mesurée de $98,3 \pm 0,7$ kDa. Cette dernière correspond bien à la masse théorique attendue de 98,5 kDa (Figure 54). De plus, l'échantillon protéique Trn1 Δ H8 est bien plus homogène en absence qu'en présence de glycérol.

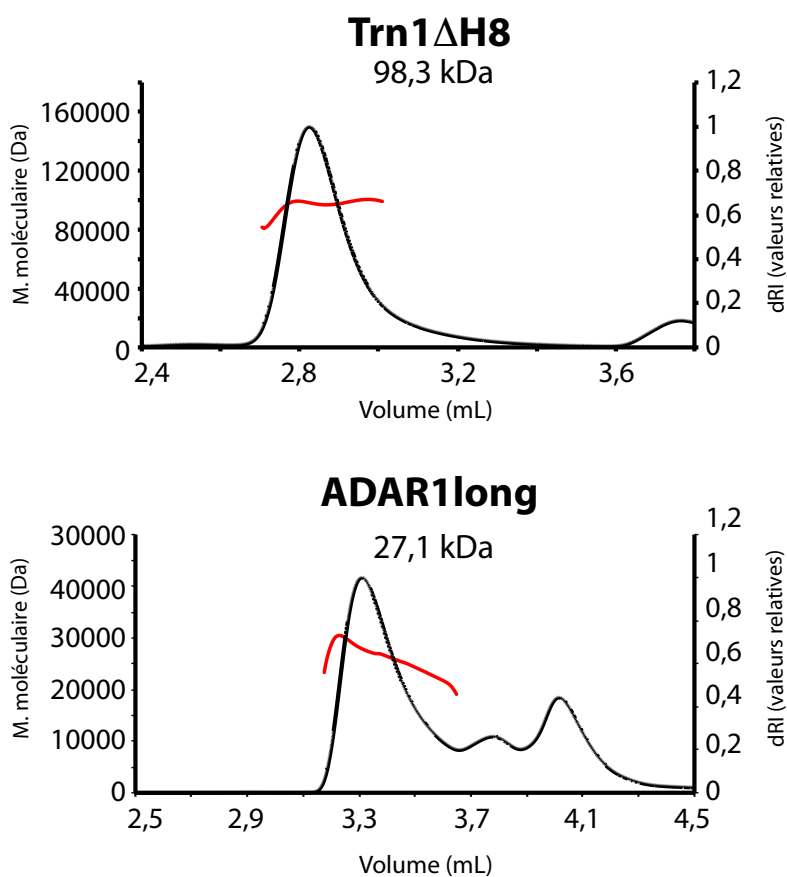


Figure 54 : Analyse par SEC-MALS de Trn1 Δ H8 et ADAR1long seules après optimisation.

20 μ L d'échantillon contenant Trn1 Δ H8 ou ADAR1long à 40 μ M ont été injectés sur la colonne équilibrée dans le tampon SAXS. Le chromatogramme représente la variation de l'indice de réfraction en noir (axe des ordonnées de droite, valeurs relatives) et la masse moléculaire en rouge (axes des ordonnées de gauche, Da) en fonction du volume d'élution en mL. La masse moléculaire de Trn1 Δ H8 calculée est identique à la masse théorique attendue. La masse moléculaire calculée d'ADAR1long correspond à quasiment le double de la masse théorique attendue.

Dans le cas d'ADAR1long, en absence de glycérol, la protéine présente un pic majoritaire correspondant à une masse moléculaire mesurée de $27,1 \pm 0,9$ kDa qui correspond à environ le double de la masse moléculaire théorique. Contrairement à ce qui a été observé avec le glycérol, ces résultats suggèrent qu'ADAR1long est sous forme dimérique en solution. Il pourrait s'agir d'interactions non spécifiques entre les extrémités flexibles, ou d'interactions spécifiques au niveau du domaine dsRBD3, certains dsRBD pouvant en effet former des dimères asymétriques (202) (Figure 54).

L'analyse de Trn1 Δ H8 en présence d'ADAR1long à un ratio 1:10 et en absence de glycérol donne un meilleur profil comparé à celui obtenu en présence de glycérol (résultats non

montr s). La masse mol culaire mesur e est de $110,8 \pm 0,9$ kDa pour le pic majoritaire qui correspond aux deux prot ines complex es (Figure 55), ADAR1long interagissant sous forme monom rique. La diff rence de masse entre Trn1 H8 en pr sence et en absence d'ADAR1long est de 12,5 kDa, ce qui est plus proche de la masse th orique d'ADAR1long (14,5 kDa). En r sum , nous observons bien la formation d'un complexe Trn1 H8/ADAR1long en SEC-MALS mais ne formons qu'environ 85 % du complexe malgr  le large exc s d'ADAR1long ajout . Nous sommes en pr sence d'un  quilibre rapide entre les formes libre et li e.

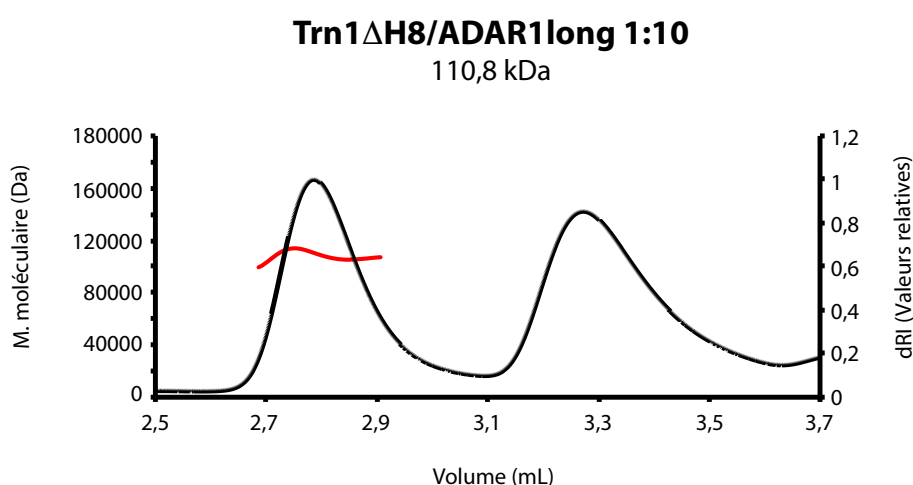


Figure 55 : Analyse par SEC-MALS de Trn1 H8 en pr sence d'ADAR1long   un ratio 1:10.

20 μ L d' chantillon contenant les prot ines Trn1 H8 (40 μ M) et ADAR1long (400 μ M) ont  t  inject s sur la colonne. La formation d'un complexe 100 % n'est toujours pas mise en  vidence avec les  p riences de SEC-MALS.

Les diff rentes  p riences de SEC-MALS ont  t  r alis es dans le but d'optimiser les conditions de formation des complexes (constructions, tampons, ratios). La baisse de la concentration en sel dans le tampon des  chantillons a permis d'am liorer l'interaction entre les deux partenaires. La suppression du gly rol dans le tampon a permis d'am liorer tr s significativement les profils des  chantillons prot iques. La quantit  de gly rol dans le tampon impactait les prot ines seules, notamment leur homog nit , mais  galement la formation du complexe Trn1 H8/ADAR1long. La diff rence de masse mol culaire du m lange Trn1 H8/ADAR1long obtenue est bien plus grande en absence qu'en pr sence de gly rol, se rapprochant ainsi de la formation d'un complexe 100 %.

Protéines	Masse théorique (kDa)	Ratio	Masse moléculaire (kDa)	Différence (kDa)	Glycérol
Trn1ΔH8	98,5	–	74,9 ± 0,9	–	+
ADAR1long	14,5	–	15,0 ± 0,1	–	+
ADAR1court	10,8	–	9,93 ± 0,5	–	+
Trn1ΔH8 / ADAR1long	113,0	1:5	84,5 ± 0,9	+ 9,6	+
Trn1ΔH8 / ADAR1court	109,3	1:5	76,8 ± 0,5	+ 1,9	+
Trn1ΔH8 / ADAR1long	113,0	1:10	86,0 ± 0,9	+ 11,2	+
Trn1ΔH8	98,5	–	98,3 ± 0,7	–	–
ADAR1long	14,5	–	27,1 ± 0,9	–	–
Trn1ΔH8 / ADAR1long	113,0	1:10	110,8 ± 0,9	+12,5	–

Tableau 15 : Synthèse des expériences SEC-MALS de Trn1ΔH8.

Tableau récapitulant l'étude de l'interaction entre Trn1ΔH8 et ADAR1long (comparé à ADAR1court) par des expériences de SEC-MALS.

En conclusion, en absence de glycérol, nous avons pu mettre en évidence la formation quasi complète du complexe Trn1ΔH8/ADAR1long lorsque la protéine ADAR1long est présente 10 fois en excès. Ces conditions demandent une grande quantité de protéine ADAR1long à de grandes concentrations. Le mélange des deux protéines à ces ratios n'est pas compatible avec une étude cristallographique. Les expériences de SEC-MALS montrent aussi que Trn1ΔH8 interagit avec ADAR1long mais pas avec ADAR1court. Par ailleurs, de précédentes études ont montré *in vivo* le transport d'ADAR1court par Trn1 (102, 126, 174). Toutefois dans ces travaux, il ne s'agissait pas de Trn1ΔH8 mais de Trn1FL. Trn1ΔH8 pourrait ainsi ne pas être adaptée à l'étude structurale du complexe entre Trn1 et ADAR1.

4 Caractérisation du complexe Trn1FL/ADAR1long

Les expériences précédentes avaient pour but de caractériser l'interaction entre la Transportine 1 et ADAR1 à l'échelle moléculaire, notamment par cristallographie. Pour ce faire, nous avons travaillé avec Trn1ΔH8 qui est largement utilisée pour les expériences de cristallisation du fait de la délétion de la boucle H8, longue et flexible. Mais étant dans l'incapacité d'obtenir un complexe stable formé complètement et donc une structure cristallographique du complexe Trn1ΔH8/ADAR1, nous avons décidé d'explorer cette interaction en utilisant Trn1FL.

4.1 Purification de Trn1FL

Le protocole de purification de Trn1FL consiste en trois étapes de purification : une colonne d'affinité glutathion sepharose, une colonne échangeuse d'anions et une colonne d'exclusion stérique.

Après la première étape, identique à celle de Trn1ΔH8, les fractions contenant Trn1FL ont été regroupées et l'étiquette GST a été coupée par la protéase HRV 3C. Après coupure de l'étiquette, une chromatographie échangeuse d'anions a été réalisée. Trn1FL est éluée à 175 mM de NaCl. Les fractions contenant Trn1FL ont été regroupées et, un maximum de 10 mL d'échantillon a été injecté sur une colonne d'exclusion stérique (Superdex S200) équilibrée dans le tampon de transport (20 mM Hepes pH 7,3, 110 mM KOAc, 2 mM MgOAc, 1 mM TCEP). La protéine est éluée après passage de 180 mL de tampon de transport. Puis, les fractions ont été regroupées et concentrées à 10 mg/mL. L'analyse de la purification par SDS- PAGE indique qu'après la dernière étape de purification, Trn1FL est suffisamment pure pour les tests de cristallisation (Figure 56).

Pour augmenter le rendement de la purification, la fraction non retenue par la colonne glutathion a été réinjectée sur la colonne pour permettre une plus grande fixation de la protéine (comme pour Trn1ΔH8). Afin de limiter les pertes dues à la précipitation, entre chaque étape de purification, une fois les fractions rassemblées, la concentration en TCEP dans l'échantillon a été ramenée à 2 mM de TCEP afin de limiter l'oxydation de la protéine. Avec toutes ces précautions, j'ai alors pu obtenir jusqu'à 3 mg de Trn1FL pure par litre de culture.

Contrairement à Trn1ΔH8, le tampon final de Trn1FL ne contient pas de glycérol car Trn1FL est soluble et stable sans glycérol et n'a pas besoin d'être autant concentrée que Trn1ΔH8 puisque les expériences de cristallisation concernant Trn1FL sont effectuées à plus faible concentration (5 mg/mL pour Trn1FL sans glycérol contre 15 mg/mL pour Trn1ΔH8 avec 20 % (v/v) glycérol).

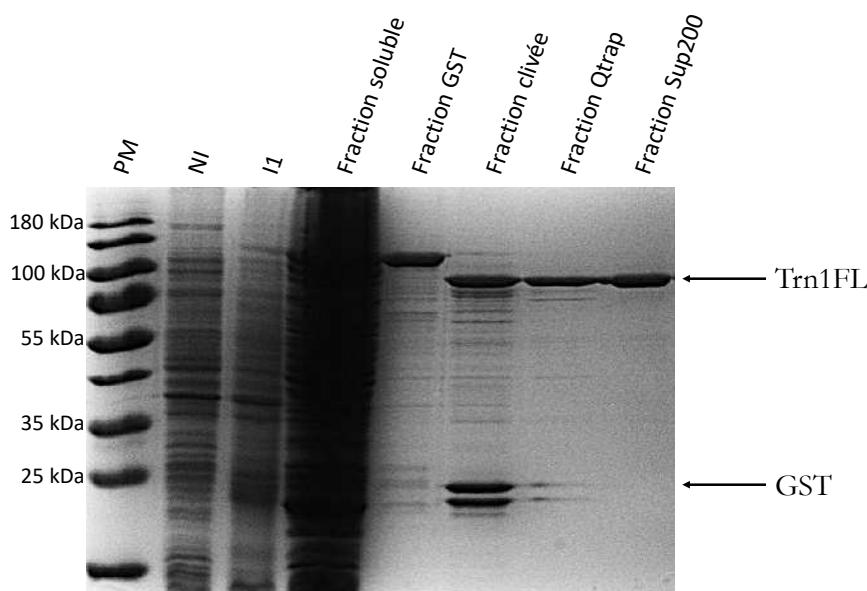


Figure 56 : Purification de Trn1FL.

La purification de Trn1FL se fait en trois  tapes chromatographiques avec une coupure de l' tiquette GST. L'analyse de la purification par SDS-PAGE 10 % (PM = marqueur de poids mol culaire, NI = extrait bact rien total avant induction, I1 = extrait bact rien total apr s 20 h d'induction   20 C) montre une puret  satisfaisante.

4.2 Caract risation physico-chimique du complexe Trn1FL/ADAR1long

Une fois la prot ine Trn1FL purifi e, nous avons dans un premier temps caract ris  l'interaction entre Trn1FL et ADAR1long, pour d terminer les conditions de m lange du complexe   utiliser lors des exp riences de cristallisation et de SAXS.

4.2.1  tude de l'interaction Trn1FL/ADAR1long par SEC-MALS

Les exp riences de SEC-MALS ont  t  r alis es avec Trn1FL (50  M) en absence et en pr sence d'ADAR1long (500  M). Ces exp riences ont  t  effectu es dans le tampon SAXS qui ne contient pas de gly cerol.

La prot ine Trn1FL seule pr sente un pic correspondant   une masse mol culaire mesur e de $101,6 \pm 0,6$ kDa qui correspond   la masse th orique attendue de 101,7 kDa (Figure 57). De plus, ces r sultats r v lent que l' chantillon prot ique Trn1FL pr sente une homog nit  compatible avec une  tude cristallographique.

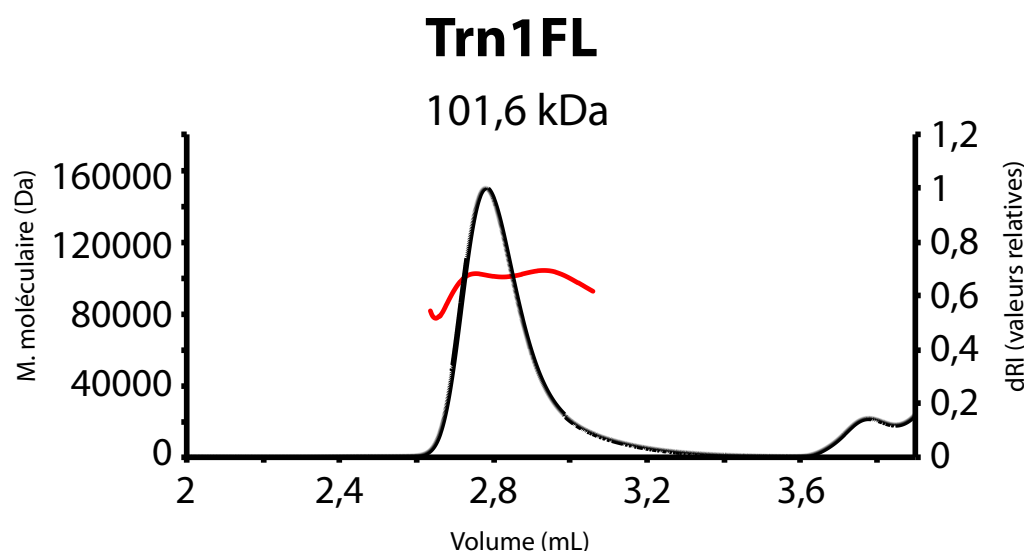


Figure 57 : Analyse par SEC-MALS de Trn1FL seule.

20 μ L d'échantillon contenant Trn1FL (40 μ M) ont été injectés sur la colonne. Le chromatogramme représente la variation d'indice de réfraction en noir (axe des ordonnées de droite, valeurs relatives) et la masse moléculaire en rouge (axes des ordonnées de gauche, Da) en fonction du volume d'élution en mL. La masse moléculaire de Trn1FL calculée est identique à la masse théorique attendue.

Nous avons ensuite réalisé l'analyse SEC-MALS du mélange Trn1FL/ADAR1long en testant quatre ratios allant de 1:1 à 1:15. L'analyse de Trn1FL (40 μ M) en présence d'ADAR1long à un ratio 1:1 donne un profil similaire à celui observé avec la protéine Trn1FL seule (Figure 58), mais avec une masse moléculaire mesurée de $117,0 \pm 0,8$ kDa (Tableau 16), soit une augmentation de la masse moléculaire d'environ 15,4 kDa par rapport à la masse moléculaire calculée de Trn1FL seule. Cette différence de masse correspond à la masse théorique d'ADAR1long (14,5 kDa). La différence de masse entre Trn1FL seule et Trn1FL en présence d'ADAR1long à un ratio 1:1 montre la formation d'un complexe entre les deux protéines.

Avec l'augmentation de la quantité d'ADAR1long ajoutée en solution, le décalage de la courbe dRI (variation d'indice de réfraction) est légèrement plus important (Figure 58). L'analyse de Trn1FL avec ADAR1long à un ratio 1:5 montre la formation du complexe entre les deux partenaires en solution avec une masse moléculaire mesurée de $118,1 \pm 0,9$ kDa (Tableau 16) soit une augmentation de +16,5 kDa à un ratio 1:5 contre une augmentation de +15,4 kDa à un ratio 1:1. Lorsque l'on augmente encore la quantité d'ADAR1long, la masse apparente du complexe continue d'augmenter légèrement. En effet, une différence de +18,3 kDa est observée lors de l'analyse de Trn1FL en présence d'ADAR1long (ratio 1:10) avec une masse moléculaire mesurée de $119,9 \pm 0,9$ kDa. Toutefois, au-delà d'un ratio 1:10, la masse apparente du complexe formé n'augmente plus (Tableau 16).

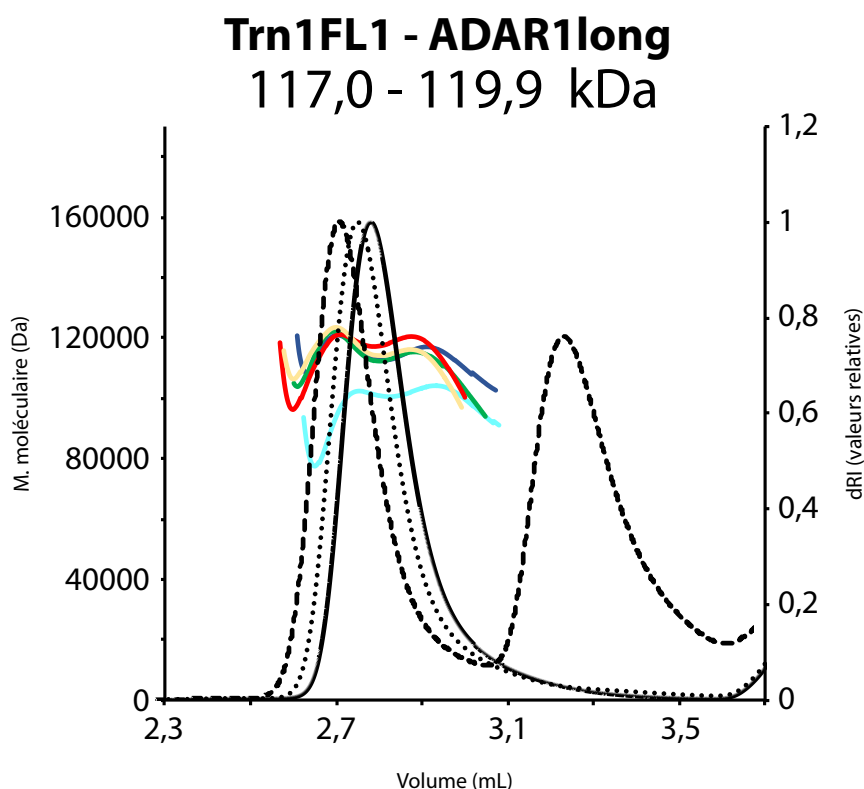


Figure 58 : Analyse par SEC-MALS de Trn1FL en présence d'ADAR1long.

20 μ L d'échantillon contenant Trn1FL (40 μ M) et ADAR1long (40-600 μ M) ont été injectés sur la colonne. Les chromatogrammes représentent la variation d'indice de réfraction (Trn1FL seule en trait plein, ratio 1:1 en point rond, ratio 1:10 en tiret) en noir (axe des ordonnées de droite, valeurs relatives) et la masse moléculaire (Trn1FL seule en cyan, ratio 1:1 en bleu, ratio 1:5 en vert, ratio 1:10 en rouge, ratio 1:15 en jaune) en Da (axes des ordonnées de gauche) en fonction du volume d'élution en mL. La présence d'ADAR1long entraîne un décalage de la courbe dRI vers la gauche indiquant la formation d'un complexe entre Trn1FL et ADAR1long. Ce décalage est observé dès l'ajout d'ADAR1long à un ratio 1:1, plus la quantité d'ADAR1long est augmentée plus ce décalage est important.

Protéines	Masse théorique (kDa)	Ratio	Masse moléculaire (kDa)	Différence (kDa)
Trn1FL	101,7	–	101,6 $\pm$ 0,6	–
ADAR1long	14,5	–	27,1 $\pm$ 0,9	–
Trn1FL / ADAR1long	116,2	1:1	117,0 $\pm$ 0,9	+ 15,4
Trn1FL / ADAR1long	116,2	1:5	118,1 $\pm$ 0,9	+ 16,5
Trn1FL / ADAR1long	116,2	1:10	119,9 $\pm$ 0,9	+ 18,3
Trn1FL / ADAR1long	116,2	1:15	119,9 $\pm$ 0,9	+ 18,3

Tableau 16 : Synthèse des expériences SEC-MALS de Trn1FL.

Tableau récapitulant les différentes expériences de SEC-MALS réalisées pour la mise en évidence de la formation du complexe Trn1FL/ADAR1long en solution. La formation du complexe est mise en évidence à partir d'un ratio 1:1.

En résumé, les différentes expériences de SEC-MALS entre Trn1FL et ADAR1long ont permis de mettre clairement en évidence la formation du complexe entre ces deux partenaires dans le tampon SAXS. A partir d'un ratio 1:1, l'interaction entre les deux protéines est clairement observable. Ces résultats suggèrent ainsi que l'affinité entre ADAR1long et Trn1FL est plus forte que celle entre Trn1ΔH8 et ADAR1long, puisque des ratios plus importants sont nécessaires pour observer la formation du complexe entre Trn1ΔH8 et ADAR1long dans les mêmes conditions. Ces résultats ont aussi permis de déterminer les conditions pour réaliser les expériences de SAXS qui seront réalisées à un ratio 1:10 dans le tampon SAXS. Avec ce ratio, on est certain de la présence en solution uniquement du complexe, ce qui est essentiel pour des expériences de SAXS, où l'entité étudiée doit être parfaitement homogène.

4.2.2 Étude structurale du complexe Trn1FL/ADAR1long en solution

Nous avons décidé de caractériser le complexe Trn1FL/ADAR1long en solution par SAXS et RMN afin d'utiliser l'ensemble de ces données pour caractériser à basse résolution l'interaction entre Trn1 et ADAR1. Des tests de cristallisation ont été menés en parallèle.

4.2.2.1 Étude du complexe Trn1FL/ADAR1long par SAXS

J'ai effectué une série de mesures en chromatographie d'exclusion stérique couplé au SAXS (SEC-SAXS) au synchrotron SOLEIL sur différents échantillons : ADAR1long seule, Trn1FL en absence et en présence d'ADAR1long. Pour chaque échantillon, la mesure des données a été réalisée sur le pic issu de la chromatographie d'exclusion stérique.

4.2.2.1.1 Étude des données de diffusion

Les courbes de diffusion des rayons X ont été traitées comme indiqué dans Matériels et Méthodes (paragraphe 2.2.3.1.1, page 100). Globalement, les courbes de diffusion des rayons X obtenues pendant l'élution permettent d'obtenir l'évolution tout au long de l'élution, du rayon de giration (R_g) et de l'intensité diffusée à l'origine $I(0)$ des échantillons. Les résultats obtenus pour Trn1FL ainsi que pour le complexe Trn1FL/ADAR1long sont représentés sur la Figure 59. Ces derniers confirment les résultats obtenus lors des expériences de SEC-MALS : ADAR1long forme un complexe stable avec Trn1FL en solution. Le complexe Trn1FL/ADAR1long est élué en un seul pic avec un rayon de giration stable le long du pic.

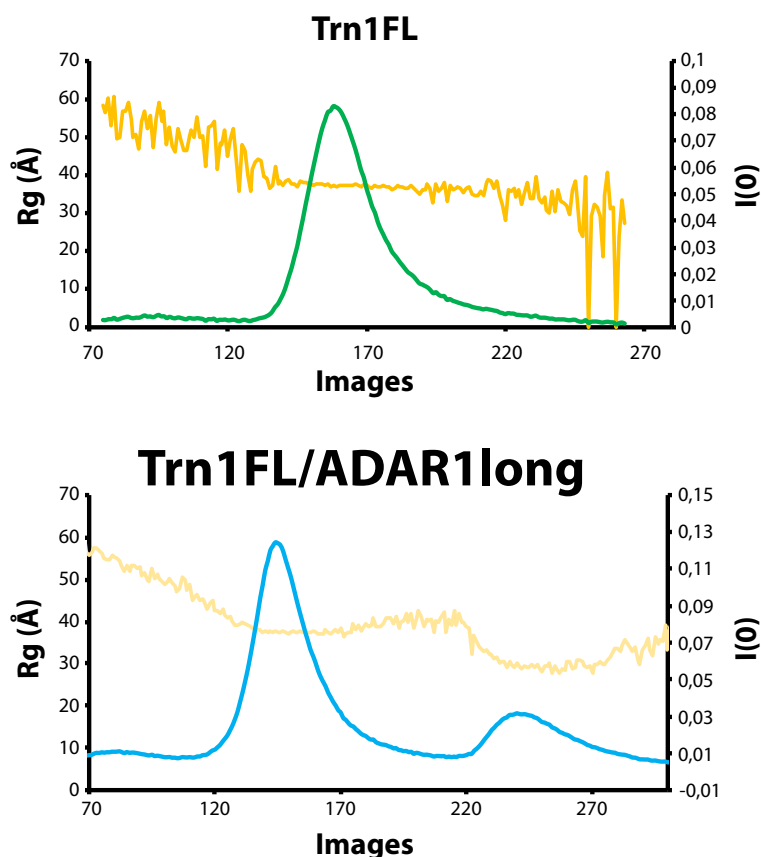


Figure 59 : Courbes des intensités diffusées et des rayons de giration de Trn1FL en absence et présence d'ADAR1long.

Courbe d'intensité de l'analyse en SAXS de Trn1FL en absence (vert) / présence (bleu) d'ADAR1long représentant l'intensité $I(0)$ diffusée en $q=0$ (en vert / en bleu) et le rayon de giration en Å (en orange / en jaune) en fonction de l'image.

La première étape du traitement des données consiste à soustraire aux données de Trn1FL ou du complexe Trn1FL/ADAR1long, l'intensité diffusée uniquement par le tampon avant l'élution de Trn1FL ou du complexe. La seconde étape du traitement des données consiste ensuite à moyenner les courbes identiques au voisinage du maximum de l'élution. Ainsi, neuf images correspondant au pic ont été moyennées pour chacun des échantillons (Figure 59). L'analyse avec le logiciel PRIMUS (191), de ces courbes moyennées représentatives de Trn1FL et de Trn1FL/ADAR1long a permis de déterminer le rayon de giration, le diamètre maximum de la molécule $D_{\max}$ et la masse moléculaire de Trn1FL et Trn1FL/ADAR1long. Concernant Trn1FL, les données obtenues ont été analysées avec le programme SHANUM (193, 194) et les données correspondant à un module du vecteur de diffusion q tel que $0,005 < q < 0,508$ ont été sélectionnées avec le programme DATCROP (193). Pour le complexe Trn1FL/ADAR1long, les données sélectionnées correspondent à q tel que $0,005 < q < 0,533$.

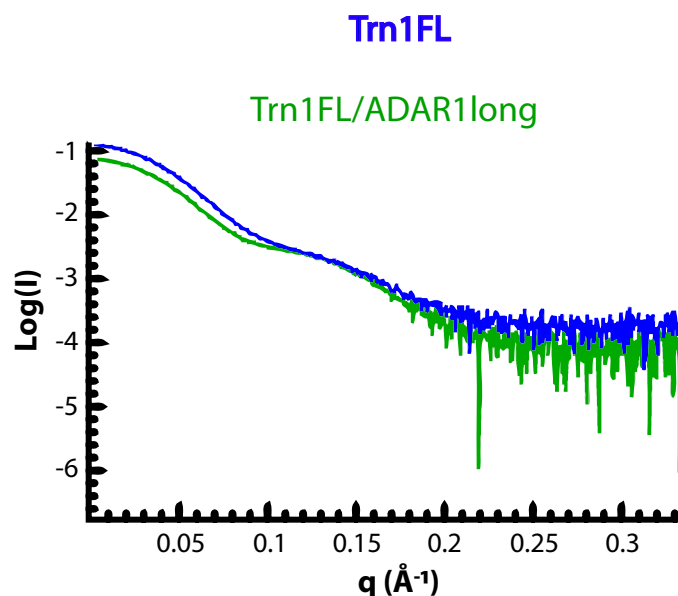


Figure 60 : Courbes expérimentales de Trn1FL en absence et en présence d'ADAR1long.

Les courbes de diffusion de Trn1FL en absence (en bleu) et en présence (en vert) d'ADAR1long sont représentées en fonction du module du vecteur de diffusion en $\text{\AA}^{-1}$.

La courbe expérimentale de Trn1FL est très proche de la courbe expérimentale du complexe Trn1FL/ADAR1long (Figure 60). La protéine Trn1FL possède un rayon de giration de $37,4 \pm 0,2 \text{ \AA}$ et le complexe Trn1FL/ADAR1long possède un rayon de giration de $37,5 \pm 0,2 \text{ \AA}$ (Figure 61).

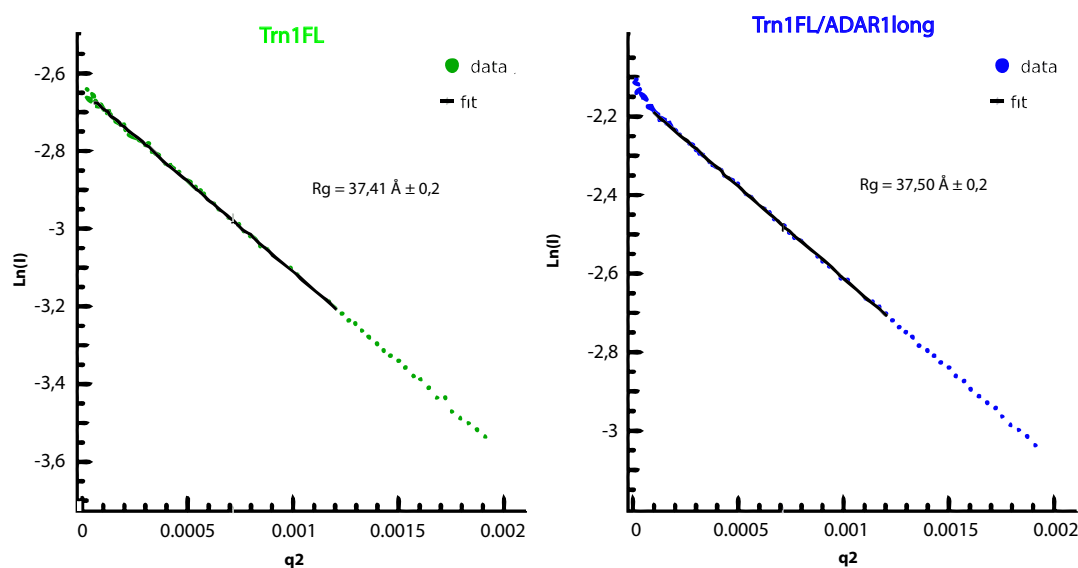


Figure 61 : Calculs des rayons de giration de Trn1FL en absence et en présence d'ADAR1long.

Le calcul des différents rayons de giration de Trn1FL en absence (en vert) ou en présence d'ADAR1long (en bleu) a été réalisé à partir de la région de Guinier des données expérimentales de SAXS. Le qR_g est compris entre 0,31 et 1,3 attestant de la fiabilité de la zone du Guinier.

La fonction paire de Trn1FL en complexe avec ADAR1long est relativement symétrique alors que celle de Trn1FL seule est clairement non symétrique et présente un léger épaulement avant le sommet de la fonction (Figure 62). La protéine Trn1FL montre une valeur de $D_{\max}$ de 120,7 Å tandis que la valeur de $D_{\max}$ pour le complexe Trn1FL/ADAR1long est de 125,7 Å (Tableau 17). Les résultats indiquent donc que Trn1FL seule et en complexe avec ADAR1long présentent des rayons de giration identiques, et des $D_{\max}$ très légèrement différents (Tableau 17). Ceci peut être expliqué par la différence de taille et de volume entre Trn1 et ADAR1long. En effet, avec plus de 101 kDa, Trn1 fait sept fois la taille d'ADAR1long (14,5 kDa). De plus, si l'on considère qu'ADAR1long peut se fixer dans la partie concave de l'arche C-terminale de Trn1, la fixation d'ADAR1long à Trn1 ne change pas de manière importante ces paramètres.

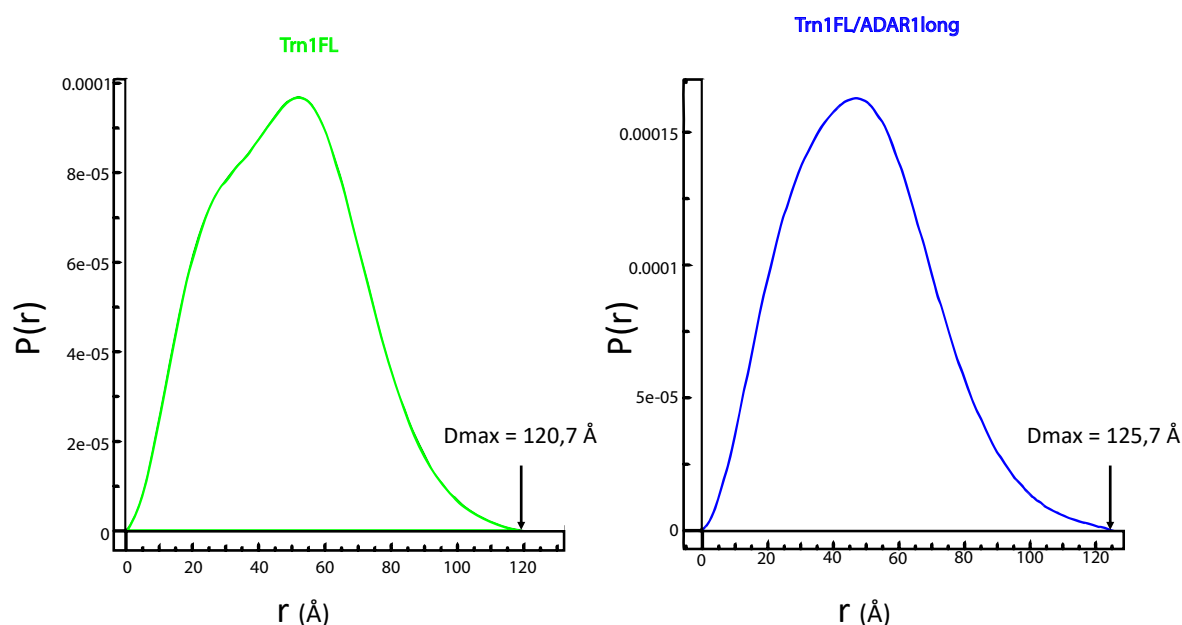


Figure 62 : Fonctions paires de Trn1FL en absence et en présence d'ADAR1long.

Les courbes de Trn1FL en absence (en vert) et en présence d'ADAR1long (en bleu) sont représentées sous la forme $P(r)$ en fonction de r (Å). La valeur de $D_{\max}$ est déterminée pour la fonction égale à zéro.

Par ailleurs, la différence de masse estimée à partir de ces données SAXS entre la protéine Trn1FL seule (99-116 kDa) et le complexe Trn1FL/ADAR1long (111-127 kDa) montre bien la fixation d'une molécule d'ADAR1long à Trn1FL (Tableau 17).

	Trn1FL	Trn1FL – ADAR1long
Collecte des données		
Images utilisées	158-167	139-148
Analyse Guinier		
Rg Guinier (Å)	37,4 ± 0,2	37,5 ± 0,2
qRg	0,31 – 1,30	0,39 – 1,30
Analyse P(r)		
Rg (Å)	37,4 ± 0,1	37,6 ± 0,1
Dmax (Å)	120,7	125,7
Détermination de la masse molaire		
MM _{sequence} (kDa)	101,7	116,2
MM _{correlation volume} (kDa)	109,1	118,7
MM _{intervalle de confiance} (kDa)	[99,2-116,0]	[111,2-127,4]
Probabilité intervalle de confiance(%)	98	92

Tableau 17 : Synthèse des expériences de Trn1FL en absence et en présence d'ADAR1long.

Tableau récapitulatif des résultats des analyses des courbes SAXS de Trn1FL en absence et en présence d'ADAR1long. Les valeurs entre crochet représentent l'intervalle de confiance de la masse moléculaire calculée à partir des données SAXS.

4.2.2.1.2 Modélisation

Nous avons ensuite essayé de déterminer un modèle de structure de Trn1FL en solution en absence et en présence d'ADAR1long à partir des données SAXS.

Concernant Trn1FL seule, nous avons utilisé notre structure cristallographique et également utilisé la déformation par modes normaux. Pour chaque modèle, une courbe théorique a été obtenue. Cette courbe a ensuite été comparée aux données expérimentales. Un test statistique appelée χ^2 (test de χ^2) nous a permis d'estimer la similarité entre la courbe expérimentale et une courbe théorique calculée à partir de chaque modèle avec le programme CRY SOL (203). Un χ^2 proche de 1, signifie que les écarts entre les données expérimentales et le modèle correspondent aux erreurs statistiques attendues. La différence entre la courbe de diffusion calculée à partir de la structure cristallographique de Trn1 Δ H8 et les données expérimentales ont montré que la structure en solution diffère substantiellement de la structure au sein du cristal ($\chi^2 = 34,9$) (Figure 63A).

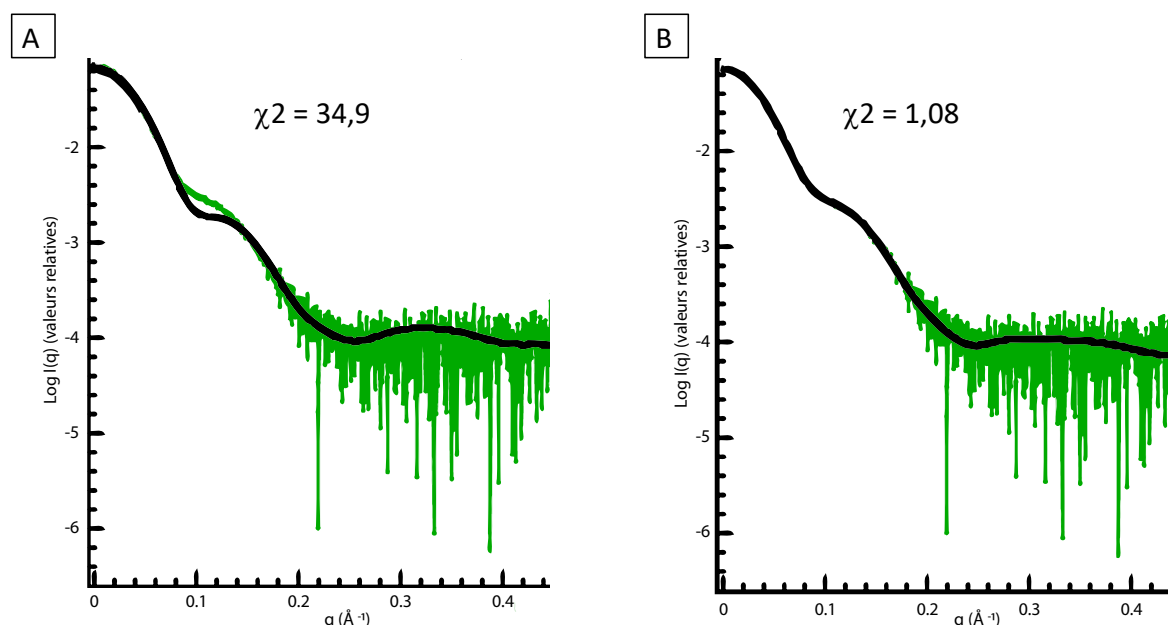


Figure 63 : Comparaison des données SAXS expérimentales de Trn1FL seule avec les courbes théoriques calculées.

Les données expérimentales sont en vert et le fit en noir. **(A)** Courbe de diffusion calculée à partir de la structure cristallographique de Trn1ΔH8. **(B)** Courbe de diffusion calculée à partir du modèle généré par modes normaux.

Cette différence peut être expliquée par la flexibilité structurale de Trn1. A cause de la flexibilité structurale intrinsèque des protéines à répétition HEAT formant des super-hélices, la structure de la super-hélice peut être plus ou moins allongée en fonction de la structure considérée (en solution ou au sein du cristal). En effet, la structure en solution étant soumise à moins de contrainte que la structure au sein du cristal, les deux structures ne présentent certainement pas le même niveau d'étirement, la structure en solution pouvant être plus allongée. A noter que pour le remplacement moléculaire, lors de la résolution de notre structure cristallographique, nous avons eu à prendre en compte cette flexibilité intrinsèque pour arriver à obtenir un bon modèle de remplacement moléculaire par déformation par modes normaux (paragraphe 1.3.2.2, page 126). Nous avons ainsi utilisé le programme SREFLEX pour la déformation par modes normaux (204). Un très bon accord entre les données expérimentales de SAXS et le modèle généré par le programme SREFLEX par déformation par modes normaux ($\chi^2 = 1,08$) a été obtenu (Figure 63B). L'hypothèse de similarité des données expérimentales et du modèle n'a pas pu être rejetée (test CorMap, 313 points, $C=11$, $p = 0,412$) (205), ce qui montre que le modèle obtenu est une bonne représentation de Trn1FL en solution (Figure 64A).

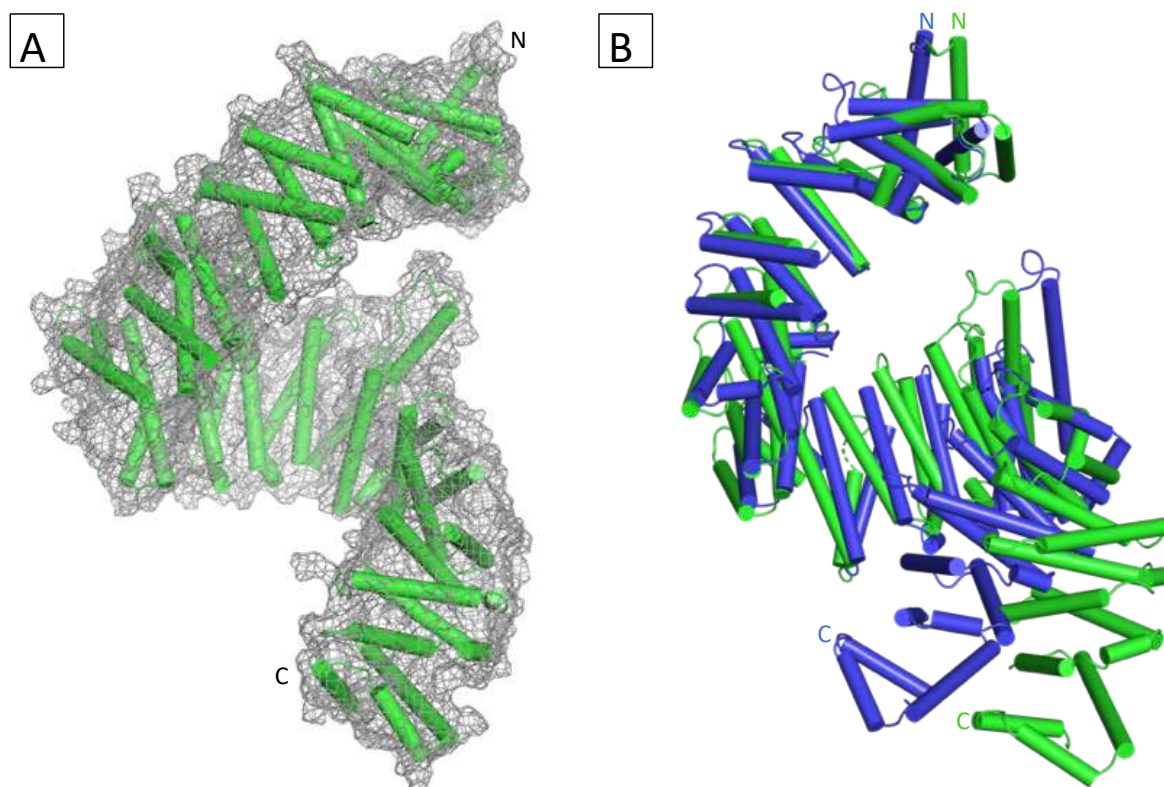


Figure 64 : Modèle de Trn1FL généré par déformation par modes normaux.

(A) Représentation du modèle de Trn1FL généré par déformation par modes normaux. (B) Comparaison du modèle de Trn1FL (en vert) généré et de la structure cristallographique de Trn1ΔH8 (en bleu).

Le modèle obtenu par déformation par modes normaux présente une structure plus allongée par rapport à celle résolue par cristallographie (Figure 64A). La comparaison des deux structures montre des différences plus importantes au niveau des extrémités N- et C-terminales qui présentent une flexibilité intrinsèque. De plus, l'arche C-terminale de la structure en solution est plus ouverte que celle de la structure cristallographique (Figure 64B).

Pour Trn1FL en complexe avec ADAR1long, nous avons utilisé notre structure cristallographique et le modèle précédemment proposé pour le complexe Trn1/dsRBD3 (174) (Figure 65). Nous avons également utilisé la déformation par modes normaux ainsi que la combinaison de différents programmes pour la construction de modèles de structure pour ce complexe.

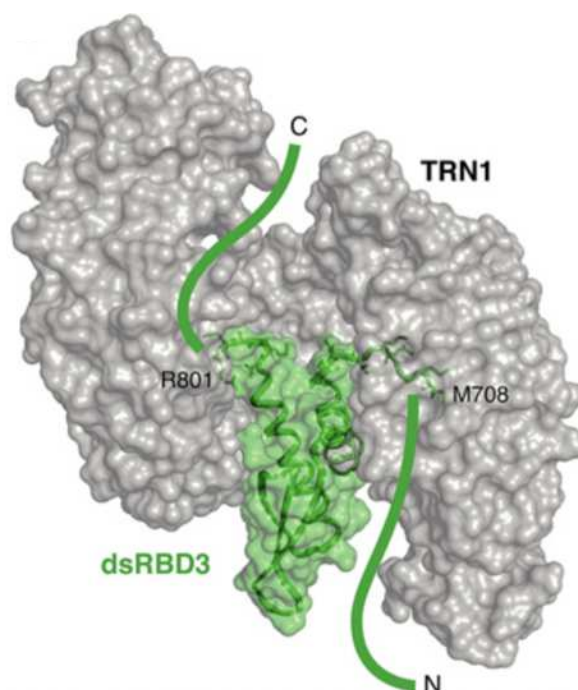


Figure 65 : Représentation en surface du modèle structural du complexe Trn1/dsRBD3 d'ADAR1. Le modèle de l'interaction entre Trn1 et le dsRBD3 d'ADAR1 (Trn1 est en gris, dsRBD3 est en vert) est proposé par Barraud *et al.* Dans ce modèle, le dsRBD3 est reconnu au niveau de l'arche C-terminale (HEAT H8-H20) de Trn1 où il occupe tout l'espace et ce sans aucune gêne stérique. D'après Barraud *et al* (174).

Un modèle obtenu par SREFLEX à partir de la structure cristallographique de Trn1 seule montre un très mauvais accord pour les données SAXS du complexe ($\chi^2 = 6,54$) (Figure 66A), confirmant qu'il s'agit bien de données d'un complexe et non de la protéine Trn1FL seule. Nous avons construit différents modèles avec une combinaison des programmes SASREF et SREFLEX (192). Les analyses ont été réalisées sans contraintes de distance entre Trn1 et ADAR1long. Une multitude de complexes différents ont été construits, donnant un accord plus ou moins bon avec les données SAXS. Seuls trois modèles représentatifs sont présentés (Figure 66 B-D). Pour les différents modèles du complexe Trn1FL/ADAR1long obtenus, la qualité d'ajustement du fit (test χ^2) montre qu'il y a un accord plutôt médiocre entre les modèles et les données expérimentales ($\chi^2 = 3,02$; 2,89 ; 1,67) (Figure 66B-D). Même pour le modèle avec le meilleur accord ($\chi^2 = 1,67$), l'hypothèse de similarité des données expérimentales et du modèle peut être rejetée (test CorMap). Ces résultats suggèrent que le modèle correspondant (Figure 66D), même s'il est le plus vraisemblable de tous les modèles obtenus (Figure 66B-C), n'est pas une représentation parfaite du complexe Trn1FL/ADAR1long en solution.

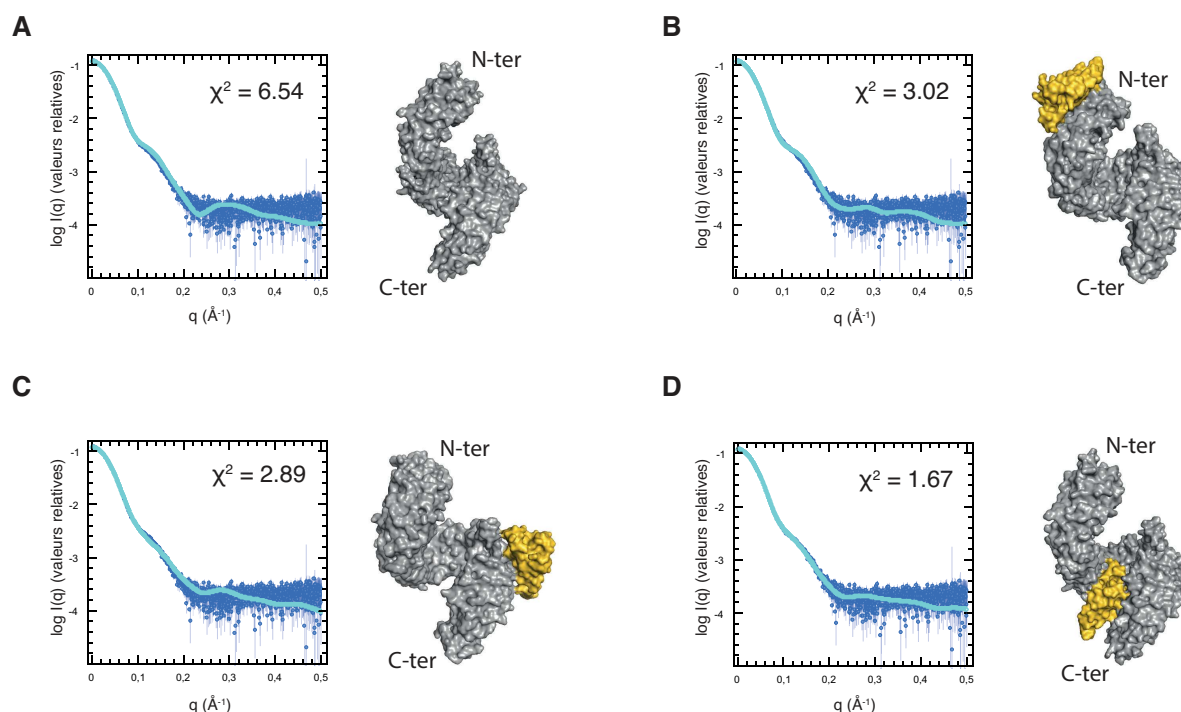


Figure 66 : Représentation des modèles du complexe Trn1FL/ADAR1long générés par déformation par modes normaux et par modélisation en corps rigide.

Les données expérimentales sont en bleu et le fit en cyan. (A) Modèle obtenu par déformations par modes normaux à partir du modèle généré de la structure cristallographique de Trn1ΔH8 avec le programme SREFLEX. (B) (C) (D) Les trois modèles représentatifs du complexe Trn1/ADAR1long obtenus par combinaison des programmes SASREF et SREFLEX.

Les expériences de SAXS avaient pour but de déterminer une structure à basse résolution permettant de situer localement le site de liaison d'ADAR1long sur Trn1FL. Ces expériences ont permis d'obtenir un bon modèle de Trn1FL en solution généré par déformation par modes normaux. Malheureusement, les données de SAXS ne nous ont pas permis de construire un meilleur modèle que celui présenté à la Figure 66D. Celui-ci a été construit par affinement en corps rigide (SASREF) et par déformation par modes normaux (SREFLEX) à partir du modèle initialement construit dans l'étude Barraud *et al* (Figure 65) sans données expérimentales, à partir uniquement de l'alignement de séquence converti en contraintes structurales (174). Les expériences de SAXS, effectuées sur Trn1FL en présence d'ADAR1long confirment, cette fois-ci avec des données expérimentales, que le modèle de l'interaction entre Trn1 et le dsRBD3 d'ADAR1 (Figure 65) est un modèle plausible de l'interaction entre les deux partenaires même si celui-ci ne rend pas parfaitement compte des données SAXS obtenues. La comparaison des différents modèles obtenus permet de valider la position d'ADAR1 au niveau de l'arche C-terminale de Trn1 (Figure 66B-D) puisqu'en comparant les différentes valeurs de χ^2 ($\chi^2 = 3,02 ; 2,89 ; 1,67$) le meilleur accord est obtenu pour une interaction à cette région.

Ces résultats donnent une validation partielle du modèle du complexe Trn1/dsRBD3 (Figure 65) qui doit être complété et affiné avec d'autres données expérimentales.

4.2.2.2 Étude du complexe Trn1FL/ADAR1long par RMN

Pour compléter les données obtenues par SAXS sur le complexe Trn1FL/ADARlong, des expériences RMN ont été réalisées par nos collaborateurs (Tobias Madl, Medical University of Graz) à partir des échantillons d'ADAR1 que nous leur avons fournis. Ces informations seront par la suite combinées aux données obtenues par SAXS pour améliorer le modèle du complexe Trn1FL/ADAR1long.

Une fois purifiée la protéine ADAR1 a été dialysée contre le tampon RMN (50 mM Tris-HCl pH 6,7, 150 mM NaCl, 2 mM TCEP). La protéine marquée Trn1FL^{15N-13C} (I, M méthyles) (Trn1FL doublement marquée uniformément en ¹⁵N, et en ¹³C au niveau des méthyles des méthionines et isoleucines) a été purifiée dans le tampon RMN. Les expériences de titration ont été réalisées avec des expériences RMN [(¹H, ¹⁵N)-HSQC, (¹H, ¹³C)-HMQC] dans lesquelles Trn1FL est titrée par ADAR1long. Ces expériences ont été comparées à des expériences de titration où Trn1FL est titrée par la protéine ADAR1ΔNLS qui correspond au domaine dsRBD3 d'ADAR1 sans les extrémités flexibles constituant le NLS d'ADAR1, et qui ne devrait donc pas ou peu interagir avec Trn1. Ces expériences ont également été comparées à des expériences de titration où Trn1FL est titrée par la protéine FUS qui possède un PY-NLS pour lequel la reconnaissance par Trn1 est bien caractérisée.

Les résultats obtenus par RMN sont présentés dans la Figure 67. Les deux protéines Trn1FL et ADAR1long interagissent lorsque cette dernière est mise en léger excès par rapport à Trn1FL (ratio 1:2). Ces expériences montrent que l'interaction de Trn1FL avec ADAR1long comprend des régions impliquées dans l'interaction avec des PY-NLS puisqu'en comparant les spectres de FUS et ADAR1long (Figure 67A-B), on remarque que des signaux impliqués dans la fixation du PY-NLS de FUS sont également impliqués dans la reconnaissance d'ADAR1long (signaux entourés en noir). Toutefois, en l'absence de l'attribution de ces méthyles, il n'est pas possible de savoir à quels résidus ils correspondent exactement.

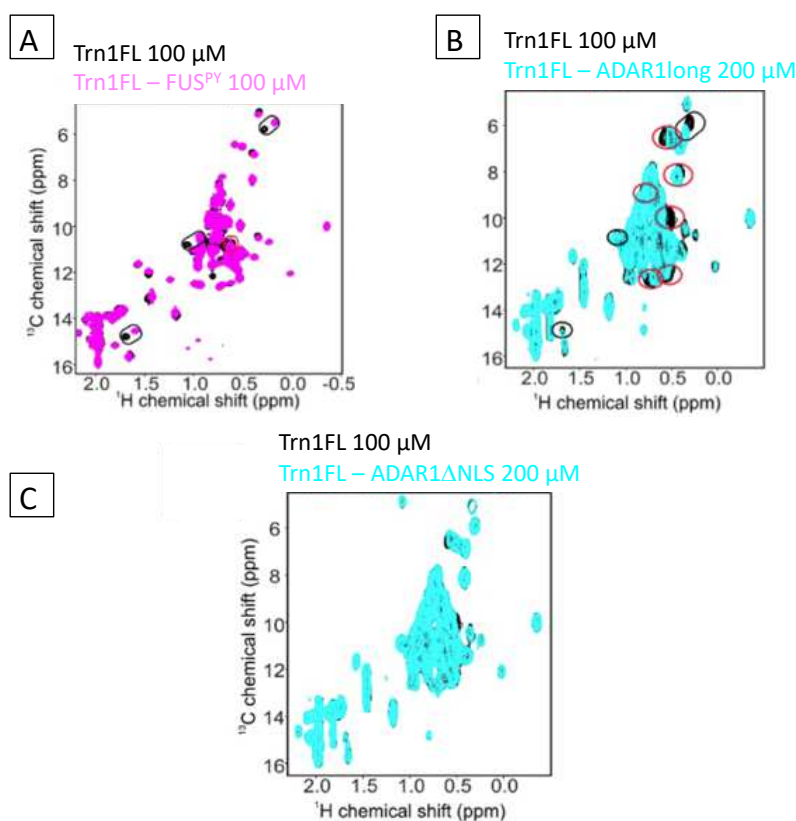


Figure 67 : Superposition des spectres HSQC ^1H - ^{15}N des expériences RMN de Trn1FL.

(A) Superposition de spectres de Trn1FL ^{15}N - ^{13}C (I, M méthyles) concentrée à 100 μM (spectre de référence en noir) en présence de 1 équivalent de la protéine FUS (en rose) réalisés à 900 MHz. (B) Superposition de spectres de Trn1FL ^{15}N - ^{13}C (I, M méthyles) concentrée à 100 μM (spectre de référence en noir) en présence de 2 équivalents de la protéine ADAR1long (en cyan) réalisés à 600 MHz. Les isoleucines de Trn1 impliquées spécifiquement dans la reconnaissance d'ADAR1 sont entourées en rouge. (C) Superposition de spectres de Trn1FL ^{15}N - ^{13}C (I, M méthyles) concentrée à 100 μM (spectre de référence en noir) en présence de 2 équivalents de la protéine ADAR1 Δ NLS (en cyan).

Ces expériences montrent aussi que la surface d'interaction de Trn1FL avec ADAR1long est plus étendue que la surface d'interaction avec les PY-NLS. En effet, plusieurs signaux d'isoleucine qui ne sont pas impliqués dans la reconnaissance des PY-NLS sont perturbés par l'ajout d'ADAR1long (signaux entourés en rouge sur la Figure 67B).

Par ailleurs, Trn1FL interagit très faiblement avec le domaine dsRBD3 d'ADAR1 sans le NLS (ADAR1 Δ NLS) (Figure 67C). Quelques isoleucines impliquées spécifiquement dans la reconnaissance d'ADAR1long (signaux entourés en rouge sur la Figure 67B) sont légèrement perturbés par l'ajout du dsRBD3 sans extension (ADAR1 Δ NLS, Figure 67C).

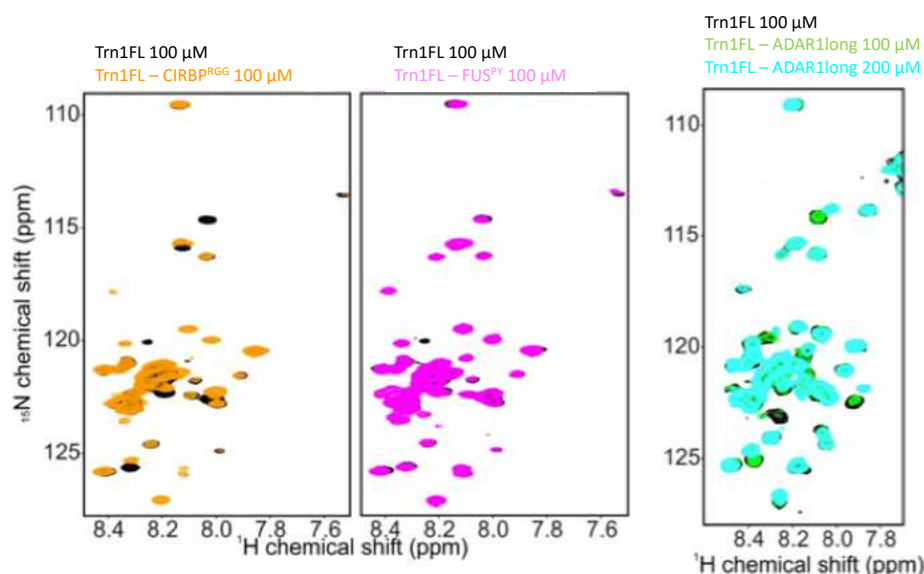


Figure 68 : Superposition de spectres HSQC ^1H - ^{15}N des expériences RMN de la boucle H8 de Trn1.

Superposition de spectres de la boucle H8 de Trn1FL ^{15}N concentrée à 100 μM (spectre de référence en noir) en présence d'un équivalent de la protéine CIRBP (en orange). Superposition de spectres de Trn1FL ^{15}N concentrée à 100 μM (spectre de référence en noir) en présence d'un équivalent de la protéine FUS (en rose). Superposition de spectres de la boucle H8 de Trn1FL ^{15}N concentrée à 100 μM (spectre de référence en noir) en présence d'un équivalent d'ADAR1long (en vert) ou de 2 équivalents d'ADAR1long (en cyan).

Les expériences RMN ont également permis d'observer l'implication de la boucle H8 de Trn1 dans la reconnaissance d'ADAR1 par Trn1. En effet, la superposition des spectres HSQC montrant l'interaction entre Trn1FL concentrée à 100 μM et ADAR1long concentrée à 100 μM ou 200 μM révèlent des variations de signaux de la boucle H8 de Trn1 (spectre de référence en noir) suite à l'ajout d'ADAR1long (en vert ou en cyan) (Figure 68). Ces mêmes variations sont également observées suite à l'ajout de la protéine CIRBP pour laquelle l'implication de la boucle H8 a été démontrée dans sa reconnaissance par Trn1 (96). Il faut noter que dans le cas des PY-NLS, la boucle H8 n'est pas impliquée dans la reconnaissance par Trn1 (Figure 68). L'absence de modifications des pics de la boucle H8 de Trn1 suite à l'ajout de la protéine FUS montre bien que la boucle H8 n'est pas impliquée dans la reconnaissance de ce PY-NLS.

En conclusion, les résultats des expériences RMN ont confirmé les résultats précédents concernant le complexe Trn1FL/ADAR1long. Les deux protéines interagissent en solution. Les résultats précédents suggéraient une meilleure affinité d'ADAR1long pour Trn1FL par rapport à Trn1 ΔH8 , ce qui est confirmé par l'étude RMN qui montre aussi de façon claire que la boucle H8 de Trn1 est impliquée directement dans la reconnaissance du NLS d'ADAR1.

4.2.3 Étude structurale du complexe Trn1FL/ADAR1long par cristallographie

En parallèle des expériences de SAXS et de RMN, nous avons poursuivi nos efforts de résolution de la structure cristallographique du complexe Trn1/ADAR1, en concentrant nos tests de cristallisation sur le complexe Trn1FL/ADAR1long.

4.2.3.1 Quantification de l'interaction Trn1FL/ADAR1long par ITC

Avant de se lancer dans la cristallisation, nous avons déterminé la constante d'association de l'interaction Trn1FL/ADAR1long par titrations calorimétriques isothermes (ITC). Les expériences ont été réalisées avec le complexe Trn1FL/ADAR1long et à titre comparatif, avec le complexe Trn1ΔH8/ADAR1long. Ces expériences d'ITC ont été réalisées avec la construction GST-His₆-ADAR1long (GST-ADAR1long) et GST-His₆-ADAR1court (GST-ADAR1court) dont l'étiquette n'est pas clivée.

Les protéines Trn1, Trn1FL ou Trn1ΔH8, ont été titrées par injection des différentes protéines d'ADAR1, GST-ADAR1long ou GST-ADAR1court, et la chaleur dégagée a été mesurée par le microcalorimètre (MicroCal iTC200). Ces titrations ont été réalisées dans le tampon MACARA (20 mM Hepes pH 7,9, 110 mM KOAc, 2 mM MgOAc, 0,5 mM EGTA, 20 mM NaCl, 1 mM PMSF, 1 mM TCEP, 0,1 % (v/v) Tween). Les aires des pics ont été calculées et la quantité de chaleur par mole d'ADAR1 injectée en fonction du ratio molaire a été obtenue. Ceci a permis de déterminer la constante d'association et le ratio molaire de la réaction.

Dans le cas de Trn1FL et GST-ADAR1long, la réaction d'association est exothermique, la constante de dissociation (K_D) est de 446 ± 88 nM et le ratio molaire est de $0,97 \pm 0,02$ (Figure 69A). Nous constatons que l'affinité entre Trn1FL et ADAR1long est assez forte avec un K_D de l'ordre de quelques centaines de nanomolaires. Un plateau est bien atteint pour cette réaction (Figure 69A) en fin de titration indiquant que le site de fixation de Trn1FL est saturé.

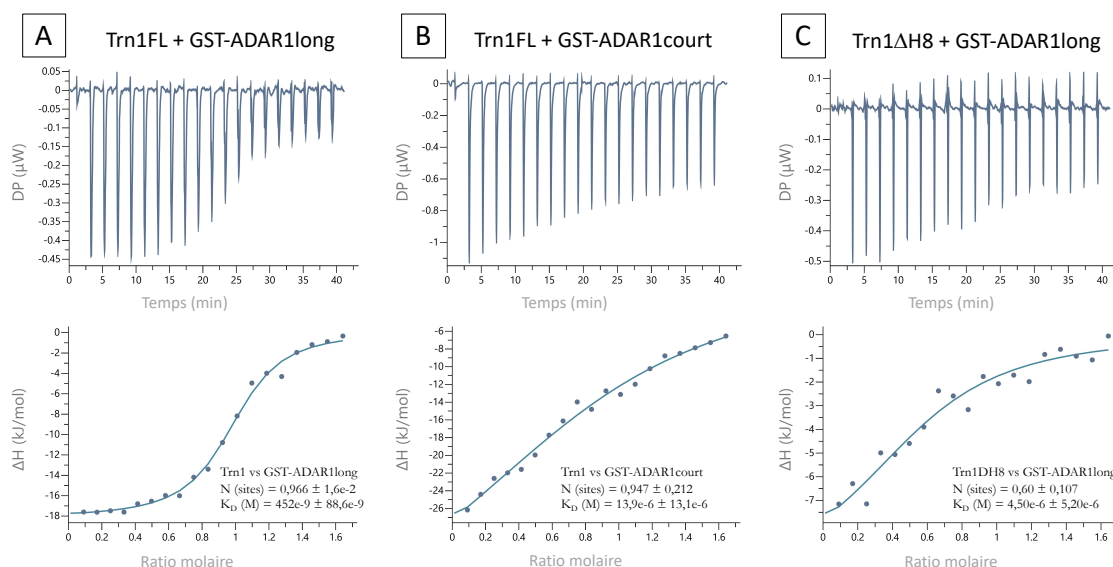


Figure 69 : Mesure ITC des complexes Trn1FL/Trn1 H8 / GST-ADAR1long/GST-ADAR1court. Titration calorim trique isotherme de Trn1FL titr e par (A) GST-ADAR1long et (B) GST- ADAR1court, et Trn1 H8 titr e par GST-ADAR1long. La quantit  de chaleur d gag e ( cal/sec) dans la cellule de mesure en fonction du temps (min) a  t  mesur e pour 20 injections successives de 4  L de GST-ADAR1long ou GST-ADAR1court   160  M dans la prot ine Trn1FL ou Trn1 H8   20  M. La variation de la quantit  de chaleur d gag e   chaque injection de GST- ADAR1long / GST-ADAR1court (kcal/mole) en fonction du ratio molaire Trn1FL / Trn1 H8 / GST-ADAR1long ou GST-ADAR1court permet de calculer le K_D .

Le K_D du complexe entre Trn1FL et GST-ADAR1court est de l'ordre d'une dizaine de micromolaires (Figure 69B), ceci confirmant que l'affinit  de Trn1 pour ADAR1long est plus forte que pour ADAR1court. L'affinit  entre Trn1 H8 et GST-ADAR1long est aussi plus faible que celle entre Trn1FL et GST-ADAR1long, avec un K_D de l'ordre de 7  M (Figure 69C).

Ces r sultats sont en accord avec les r sultats obtenus dans les paragraphes pr c dents, et confirment l'implication de la boucle H8 de Trn1 dans l'interaction avec ADAR1 ainsi que des extr mit s N- et C-terminales pr sentes dans le NLS d'ADAR1long. Il est donc n cessaire de travailler avec Trn1FL et ADAR1long pour mener une  tude structurale.

4.2.3.2 Utilisation de chim re pour la cristallisation du complexe Trn1FL/ADAR1long

Pour essayer d'obtenir un cristal du complexe Trn1FL/ADAR1long, j'ai commenc  par construire une prot ine chim re Trn1FL-ADAR1long pour la r solution de la structure cristallographique du complexe entre les deux partenaires. Cette chim re a  t  construite sur le

même principe que les précédentes, ADAR1long est greffée au niveau de l'extrémité C- terminale de Trn1FL *via* un petit espaceur (His-Met-Gly-Gly) qui permet de rapprocher les deux protéines.

4.2.3.2.1 Préparation et purification de la chimère Trn1FL-ADAR1long

Le protocole suivi pour l'expression de Trn1FL-ADAR1long est celui de la protéine Trn1FL. Ce protocole d'expression permet une surexpression suffisante de la chimère (Figure 70).

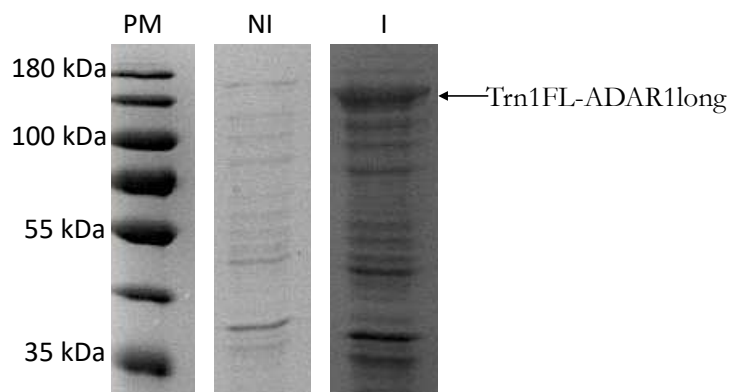


Figure 70 : Expression de la chimère Trn1FL-ADAR1long.

Le résultat obtenu pour l'expression de Trn1FL-ADAR1long est révélé par analyse SDS-PAGE 10 % (PM = marqueur de poids moléculaire, NI = l'extrait bactérien total avant l'induction, I = l'extrait bactérien total après 20 h d'induction à 20°C).

4.2.3.2.2 Purification de la chimère Trn1FL-ADAR1long

Le protocole de purification de la chimère Trn1FL-ADAR1long est le même que celui de la protéine Trn1FL qui consiste en trois étapes de purification : une colonne d'affinité glutathion sepharose, une colonne échangeuse d'anions et une colonne d'exclusion stérique.

Lors de la première étape de la purification, il est nécessaire de réinjecter la fraction soluble une deuxième fois pour pouvoir récupérer un maximum de matériel du fait que la colonne GST sepharose ne permet pas une fixation optimale de la protéine Trn1FL-ADAR1long. Après la coupure de l'étiquette GST par la protéase HRV 3C, la deuxième étape de purification permettant de séparer l'étiquette de la chimère à l'aide d'un gradient de NaCl a été réalisée. La chimère Trn1FL-ADAR1long est éluée distinctement de l'étiquette GST à 175 mM de NaCl. Une dernière étape de chromatographie d'exclusion stérique a été réalisée sur la colonne pré-équilibrée dans le tampon de transport. La chimère est éluée après passage de 150 mL de tampon de transport. Les différentes étapes de purification ont été analysées par SDS-PAGE qui montre qu'en fin de purification, la chimère Trn1FL-ADAR1long n'est pas aussi pure que

les chim res pr c dentes (Trn1 H8-ADAR1_pep ou Trn1 H8-A1_pep), mais suffisamment pure pour  tre compatible avec une  tude cristallographique (Figure 71). Le rendement de la purification pour la chim re Trn1FL-ADAR1long est d'environ 2 mg de prot ine pure par litre de culture bact rienne. Ce rendement est plus faible que celui obtenu pour les chim res pr c dentes, n anmoins, les diff rents tests n cessitent moins de mat riel pour la chim re Trn1FL-ADAR1long.

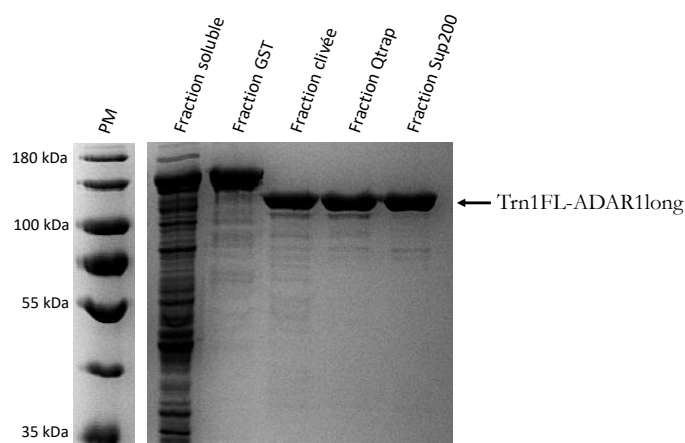


Figure 71 : Purification de la chim re Trn1FL-ADAR1long.

La purification de Trn1FL-ADAR1long est r alis e en trois  tapes chromatographiques avec une coupure de l' tiquette GST. L'analyse de la purification par SDS-PAGE 10 % (PM = marqueur de poids mol culaire) montre une puret  satisfaisante.

4.2.3.2.3 Cristallisation de la chim re Trn1FL-ADAR1long

Une fois la prot ine purifi e, nous avons ensuite entrepris des tests de cristallisation de la chim re Trn1FL-ADAR1long. La d marche est bas e sur l' tude des structures cristallographiques de Trn1FL dans la litt rature. Les premiers tests de cristallisation ont  t  effectu s avec Trn1FL-ADAR1long concentr e entre 5 et 7 mg/mL. Un premier criblage large   20 C a  t  r alis  en utilisant diff rents kits commerciaux.

Une premi re piste a  t  obtenue   partir de ce criblage large. Il s'agissait de sph rolites obtenus dans la condition 15 % (w/v) PEG 3350, 0,10 M Mg formate (Figure 72A) avec la chim re Trn1FL-ADAR1long concentr e   5 mg/mL. Ces sph rolites ont ensuite  t  optimis s en commen ant par affiner la condition dans laquelle ils  taient apparus. Cette premi re  tape d'optimisation a permis d'obtenir des microcristaux dans la condition 0,10 M Bis-Tris propane pH 7,50, 20,00 % (w/v) PEG 3350, 0,20 M Mg formate (Figure 72B). Puis,   partir de cette derni re condition, les microcristaux (Figure 72B) ont  t  pr lev s et utilis s pour la r alisation du microseeding. Cette  tape d'optimisation a permis d'obtenir encore une fois des

microcristaux plus gros (Figure 72). Finalement, une dernière étape d'optimisation a permis d'obtenir des cristaux en forme de plaque (Figure 72D), ces derniers ont été obtenus en effectuant une deuxième étape de microseeding à partir des microcristaux de l'étape précédente (Figure 72C).

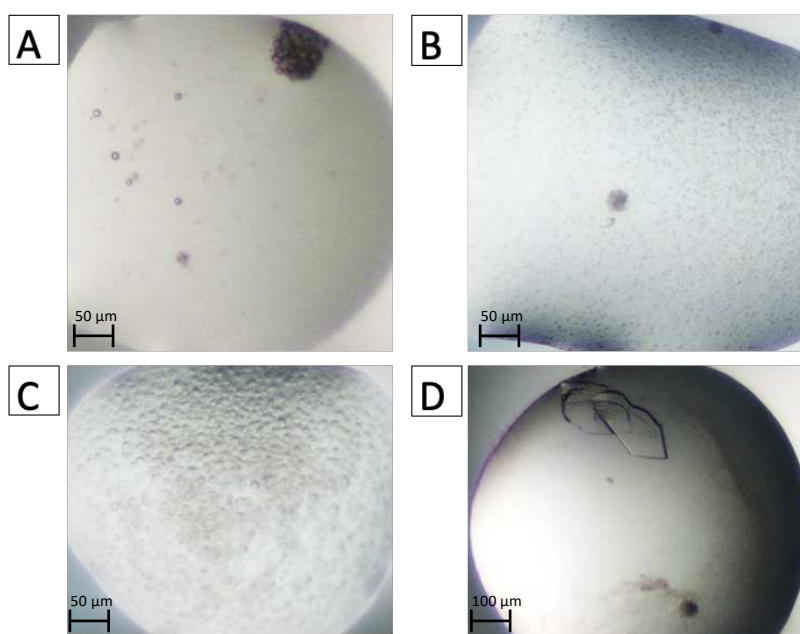


Figure 72 : Cristallisation de la chimère Trn1FL-ADAR1long.

(A) Les sphérolites obtenus dans la condition 15 % (w/v) PEG 3350, 0,10 M Mg formate. (B) Les microcristaux obtenus dans la condition 0,10 M Bis-Tris propane pH 7,50, 20 % (w/v) PEG 3350, 0,20 M Mg formate avec la première étape d'optimisation des cristaux obtenus en (A). (C) Les microcristaux de plus grande taille obtenus dans la condition 0,14 M K_2HPO_4 , 1,26 NaH_2PO_4 avec la deuxième étape d'optimisation des cristaux obtenus en (B). (D) Les cristaux en forme de plaque obtenus dans la condition 0,10 M Na Acetate pH 5,00, 20 % (w/v) PEG 6K, 0,20 M $CaCl_2$ à partir des cristaux obtenus en (C).

Les cristaux optimisés en forme de plaque (Figure 72D) ont été prélevés et cryoprotégés par incubation dans une goutte de solution de cristallisation contenant 20 à 25 % (v/v) de glycérol avant d'être congelés dans l'azote liquide. Malheureusement, il s'est avéré que c'étaient des cristaux de sels. En effet, les microcristaux obtenus après la deuxième étape d'optimisation avec le microseeding sont apparus dans une condition avec une forte concentration de phosphate (Figure 72C), des cristaux de phosphate de calcium qui sont probablement apparus lors de la dernière étape d'optimisation (Figure 72D).

Par manque de temps je n'ai pas pu poursuivre ces tests de cristallisation. Nous n'avons ainsi pas pu résoudre la structure cristallographique du complexe Trn1FL/ADAR1long avec l'utilisation de la chimère Trn1FL-ADAR1long.

4.2.3.3 Tests de co-cristallisation du complexe Trn1FL/ADAR1long

En parallèle, j'ai tenté de co-cristalliser le complexe Trn1FL/ADAR1long en formant le complexe à partir d'un mélange des partenaires purifiés. J'ai utilisé la construction Trn1FL et la construction GST-ADAR1long pour laquelle l'étiquette GST n'est pas clivée puisqu'elle présente la plus forte affinité pour Trn1FL parmi les différentes constructions testées (résultats non montrés). Les deux protéines ont été produites et purifiées comme décrit précédemment, seule la dernière étape de purification est réalisée dans le tampon MACARA utilisé pour les expériences ITC. Le tampon MACARA diffère du tampon de transport utilisé jusqu'à présent par son pH de 7,9 au lieu de 7,3. Ce tampon a été choisi pour la réalisation de ces expériences car, après optimisation, il s'est avéré que le pH du tampon était important pour la formation du complexe (résultats non montrés). A pH 7,9 la formation du complexe Trn1FL/ADAR1long est observée dès un ratio 2:1 contre un ratio 1:1 à pH 7,3.

4.2.3.3.1 Préparation du complexe Trn1FL/GST-ADAR1long

Les protéines Trn1FL et GST-ADAR1long ont été mélangées à 50 μ M, soit un ratio de 1:1, et le mélange a été injecté sur une colonne Superdex (Increase S200). En présence de GST-ADAR1long, nous observons un pic décalé vers la gauche par rapport à celui de Trn1FL seule. Ce pic présente un volume d'élution (9,01 mL) inférieur à celui de Trn1FL (10,30 mL) ce qui signifie qu'un complexe s'est formé (Figure 73A). De plus, le complexe est observable sur le gel SDS-PAGE (Figure 73B). Ce premier pic comprend le complexe Trn1FL/GST-ADAR1long avec une pureté compatible avec des études de cristallographie. Il est directement suivi d'un pic qui représente le complexe Trn1FL/GST-ADAR1long mais cette fois-ci contaminé par la GST qui s'est auto-coupée de la protéine.

La cristallisation du complexe Trn1FL/ADAR1long obtenu sur gel filtration a été menée sur un complexe concentré à 50 μ M avec le kit INDEX (Hampton Research). Des premières pistes de cristallisation ont été obtenues, mais l'optimisation n'a malheureusement pas pu être poursuivie par manque de temps.

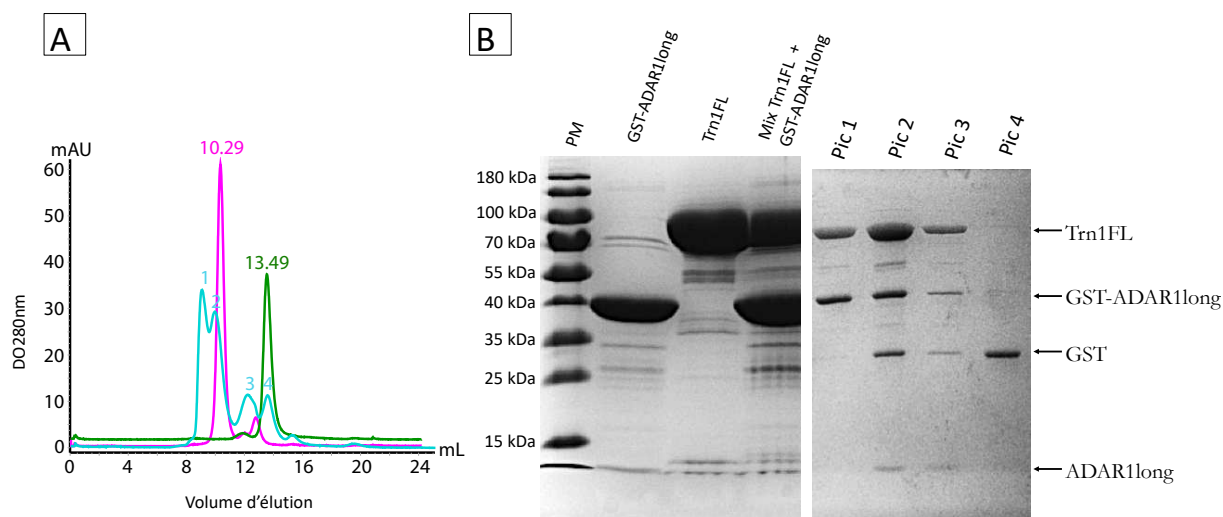


Figure 73 : Formation du complexe Trn1FL/GST-ADAR1long.

(A) Chromatogrammes d'exclusion stérique (Trn1FL seule en vert, GST-ADAR1long seule en rose, mélange Trn1FL + GST-ADAR1long en cyan). Les expériences ont été menées avec la construction GST-His₆-ADAR1long dont l'étiquette est conservée : GST-ADAR1long, en quantité équimolaire avec un ratio 1:1. (B) Analyse SDS-PAGE (12 %) de la rétention sur tamis moléculaire de l'expérience réalisée en (A).

5 Conclusion

En conclusion, dans cette première partie portant sur la reconnaissance d'ADAR1 par Trn1, l'objectif était de résoudre la structure cristallographique de Trn1 en complexe avec ADAR1. Cette structure devait apporter des informations expérimentales qui auraient pu être comparées aux données indirectes sur lesquelles sont basées la première hypothèse du modèle d'inactivation du NLS en présence d'ARN (la protéine ADAR1 comprend une surface d'interaction distincte pour Trn1 et l'ARNdb, et la protéine ADAR1 est reconnue au niveau de l'arche C-terminale de Trn1 où sont également reconnus les PY-NLS). Malgré différentes approches utilisées, et différentes constructions, nous n'avons pas pu résoudre la structure cristallographique de ce complexe. J'ai toutefois résolu la structure de Trn1ΔH8 à 1,81 Å qui a permis d'obtenir des cartes de densité facile à interpréter. La comparaison de cette structure aux deux autres structures de Trn1 seule, a montré des différences de connexion entre les hélices B et A des HEAT H3 et H4 qui sont dues à un empilement différent des trois structures. Par ailleurs, l'étude du complexe Trn1/ADAR1 nous a permis de quantifier par des expériences d'ITC l'interaction entre ces deux partenaires : le K_D est de l'ordre de quelques centaines de nanomolaires. Les expériences de RMN et de SEC-MALS ont permis de montrer l'implication de la boucle H8 dans la reconnaissance d'ADAR1 par Trn1, et ces résultats ont été confirmés par des expériences d'ITC qui ont montrées une meilleure affinité d'ADAR1 pour Trn1FL que

pour Trn1 Δ H8. De plus, les expériences de SAXS ont permis d'obtenir un modèle de l'interaction entre Trn1 et le dsRBD3 d'ADAR1 qui ne rend pas parfaitement compte des données SAXS, mais qui confirme avec des données expérimentales, que le modèle d'interaction proposé par Barraud *et al* est un modèle plausible. Ces expériences ont toutefois validé la reconnaissance d'ADAR1 au niveau C-terminale de Trn1. Malheureusement les données obtenues avec les expériences RMN n'ont pas pu confirmer le modèle obtenu par SAXS, car l'attribution des signaux méthyles affectés par l'interaction d'ADAR1 n'est pas connue. Cette attribution est néanmoins en cours de réalisation par nos collaborateurs, ce qui permettra par la suite de construire un meilleur modèle de l'interaction entre Trn1 et ADAR1.

II. Reconnaissance d'un set de non-PY-NLS par Trn1

En parallèle de l'étude portant sur l'interaction entre la Transportine 1 et l'enzyme d'édition ADAR1, nous avons mené une étude sur la reconnaissance d'un set de non-PY-NLS par Trn1. Il s'agissait d'étendre l'étude de la reconnaissance d'ADAR1 à d'autres substrats de Trn1 qui ne sont jusque-là pas encore caractérisés, du moins pas aussi bien que les PY-NLS. Pour cela, nous avons entrepris de résoudre les structures cristallographiques de différents complexes de Trn1 avec des non-PY-NLS. L'objectif est de définir des règles communes de reconnaissance des non-PY-NLS par Trn1.

Pour l'analyse de la caractérisation des non-PY-NLS par Trn1, un ensemble de non-PY-NLS importé par Trn1 a été étudié. Il s'agit de peptides portant le signal de localisation nucléaire reconnu par Trn1 et ne faisant pas partie de la classe des PY-NLS. Cet ensemble est composé de quatre protéines dont le NLS reconnu par Trn1 ne porte pas de motifs PY : les protéines REV (VIH), L2 (HPV16), L2 (HPV18) et CD44 (HUMAIN). Ces protéines ont été choisies car elles représentaient des séquences identifiées et bien définies reconnues par Trn1, et aussi à cause de la petite taille du NLS reconnu. Les interactions entre Trn1 et ces différentes protéines ont été démontrées par différentes méthodes biochimiques (80).

1 Caractérisation des complexes avec l'utilisation de peptides synthétiques

Comme pour l'étude de l'interaction du complexe Trn1 Δ H8/ADAR1, nous avons dans un premier temps utilisé des peptides pour tenter de résoudre leurs structures en complexe avec Trn1 Δ H8 par cristallographie aux rayons X. Uniquement trois des quatre non-PY-NLS ont été étudiés avec cette approche (REV, HPV16 et CD44) et dans chaque cas, le peptide est composé uniquement de la séquence NLS permettant l'import nucléaire par la voie Trn1-dépendante (Tableau 18). Les différents peptides ont été synthétisés chimiquement et leur pureté est garantie par le fournisseur.

Peptides	Séquences	Caractéristiques
REV_pep	MRQARRNRRRWREER	pI : 13.02 MM = 2183 g/mol
HPV16_pep	MRHKRSKRTKR	pI = 12.96 MM = 1554 g/mol
CD44_pep	NSRRRCGQKKLVINSNGAVEDRKPS	pI = 12.07 MM = 2998 g/mol

Tableau 18 : Tableau regroupant les différents peptides synthétiques étudiés.

Trois peptides synthétiques sont étudiés pour cette partie (pI = point isoélectrique, MM = masse moléculaire en g/mol (fournisseur)).

1.1 Caractérisation physico-chimique des complexes Trn1 Δ H8/peptide

Avant la réalisation des tests de cristallisation des complexes Trn1 Δ H8/peptide, nous avons vérifié l'interaction entre Trn1 Δ H8 et les différents peptides (REV_pep, HPV16_pep et CD44_pep) en solution dans le tampon A.

1.1.1 Étude des interactions Trn1 Δ H8/peptide par DSF et GST Pull-down

Nous avons commencé par réaliser des expériences de DSF pour analyser l'effet de l'ajout des peptides sur la stabilité de la protéine Trn1 Δ H8. Pour cela, la détermination du T_m de Trn1 Δ H8 seule puis suite à l'ajout des peptides à différentes concentrations a été réalisée (ratio 1:1 à 1:20). La comparaison de ces différentes expériences permet de rendre compte des différentes interactions Trn1 Δ H8/peptide mais aussi d'identifier le peptide avec potentiellement la meilleure affinité parmi ceux étudiés.

En absence de peptide, Trn1ΔH8 a un T_m de $59,5 \pm 0,2$ °C. En présence du peptide CD44_pep ($T_m = 59,8 \pm 0,2$ °C), nous ne constatons aucune variation du T_m de Trn1ΔH8 quel que soit le ratio Trn1ΔH8/CD44_pep utilisé (Figure 74A). Le peptide CD44 n'interagit donc pas avec Trn1ΔH8 dans les conditions utilisées.

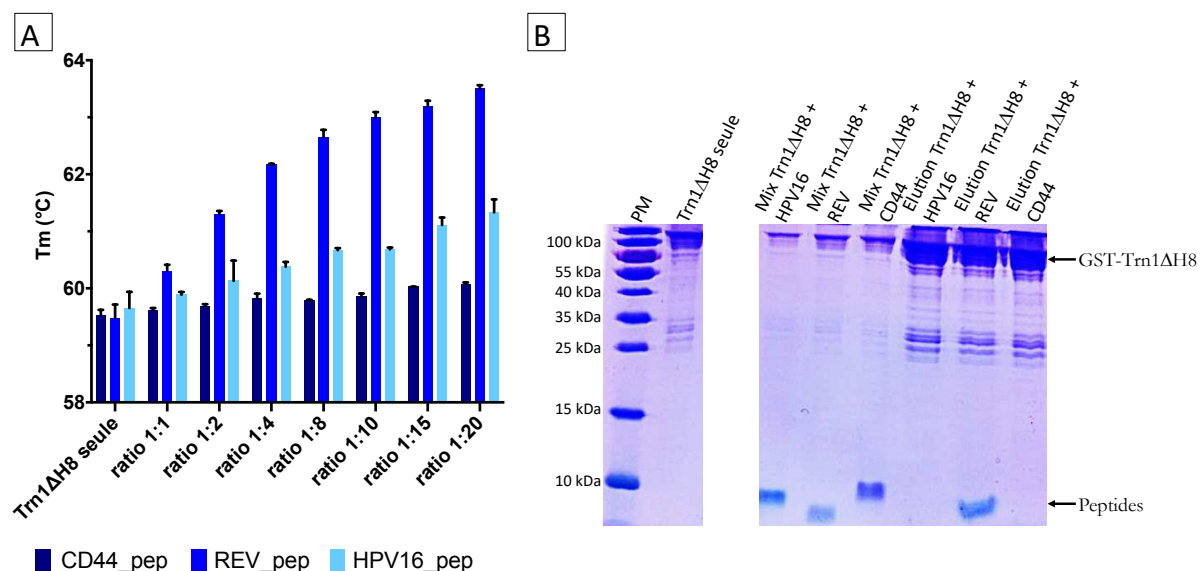


Figure 74 : Caractérisation des interactions Trn1ΔH8/peptides.

(A) Caractérisation des interactions par DSF. Les T_m ont été mesurés pour Trn1ΔH8 (2 μM) en présence et en absence des peptides (2 à 40 μM). **(B) Analyse SDS-PAGE des expériences de caractérisation par GST Pull-down.** Les expériences ont été effectuées avec Trn1ΔH8 (27 μM) en absence et en présence des peptides (216 μM) soit à un ratio 1:8. L'analyse a été réalisée par SDS-PAGE 16 % (PM = poids moléculaire, mix = mélange des deux partenaires, Éluat = éluat de l'expérience GST Pull-down.)

En présence des peptides REV_pep ($T_m = 62,3 \pm 0,4$ °C) et HPV16_pep ($T_m = 60,6 \pm 0,5$ °C), nous constatons une augmentation claire du T_m de Trn1ΔH8 (Figure 74A). Plus la concentration du peptide est élevée, plus la valeur du T_m augmente. Ceci indique que la présence des peptides REV_pep et HPV16_pep permet de stabiliser Trn1ΔH8, et montre que ces peptides interagissent avec Trn1ΔH8. Le peptide REV est celui qui provoque la plus grande variation du T_m .

Nous avons ensuite effectué des expériences de GST pull-down entre Trn1ΔH8 et les différents peptides. Ces expériences ont été réalisées en suivant le protocole des expériences réalisées avec le peptide ADAR1_pep (paragraphe 2.1, page 135). Le contenu de l'éluat analysé a permis de mettre en évidence l'interaction entre GST-Trn1ΔH8 et le peptide REV_pep. En effet, sur le

gel SDS-PAGE (Figure 74B), une deuxième bande correspondant au peptide REV_pep est observée en plus de celle correspondant à GST-Trn1ΔH8. Ce résultat vient confirmer ce que nous avons constaté avec les expériences de DSF. Ces résultats confirment également les résultats obtenus par DSF concernant le peptide CD44_pep. En revanche, pour le peptide HPV16_pep pour lequel une interaction avec Trn1ΔH8 avait été mise en évidence par DSF, aucune bande supplémentaire n'est observée (Figure 74B). Il semblerait que l'interaction entre GST-Trn1ΔH8 et HPV16_pep ne soit pas assez forte, et les trois lavages avant l'élution ont sûrement entraîné la dissociation des deux partenaires et donc l'élimination de HPV16_pep qui n'est ainsi plus observé sur le gel.

1.1.2 Étude des interactions Trn1ΔH8/peptide par BLI

Suite aux résultats qualitatifs obtenus par DSF et GST Pull-down, nous avons voulu quantifier les interactions des complexes Trn1ΔH8/peptide par la technique BLI (*BioLayer Interferometry*). Nous avons mesuré la constante de dissociation K_D de l'interaction Trn1ΔH8/REV_pep. Pour la réalisation des expériences de BLI, Trn1ΔH8 et les peptides ont été dilués dans le tampon BLI (tampon PBS pH 7,4) pour pouvoir effectuer la réaction de biotinylation de Trn1ΔH8.

Nous avons commencé par la mise en évidence de l'interaction entre la protéine Trn1ΔH8 et les peptides. A titre comparatif, les mêmes expériences ont aussi été réalisées avec le peptide ADAR1_pep utilisé pour la caractérisation du complexe Trn1ΔH8/ADAR1_pep. Pour l'ensemble de ces expériences, la protéine Trn1ΔH8 biotinylée est concentrée à 10 μ M, et chacun des peptides est testé à 5 μ M.

Les résultats obtenus par BLI permettent de confirmer certains des résultats obtenus par DSF (Figure 74A). En effet, ils montrent une interaction entre Trn1ΔH8 et REV_pep ainsi que HPV16_pep et une absence d'interaction entre Trn1ΔH8 et CD44_pep (Figure 75A). Il est intéressant de noter que ces résultats révèlent également une interaction entre Trn1ΔH8 et ADAR1_pep, peptide pour lequel aucune interaction n'avait pu être mise en évidence avec les expériences de DSF ou de GST Pull-down (Figure 45).

Nous avons ensuite déterminé le K_D de l'interaction Trn1ΔH8/REV_pep. Pour cela, Trn1ΔH8 biotinylée a été utilisée à 1 μ M et REV_pep a été ajouté à des concentrations allant de 40 nM à 5 μ M. L'interaction Trn1ΔH8/REV_pep se fait avec une association rapide et une dissociation

lente des deux partenaires (Figure 75B-C). Cette observation est en ad quation avec la valeur de K_D obtenue qui est de 386 ± 5 nM (Figure 75D).

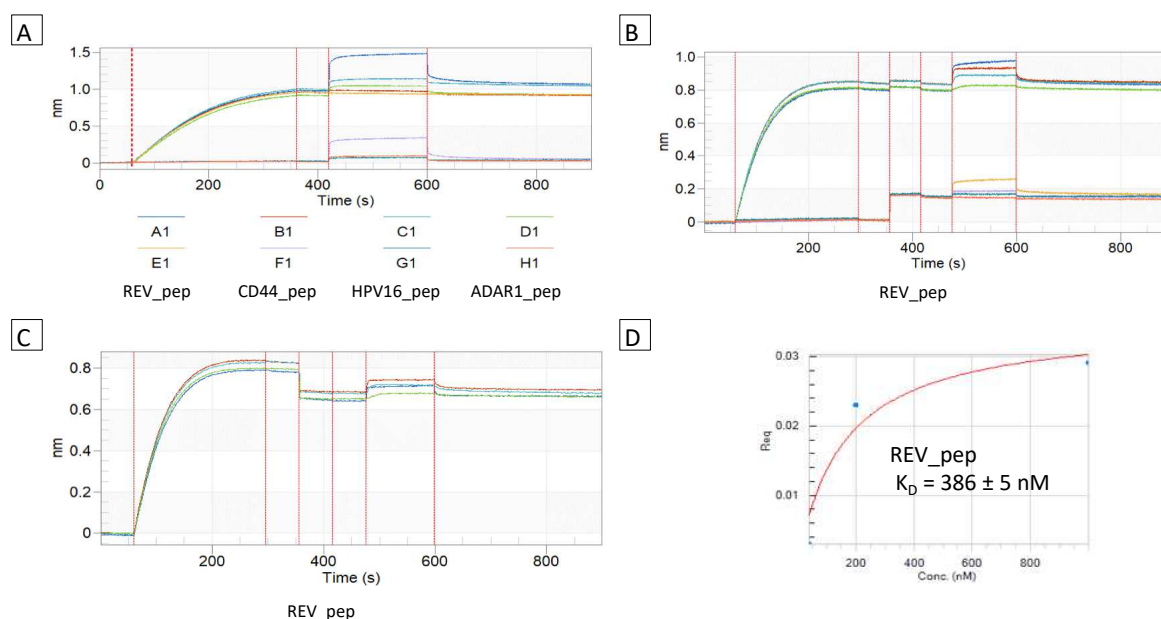


Figure 75 :  tude quantitative de l'interaction Trn1ΔH8/peptides.

(A) Sensorgrammes obtenus par BLI pour l' tude des interactions Trn1ΔH8/peptide. (B- C) Sensorgrammes obtenus par BLI pour la d termination du K_D de l'interaction Trn1ΔH8 / REV_pep. Les sensorgrammes repr sentent les courbes d'association/dissociation en fonction du temps. (D) Estimation de la constante d'affinit  K_D de l'interaction Trn1ΔH8/REV_pep.

En r sum , la caract risation physico-chimique des interactions Trn1ΔH8/peptide  tudi es par DSF, GST Pull-down et BLI a permis de mettre en  vidence que les interactions entre Trn1ΔH8 et les peptides REV_pep et HPV16_pep. REV_pep pr sente la plus grande affinit . L'interaction de Trn1ΔH8/HPV16_pep est plus faible que celle de Trn1ΔH8/REV_pep puisqu'elle n'est pas observ e par GST Pull-down. En revanche, aucune interaction entre Trn1ΔH8 et CD44_pep n'a pu  tre mise en  vidence quelle que soit la technique utilis e. Il semblerait donc que Trn1ΔH8 n'interagisse pas du tout avec le peptide CD44_pep. L' tude de ce dernier n'a pas  t  poursuivie pour cette raison.

1.2 Cristallisation des complexes TrnΔH8/peptide

Les tests de co-cristallisation des diff rents complexes (Trn1ΔH8/REV_pep et Trn1ΔH8/HPV16_pep) ont  t  r alis s en criblage large avec Trn1ΔH8 et les peptides respectivement concentr s entre 0,15 et 0,17 mM et entre 0,75 et 1,7 mM (ratios Trn1ΔH8/peptide allant de 1:5   1:10). Les m langes Trn1ΔH8/peptide ont  t  r alis s dans le

tampon A. Ces exp riences ont permis d'obtenir des cristaux de grande taille pour chaque peptide test  (Figure 76).

Pour chacun des peptides, diff rents cristaux ont  t  mont s sur des boucles de nylon, cryoprot g s par incubation dans une goutte de solution de cristallisation contenant 20– 30 % (v/v) de glyc rol ou d' thyl ne glycol. Apr s collecte et traitements des diff rents jeux de donn es, aucune densit  n'a  t  retrouv e pour le peptide, quel que soit le peptide utilis . Ces cristaux ne constituaient donc pas des co-cristaux de complexes Trn1 H8/peptide. Ceci indique que le m lange Trn1 H8/peptide dans les conditions utilis es n'a pas permis la formation de complexes ou que les complexes n'ont pas r sist  aux conditions de cristallisation.

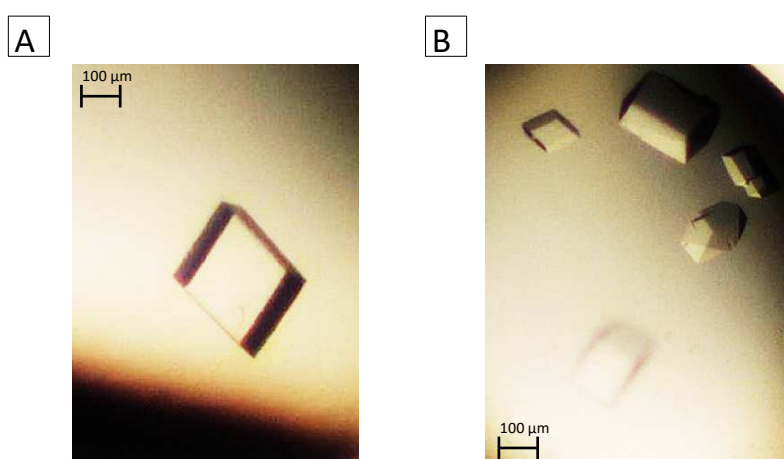


Figure 76 : Tests de co-cristallisation des complexes Trn1 H8/peptide.

(A) Le monocristal cubique obtenu pour le m lange Trn1 H8/REV_pep dans la condition 0,10 M Bis- Tris pH 6,50, 1,50 M (NH  ) SO  , 0,10 M NaCl. (B) Les cristaux de diff rentes formes obtenus pour le m lange Trn1 H8/HPV16_pep dans la condition 0,10 M Hepes pH 7,00, 1 % (w/v) PEG MME 2K, 1,00 M Na   Succinate.

Nous avons ensuite r alis  des exp riences de « trempage » avec des cristaux fraichement form s de Trn1 H8. Les cristaux de Trn1 H8 ont  t  pr lev s sur des boucles de nylon puis tremp s dans des solutions cryoprotectrices suppl ment es d'un des peptides en exc s avant d' tre congel s dans l'azote liquide. Malheureusement, toutes les exp riences de « trempage » r alis es n'ont pas permis d'obtenir de structures de complexes Trn1 H8/peptide. Malgr  des temps de trempage et des concentrations  lev es en peptides augment s, apr s traitement des diff rents jeux de donn es collect s, aucune densit  suppl mentaire pouvant correspondre   celle des peptides n'a  t  visualis e.

2 Caractérisation des complexes avec l'utilisation de protéines chimères

Suite aux résultats précédents, le set de peptides étudiés a été changé du fait de l'absence d'interaction avec le peptide CD44_pep. Il a été remplacé par le peptide HPV18_pep. Nous avons également changé d'approche pour la résolution des structures cristallographiques des différents complexes Trn1ΔH8/peptide étudiés. J'ai construit trois protéines chimères Trn1ΔH8-REV_pep, Trn1ΔH8-HPV16_pep et Trn1ΔH8-HPV18_pep sur le même principe que la chimère Trn1ΔH8-ADAR1_pep (Voir Matériels et Méthodes : paragraphe 1.4, page 71, Figure 26A). Les séquences des NLS des peptides REV_pep, HPV16_pep et HPV18_pep (Tableau 19) entourées d'une vingtaine d'acides aminés en N- et C-terminal, ont été ajoutées à l'extrémité C-terminale de Trn1ΔH8. La taille et la composition du peptide espaceur en N-terminal du NLS ont été choisies pour être flexibles et non encombrantes, permettant ainsi une interaction dans l'arche C-terminale de Trn1ΔH8. Le petit espaceur en N-terminal de REV et les séquences en C-terminal appartiennent à la séquence des protéines et apporteraient une plus grande surface d'interaction lors de la formation du complexe. Dans le cas de HPV16 et HPV18, comme le NLS débute avec la méthionine initiatrice de ces protéines, l'espaceur N-terminal comprend principalement des répétitions de petits résidus flexibles (glycines et sérines) (Tableau 19).

Protéine	Espaceur N-terminal	Séquence NLS	Région C-terminale
REV	AVRLIKLLYQSNPPPNPEGT	RQARRNRRRRWR	ERQRQIHSISERILGTYLGR
HPV16	GSADNVTGGSGGSGGSGGSG	MRHKRS AKRTKRA	SATQLYKTCKQAGTCPPDII
HPV18	GTANDVSGGSGGSGGSGGG	MVSHRAARRKRA	SVTDLYKTCKQSGTCPPDVV

Tableau 19 : Séquences des NLS utilisés pour la construction des chimères.

La séquence minimale du NLS est indiqué en gras, ces peptides ont été fusionnés à l'extrémité C-terminale de Trn1ΔH8.

Le protocole suivi pour l'expression de ces chimères est le même que celui utilisé pour la protéine Trn1ΔH8 avec de légères modifications, notamment l'augmentation de la concentration d'IPTG (paragraphe 1.4.2, page 74). Ce protocole d'expression permet une surexpression suffisante des chimères (Figure 77).

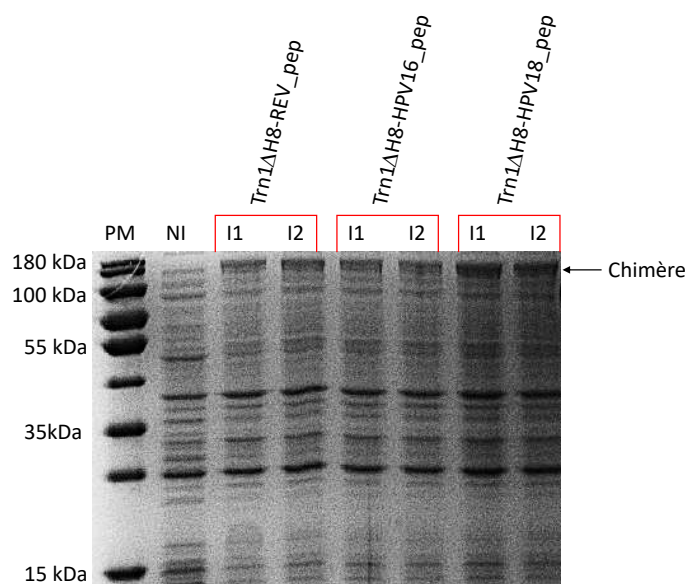


Figure 77 : Expression des chimères Trn1ΔH8-peptides.

Les résultats obtenus pour les expressions de ces chimères ont été révélés par analyse SDS-PAGE 10 % (PM = marqueur de poids moléculaire, NI = l'extrait bactérien total avant l'induction, I1=I2 = l'extrait bactérien total après 20 h d'induction à 30 °C (les quantités de matériel déposées sont identiques pour I1 et I2)).

2.1 Purification des protéines chimères

Le protocole de purification des trois chimères est le même que celui de la protéine Trn1ΔH8 qui consiste en une étape de purification sur une première colonne d'affinité glutathion sepharose, puis sur une seconde colonne échangeuse d'anions après coupure de l'étiquette GST et enfin sur une dernière colonne d'exclusion stérique.

Les différentes étapes de purification sont analysées par SDS-PAGE qui montrent qu'en fin de purification, les trois sont suffisamment pures pour la réalisation de tests de cristallisation (Figure 78). Le rendement de la purification pour les chimères Trn1ΔH8-HPV16_pep, Trn1ΔH8-HPV18_pep et Trn1ΔH8-REV_pep est respectivement d'environ 3 mg, 2 mg et 0,5 mg de protéine pure par litre de culture bactérienne.

Dans le cas de ces chimères, la séparation avec l'étiquette GST est plus efficace par rapport à ce qui avait été observé avec les chimères Trn1ΔH8-ADAR1_pep et Trn1ΔH8-A1_pep. Nous pouvons remarquer la présence de l'étiquette GST dans la « fraction Qtrap » (Figure 78) uniquement parce que le pic représentant la GST et la chimère d'intérêt se chevauchent et que les fractions sont collectées de manière plus large pour récolter un maximum de matériel.

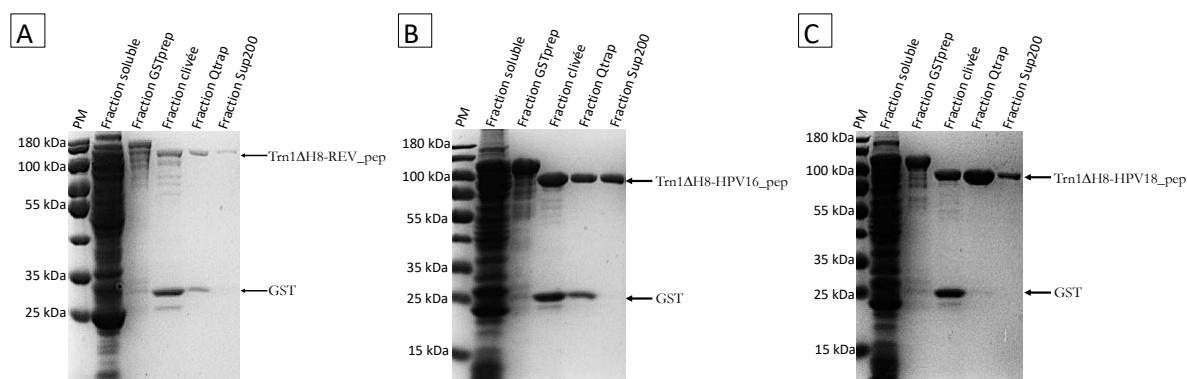


Figure 78 : Purification des chim res Trn1 H8-peptides.

L'analyse de la purification par SDS-PAGE 10 % (PM = marqueur de poids mol culaire) montre une puret  satisfaisante pour les chim res (A) Trn1 H8-REV_pep, (B) Trn1 H8-HPV16_pep et (C) Trn1 H8-HPV18_pep. Cependant, le rendement final de la Trn1 H8-REV_pep est insatisfaisant.

La chim re Trn1 H8-REV_pep pr sente un tr s faible rendement   cause de la propri t  d'oligom risation de la prot ine REV (206). En effet, le fort degr  d'oligom risation de la prot ine REV entraine la formation de diff rentes formes de la chim re qui sont alors  lu es en plusieurs pics lors de la chromatographie d'exclusion st rique. Cette oligom risation entraine alors une diminution du rendement final de purification.

2.2  tude de la stabilit  des prot ines chim res

Nous avons ensuite r alis  des exp riences de DSF pour v rifier la stabilit  des trois chim res. Pour cela, nous avons effectu  des mesures de T_m pour les prot ines chim riques concentr es   2 μ M.

Les prot ines chim riques ont une stabilit  comparable   celle de la prot ine Trn1 H8. Les valeurs des diff rents T_m des chim res sont quasiment identiques   celle du T_m de Trn1 H8 qui est de $60,2 \pm 0,3^\circ\text{C}$ (Figure 79). Seule la chim re Trn1 H8-A1_pep pr sente une meilleure stabilit  par rapport   Trn1 H8, sugg rant une interaction entre le PY-NLS de hnRNP A1 greff  en C-terminal de la Transportine 1 et Trn1 H8. Ces r sultats indiquent, pour chacune des trois chim res Trn1 H8-REV_pep, Trn1 H8-HPV16_pep et Trn1 H8-HPV18_pep, que le non- PY-NLS greff  en C-terminal de la Transportine 1 et Trn1 H8 n'interagissent pas dans les conditions utilis es.

Pour rappel, les tests de cristallisation concernant Trn1 H8-A1_pep avaient permis la production de cristaux, mais qui n'avaient pas pr sent  une bonne qualit  de diffraction.

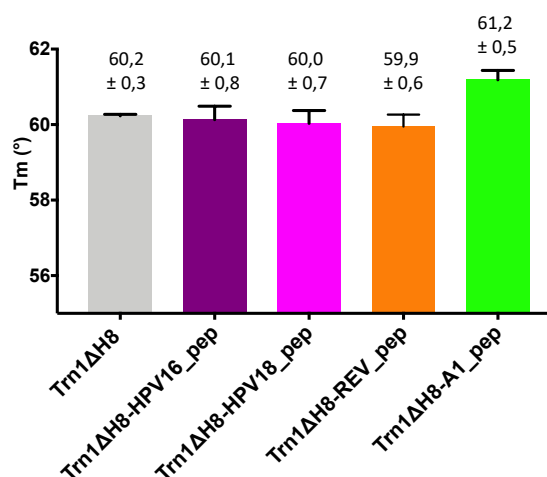


Figure 79 : Étude de la stabilité des chimères Trn1ΔH8-peptide.

Les températures de fusions (Tm en °C) ont été mesurées à une concentration finale de 2 μM pour toutes les protéines.

2.3 Cristallisation des protéines chimères

Par la suite, nous avons effectué des tests de cristallisation pour tenter de résoudre les structures cristallographiques des complexes Trn1ΔH8/non-PY-NLS.

2.3.1 Cristallisation de la protéine chimère Trn1ΔH8-HPV16_pep

Les tests de cristallisation de la chimère Trn1ΔH8-HPV16_pep ont été effectués à une concentration comprise entre 15 et 17 mg/mL. Dans un premier temps, un criblage large a été réalisé en utilisant différents kits commerciaux. Des microcristaux ont été obtenus dans la condition 0,08 M Na Cacodylate pH 6,50, 20,00 % (v/v) glycérol, 14,40 % (w/v) PEG 8K, 0,16 M Ca Acetate (Figure 80A). Ces cristaux sont apparus dans une goutte assez précipitée, nous avons alors diminué la concentration de la protéine. Cette première étape d'optimisation a permis d'obtenir des cristaux très irréguliers en forme de croix (Figure 80B) dans la condition 0,10 M Bis-Tris pH 5,50, 25,00 % (w/v) PEG 3350, 0,20 M MgCl₂. Ces cristaux n'ont pas pu être testés du fait de leur grande fragilité. En effet, au moment de les monter sur les boucles de nylon, ils ont partiellement fondu dans la goutte. Nous avons alors affiné les paramètres de la condition dans laquelle ils étaient apparus. Ceci a permis d'obtenir des petits cristaux en « amande » (Figure 80C) dans la condition 0,10 M Bis-Tris pH 5,50, 20,00 % (w/v) PEG 3350, 0,15 M MgCl₂. Une dernière étape d'optimisation de ces cristaux en « amande » consistant à utiliser un kit d'additifs, a permis de produire des cristaux plus gros de la même (Figure 80D).

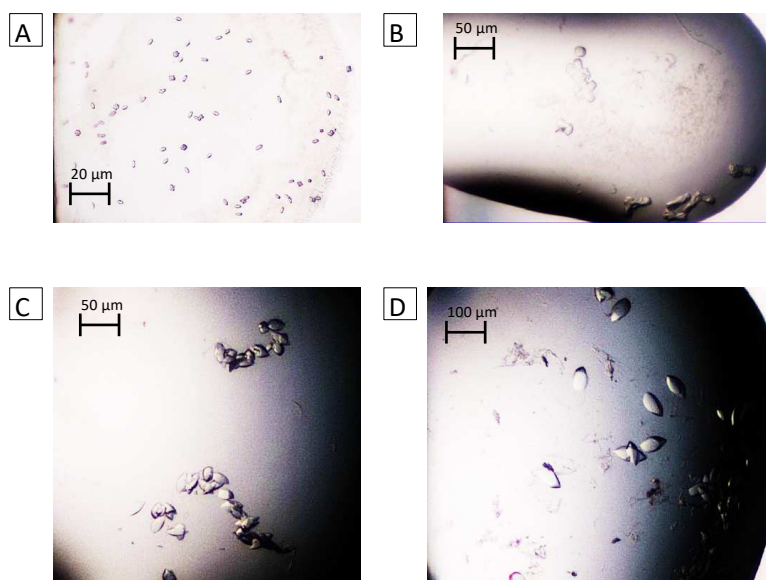


Figure 80 : Cristallisation de la chim re Trn1ΔH8-HPV16_pep.

(A) Les micro-cristaux obtenus dans (0,08 M Na Cacodylate pH 6,50, 20,00 % (v/v) gly cerol, 14,40 % (w/v) PEG 8K, 0,16 M Ca Acetate). (B) Les cristaux irr guliers obtenus dans (0,10 M Bis-Tris pH 5,50, 25,00 % (w/v) PEG 3350, 0,20 M MgCl₂) en optimisant les microcristaux obtenus en (A). (C) Les cristaux en « amande » produits dans (0,10 M Bis-Tris pH 5,50, 20,00 % (w/v) PEG 3350, 0,15 M MgCl₂) en optimisant les cristaux obtenus en (B). (D) Les cristaux en « amande » produits avec la derni re  tape d’optimisation   partir des cristaux obtenus en (C).

Les cristaux (Figure 80D) ont  t  pr lev s et cryoprot g s par incubation dans une goutte de solution de cristallisation contenant 20   30 % (v/v) de gly cerol avant d’ tre congel s dans l’azote liquide. Ces cristaux n’ont pas permis de r soudre la structure cristallographique de la chim re Trn1ΔH8–HPV16_pep, puisqu’ils diffractaient   tr s faible r solution (au-del  de 9  ). L’analyse des clich s de diffraction indique qu’il s’agit tout de m me de cristaux de prot ines. Par ailleurs, d’autres conditions ont  t  optimis es mais ne nous ont pas permis d’obtenir de cristaux pouvant  tre mont s et test s au synchrotron.

2.3.2 Cristallisation de la prot ine chim re Trn1ΔH8-HPV18_pep

Pour la cristallisation de la chim re Trn1ΔH8-HPV18_pep, les exp riences ont  t  r alis es   20 C avec une concentration de la chim re   15 mg/mL. Des cristaux en forme de « croix » tr s irr guliers (Figure 81A) qui avaient d j   t  obtenus avec la chim re Trn1ΔH8-HPV16_pep (Figure 80B) ont  t  produits. J’ai optimis  ces derniers avec l’utilisation d’additifs qui ont permis la production de cristaux en « amande » plus r guliers et de meilleure apparence (Figure 81B)

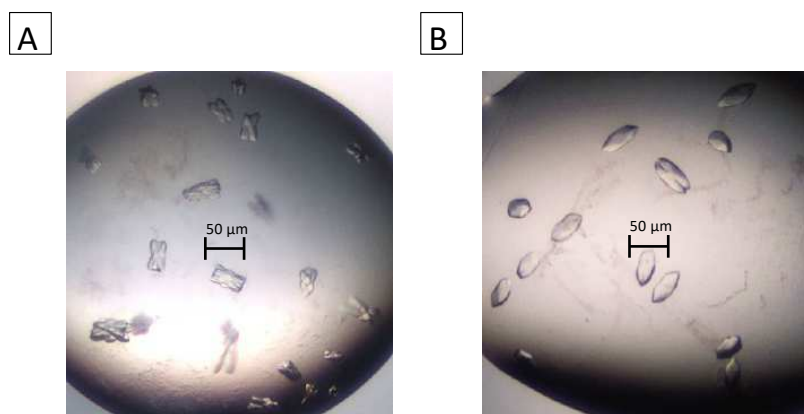


Figure 81 : Cristallisation de la chimère Trn1ΔH8-HPV18_pep.

(A) Les cristaux en « croix » produits dans (0,10 M Hepes pH 7,50, 22,00 % (w/v) PAA 5100, 0,02 M MgCl₂). (B) Les cristaux en « amande » produits après optimisation à partir des cristaux obtenus en (A).

Les cristaux optimisés (Figure 81B) ont été prélevés et cryoprotégés par incubation dans une goutte de solution de cristallisation contenant 25 à 30 % (v/v) de glycérol avant d'être congelés dans l'azote liquide. Les différents tests sur ces cristaux ont permis de collecter différents jeux de données de résolution allant de $\sim 4,1$ à $3,3$ Å. Après traitements des données, aucune densité correspondant au peptide HPV18_pep n'a été observée bien qu'il soit accroché en C-terminal de Trn1ΔH8.

2.3.3 Cristallisation de la protéine chimère Trn1ΔH8-REV_pep

Dans le cas de la chimère Trn1ΔH8-REV_pep, peu de tests de cristallisation ont été menés du fait du faible rendement obtenu lors de sa purification. Les tests de cristallisation de Trn1ΔH8-REV_pep ont été réalisés avec un criblage large à 20°C. Pour cela, Trn1ΔH8-REV_pep était concentrée entre 12 et 15 mg/mL et trois kits commerciaux ont été utilisés : INDEX HT, JCSG+ et MIDAS+. Ces expériences de cristallisation n'ont malheureusement pas permis d'obtenir de cristaux pouvant être montés sur boucle de nylon et testés au synchrotron.

En résumé, l'utilisation de protéines chimères a permis d'obtenir des cristaux pour les chimères Trn1ΔH8-HPV16_pep et Trn1ΔH8-HPV18_pep. Les cristaux obtenus avec la chimère Trn1ΔH8-HPV16_pep n'ont pas eu une forte capacité à diffracter. Les jeux de données analysés à partir des cristaux de Trn1ΔH8-HPV18_pep n'ont pas mis en évidence de densité supplémentaire pouvant correspondre au peptide. Pour la construction des chimères, il est

possible que la longueur du petit espaceur en N-terminal des peptides inspirée de la structure entre Trn1 et un PY-NLS ne soit pas suffisante dans le cas de ces non-PY-NLS. L'affinité de la Transportine 1 pour les PY-NLS est plus forte que celle des non-PY-NLS. Des espaceurs plus longs auraient peut-être apporté plus de points d'interaction avec Trn1 ou plus de flexibilité. Il est également possible que l'interaction de ces non-PY-NLS implique la boucle H8 de Trn1 comme il a été observé pour ADAR1 expliquant notamment l'absence de densité supplémentaire correspondant au peptide dans le cas de HPV18. Ainsi, ces non-PY-NLS auraient une meilleure affinité pour Trn1FL que pour Trn1 Δ H8.

3 Caractérisation des complexes avec l'utilisation de peptides recombinants

En parallèle de l'approche utilisant des chimères, nous avons poursuivi une autre approche consistant à produire des peptides à partir de protéines recombinantes afin de contrôler leur production et leur purification. Pour cette approche, nous nous sommes focalisés sur les protéines REV et HPV16.

Concernant les protéines recombinantes REV et HPV16, nous avons utilisé le système pRSF qui permet de travailler avec des peptides sous forme de protéines de fusion avec une étiquette SUMO en N-terminal. Le protocole de purification est le même pour les deux peptides. Pour ces peptides recombinants l'étiquette SUMO est conservée pour une meilleure solubilité des peptides.

3.1 Purification des protéines SUMO-REV-peptide et SUMO- HPV16-peptide

La première étape de purification d'affinité a été réalisée sur une colonne d'affinité de nickel sepharose. Les protéines SUMO-peptides recombinantes SUMO-REV_NLS et SUMO- HPV16_NLS sont respectivement éluées à 140 et 175 mM d'Imidazole. Les fractions contenant la protéine SUMO-peptide d'intérêt ont ensuite été rassemblées, puis l'étape de chromatographie d'interaction hydrophobe a été effectuée. Les protéines SUMO-peptides sont éluées par décroissance de la force ionique, les protéines SUMO-REV_NLS et SUMO- HPV16_NLS sont respectivement éluées à 175 et 450 mM de sulfate d'ammonium. Les fractions de l'étape précédente ont été rassemblées et finalement injectées sur une colonne d'exclusion stérique équilibrée (Superdex S75) dans le tampon A.

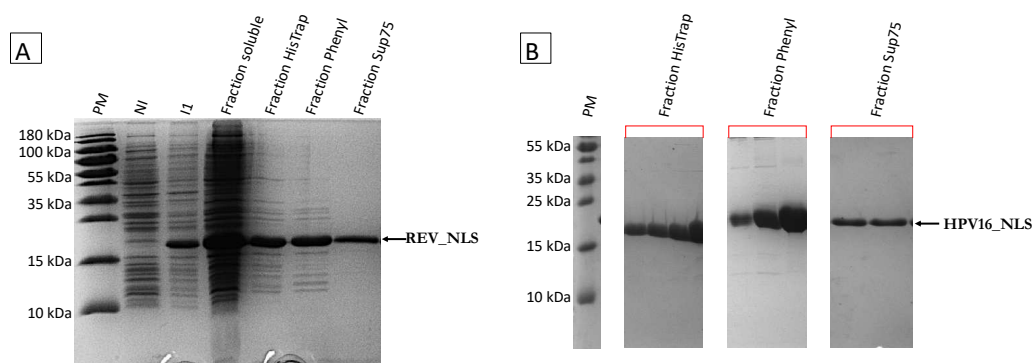


Figure 82 : Purification des protéines SUMO-REV-NLS et SUMO-HPV16-NLS.

La purification des deux protéines est réalisée en trois étapes chromatographiques sans coupure de l'étiquette SUMO. L'analyse des fractions par SDS-PAGE 16 % (PM = marqueur de poids moléculaire) indique une bonne pureté des deux protéines.

Les protéines SUMO-REV-NLS et SUMO-HPV16-NLS sont respectivement éluées après passage de 170 et 165 mL de tampon A. En fin de purification, l'analyse par SDS-PAGE montre, pour les deux protéines une pureté suffisante pour les tests de cristallisation avec le partenaire Trn1ΔH8 (Figure 82). Le rendement de la purification pour les deux protéines est d'environ 5 mg de protéine pure par litre de culture bactérienne.

3.2 Tests de co-cristallisation de Trn1ΔH8 et des peptides recombinants.

Les tests de co-cristallisation ont consisté à mélanger Trn1ΔH8 (0,14 – 0,17 mM) avec les peptides SUMO-REV-NLS/HPV16-NLS (0,42 – 1,40 mM) avant la réalisation des boîtes de cristallisation.

Dans le cas du peptide SUMO-REV-NLS, ces expériences ont permis d'obtenir des cristaux dont un en forme de plaque dans la condition 0,08 M Na Cacodylate, pH 6,50, 20,00 % (v/v) glycérol, 14,40 % (w/v) PEG 8K, 0,16 M Ca Acetate (Figure 83A). En augmentant la concentration de peptide, des cristaux en forme de grandes baguettes en « cluster » ont été obtenus dans la condition 0,10 M K Phosphate Citrate, pH 4,20, 20,00 % (w/v) PEG 8K, 0,20 M NaCl (Figure 83B). Ces cristaux apparaissent dans des gouttes contenant du précipité ce qui rend leur prélèvement plus difficile. Néanmoins, ils ont été montés sur boucles et cryoprotégés dans une solution contenant la solution de cristallisation ainsi que du cryoprotectant (20-25 % (v/v) glycérol ou 15-20 % (v/v) d'éthylène glycol) avant d'être congelés. Les différents jeux de données traités et analysés n'ont pas présenté de densité supplémentaire pouvant correspondre à celle du peptide. Ces expériences n'ont donc pas permis d'obtenir de structure cristallographique du complexe Trn1ΔH8 /SUMO-REV-NLS.

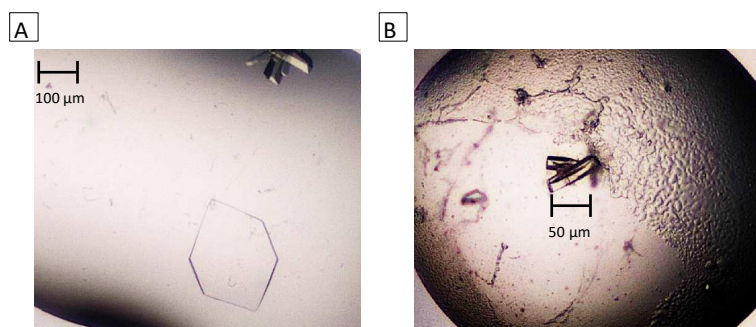


Figure 83 : Tests de co-cristallisation du complexe Trn1 H8/SUMO-REV_NLS.

(A) Les cristaux (dont un en forme de plaque) obtenus dans la condition (0,08 M Na Cacodylate, pH 6,50, 20,00 % (v/v) gly cerol, 14,40 % (w/v) PEG 8K, 0,16 M Ca Acetate). (B) Les cristaux en « cluster » obtenus dans la condition (0,10 M K Phosphate Citrate, pH 4,20, 20,00 % (w/v) PEG 8K, 0,20 M NaCl).

Les tests de co-cristallisation concernant le peptide SUMO-HPV16_NLS ont permis d'obtenir des cristaux de tailles et formes diff rentes (Figure 84A) dans la condition 0,10 M Hepes pH 7,50, 22,00 % (w/v) PAA 5100, 0,02 M MgCl . Ces cristaux ont  t  optimis s en affinant la condition dans laquelle ils  taient apparus. Leur pouvoir de diffraction  tait assez faible, autour de ~ 5-6  . Cette optimisation a ainsi permis de produire un monocristal cubique (Figure 84 B) dont le pouvoir de diffraction  tait meilleur, autour de ~3-4  .

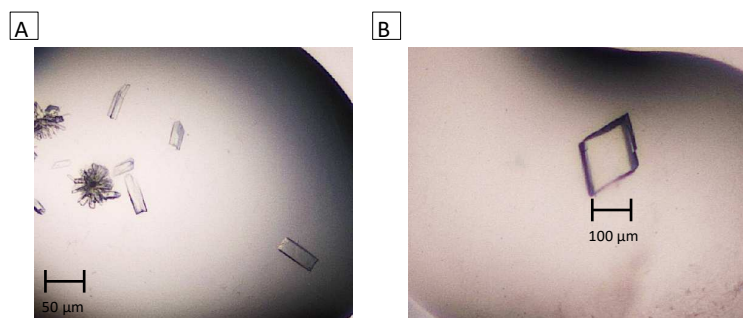


Figure 84 : Tests de co-cristallisation du complexe Trn1 H8/SUMO-HPV16_NLS.

(A) Les cristaux de diff rentes formes obtenus dans la condition (0,10 M Hepes, pH 7,50, 22,00 % (w/v) PAA 5100, 0,02 M MgCl ). (B) Le monocristal obtenu de l'optimisation   partir des cristaux obtenus en (A).

Cependant, comme pour les exp riences de co-cristallisation avec le peptide REV_NLS, ces exp riences n'ont pas permis d'obtenir la structure cristallographique du complexe Trn1 H8/HPV16_NLS. Apr s l'analyse des donn es de diffraction, aucune densit  suppl mentaire pouvant correspondre   celle du peptide HPV16_NLS n'a  t  observ e.

Le but de l'utilisation des peptides recombinants  tait de pouvoir contr ler la production de ces derniers afin d'augmenter les ratios des m langes pour les exp riences de co-cristallisation

favorisant ainsi la formation de co-cristaux Trn1 Δ H8/peptide. Ces expériences ont permis la production de cristaux qui après analyses, se sont avérés ne pas être des co-cristaux des différents complexes étudiés. Ainsi, l'utilisation de protéine recombinante bien qu'ayant permis de contrôler la production des peptides et d'augmenter leur concentration pour atteindre des ratios Trn1 Δ H8/peptide plus importants, n'a pas permis de résoudre la structure cristallographique des différents complexes.

4 Conclusion

Ce chapitre concernait l'étude de complexes entre Trn1 Δ H8 et un set de non-PY-NLS. Il s'agissait d'étendre l'étude sur la caractérisation de l'interaction entre la Transportine 1 et ADAR1 à des non-PY-NLS. Pour cela, nous avons entrepris de résoudre la structure cristallographique de complexes entre Trn1 Δ H8 et différents peptides NLS. Nous avons essayé différentes approches mais aucune n'a été fructueuse et n'a permis de résoudre les structures de ces complexes. Nous avons néanmoins pu déterminer que le NLS avec la meilleure affinité pour Trn1 Δ H8 (parmi ceux étudiés) est celui de REV dont le K_D avec Trn1 Δ H8 est de l'ordre de quelques centaines de nanomolaires.

III. Reconnaissance d'un ARN double brin par ADAR1

Cette troisième et dernière partie porte sur la reconnaissance de l'ARNdb par le dsRBD3 de la protéine ADAR1. Il s'agit d'explorer la deuxième hypothèse sur laquelle est basé le modèle de l'inactivation du NLS en présence d'ARN : le dsRBD3 d'ADAR1 interagit avec l'ARNdb de façon canonique sans réarrangement de structure (voir Introduction : paragraphe 3.2.3.1.2, page 55 et Figure 24, page 54) empêchant ainsi l'interaction avec Trn1. Pour mettre en évidence et caractériser l'interaction entre le dsRBD3 et un ARNdb, nous avons mené une étude cristallographique. La construction du dsRBD3 d'ADAR1, ADAR1 Δ NLS (ADAR1_716-797) dont le NLS constitué par les extrémités flexibles de part et d'autre du dsRBD3 ont été délétées, est celle utilisée pour la caractérisation de cette interaction. Nous avons dans un premier temps réalisé des expériences de cristallisation sur ADAR1 Δ NLS seule, et ensuite des expériences de co-cristallisation avec un ARNdb.

ADAR1 Δ NLS est produite de manière recombinante en suivant le protocole décrit dans le chapitre Matériels et Méthodes (paragraphe 1.1.1, page 112). La préparation de l'ARNdb est aussi présentée dans le chapitre Matériels et Méthodes (paragraphe 1.2, page 115).

1 Purification d'ADAR1 Δ NLS

La purification d'ADAR1 Δ NLS est réalisée en trois étapes : une chromatographie d'affinité de nickel, la coupure de l'étiquette His₆, une seconde chromatographie d'affinité de nickel et une chromatographie d'exclusion stérique.

La première étape de purification d'affinité a été réalisée sur la colonne de nickel sépharose pré-équilibrée dans le tampon G (50 mM Tris-HCl pH 8,0, 1 M NaCl). La protéine ADAR1 Δ NLS est éluée à 140 mM d'imidazole. Les fractions contenant la protéine ont été rassemblées et l'étiquette His₆ a été clivée à l'aide de la protéase TEV. Par la suite, une chromatographie d'affinité de nickel inverse a été effectuée. Cette deuxième étape de purification a permis de séparer l'étiquette s'accrochant à la colonne, de la protéine qui dépourvue de l'étiquette ne s'y accroche pas. Ainsi, la fraction non retenue a été récupérée et concentrée. Enfin une chromatographie d'exclusion stérique a été réalisée sur une colonne Superdex S75 (Sup75) pré-équilibrée dans le tampon I (50 mM Tris-HCl pH 8,0, 100 mM KCl, 1mM TCEP). ADAR1 Δ NLS est éluée après passage de 195 mL de tampon I. L'analyse de la purification par SDS-PAGE montre en fin de purification, une pureté compatible avec des expériences de cristallisation (Figure 85). Le rendement de la purification pour la protéine est d'environ 7 mg de protéine pure par litre de culture bactérienne.

La lyse bactérienne ainsi que la première étape de purification ont été réalisées sans ajout d'agent réducteur dans le tampon de lyse ainsi que dans le tampon G. Ce manque d'agent réducteur dans le tampon de lyse entraîne l'observation d'une forme dimérique de la protéine ADAR1 Δ NLS qui est représentée par une bande (autour de 20 kDa) quasiment aussi intense que celle de la protéine ADAR1 Δ NLS (autour de 10 kDa) (Figure 85). Cette deuxième bande n'est plus observée à la suite de l'ajout de l'agent réducteur (DTT).

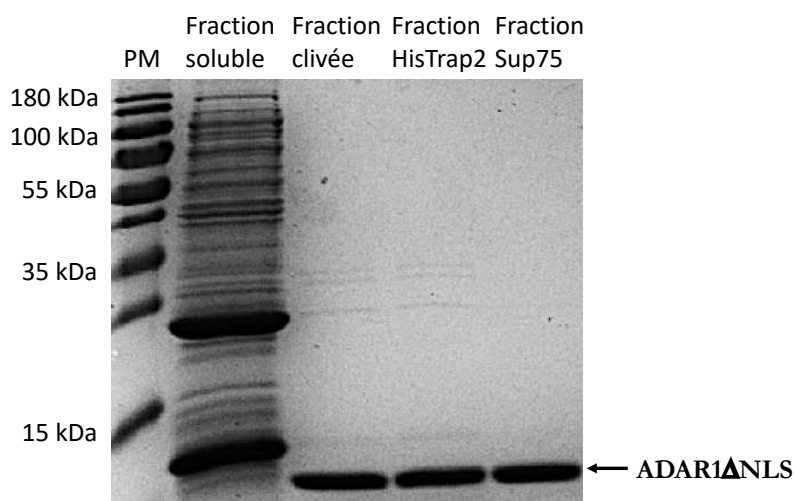


Figure 85 : Purification de la protéine ADAR1ΔNLS.

La Purification d'ADAR1ΔNLS est réalisée en trois étapes chromatographiques avec une coupure de l'étiquette His₆. L'analyse des fractions par SDS-PAGE 16 % (PM = marqueur de poids moléculaire) montre une pureté compatible avec des tests de cristallisation.

2 Structure cristallographique d'ADAR1ΔNLS

Après la purification d'ADAR1ΔNLS en quantité suffisante dans le tampon I, j'ai entrepris de résoudre sa structure cristalline.

2.1 Cristallisation d'ADAR1ΔNLS

Les premiers tests de cristallisation d'ADAR1ΔNLS ont été effectués dans une gamme de concentration de la protéine allant de 10 à 15 mg/mL. Le criblage large a été réalisé avec différents kits commerciaux disponibles au laboratoire et tous les tests ont été pratiqués à 20°C. La formation des gouttes a été faite selon la méthode de la diffusion de vapeur en goutte assise.

Des cristaux en « baguette » ont été obtenus dans différentes conditions notamment celles-ci : 0,10 M Hepes, pH 7,50, 0,80 M KH₂PO₄, 0,80 M NaH₂PO₄ et 1,00 M Tris-HCl pH 8,5, 1,0 M Ammonium sulfate. Ces cristaux sont produits avec la protéine concentrée à 15 mg/mL (Figure 86A) ou 10 mg/mL (Figure 86B). Ces cristaux ont été prélevés sur une boucle en nylon et cryoprotégés en les incubant dans une goutte de solution de cristallisation contenant 20 à 25 % (v/v) de glycérol avant de les congeler dans l'azote liquide. Seulement, ces premiers cristaux ont eu un pouvoir de diffraction assez faible, ils ont diffracté au-dessus de 3 Å. J'ai ensuite entrepris leur optimisation afin d'obtenir un meilleur pouvoir de diffraction mais également réduire le nombre de cristaux produits dans les gouttes.

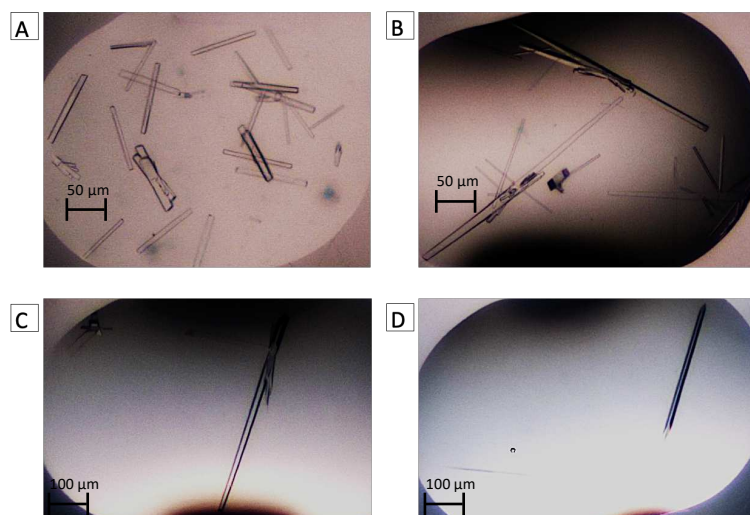


Figure 86 : Cristallisation d'ADAR1ΔNLS.

Les cristaux d'ADAR1ΔNLS (A) (15 mg/mL) obtenus dans (0,10 M Hepes pH 7,50, 0,80 M KH_2PO_4 , 0,80 M NaH_2PO_4). (B) (10 mg/mL) obtenus dans (1,00 M Tris-HCl pH 8,5, 1,0 M Ammonium sulfate). (C) (8 mg/mL) obtenus dans (0,10 M Hepes pH 7,50, 0,80 M KH_2PO_4 , 0,80 M NaH_2PO_4). (D) (5 mg/mL) obtenus dans (0,10 M Na_3 citrate pH 4,00, 20,00 % (w/v) PEG 6K, 1,00 M LiCl).

J'ai commencé par diminuer la concentration d'ADAR1ΔNLS à 8 mg/mL (Figure 86C) puis à 5 mg/mL (Figure 86D). Cette optimisation a permis la production de cristaux, toujours en forme de « baguettes » dans différentes conditions dont une déjà obtenue par criblage large : 0,10 M Hepes pH 7,50, 0,80 M KH_2PO_4 , 0,80 M NaH_2PO_4 et 0,10 M Na_3 citrate pH 4,00, 20,00 % (w/v) PEG 6K, 1,00 M LiCl.

Les cristaux obtenus dans la condition (0,10 M Na_3 citrate pH 4,00, 20,00 % (w/v) PEG 6K, 1,00 M LiCl) ont ensuite été prélevés sur des boucles de nylon et cryoprotégés par incubation dans une goutte de solution de cristallisation contenant 18 à 20 % (v/v) de glycérol avant d'être congelés dans l'azote liquide. Ces cristaux ont diffracté à une résolution entre 3,0 et 1,5 Å.

2.2 Collecte et traitement des données de diffraction

La collecte du meilleur jeu de données obtenu à partir de ces cristaux a été réalisée à SOLEIL sur la ligne de lumière PX2. Le jeu de données a été indexé et intégré avec les scripts XDSME associés au logiciel XDS. Ce programme a déterminé les groupes d'espace $P3_121$ ou $P3_221$ (indiquant une maille trigonale) comme étant les groupes d'espace les plus probables du jeu de données. Le remplacement moléculaire effectué par la suite a montré que ce jeu de données appartient au groupe d'espace $P3_121$. Après l'évaluation de la qualité du jeu de données avec le programme X-triage de la suite de logiciels PHENIX, la mise à l'échelle des données a été

faite avec le programme AIMLESS de la suite CCP4. La limite de résolution a été ajustée à 1,57 Å en se basant sur le facteur $CC_{1/2}$ (Tableau 20), ce qui donne des statistiques satisfaisantes.

Collecte des données	ADAR1ΔNLS
Beamline	PX2 (SOLEIL)
Longueur d'onde (Å)	0,9801
Limite de résolution (Å)	19,75 - 1,57 (1,62 - 1,57)
Groupe d'espace	P 3 ₁ 2 1
Paramètres de maille	
a b c (Å)	a = 44,2 b = 44,2 c = 131,8
α β γ (°)	$\alpha = \beta = 90$ $\gamma = 120$
Multiplicité	17,5 (9,2)
Réflexions mesurées	378752 (18720)
Réflexions uniques	21628 (2019)
R _{merge} (%)	0,15 (1,11)
R _{pim} (%)	0,036 (0,38)
Complétude (%)	98,38 (94,97)
I/s(I)	10,9 (0,64)
CC _{1/2}	0,995 (0,542)

Tableau 20 : Tableau regroupant les statistiques (XDS) du jeu de données d'ADAR1ΔNLS.

Les valeurs entre parenthèses correspondent aux statistiques de la dernière tranche de résolution (1,62 – 1,57 Å).

2.3 Phasage par remplacement moléculaire

Après le traitement des données par XDS, nous avons ensuite procédé au phasage par remplacement moléculaire. La détermination du coefficient de Matthews qui indique le nombre de copies du modèle à chercher dans l'unité asymétrique est réalisée avant la procédure de remplacement moléculaire. Nous avons ainsi établi qu'il y avait deux monomères dans l'unité asymétrique (Tableau 21).

Nombre de molécules / unité asymétrique	Coefficient de Matthews	Pourcentage de solvant	probabilité
1	4,03	69,47	0,01
2	2,01	38,93	0,99
3	1,34	8,40	0,00

Tableau 21 : Tableau de probabilité du nombre de molécules par unité asymétrique de la structure d'ADAR1ΔNLS (déterminée par la méthode Matthews (185)).

Par la suite, nous avons procédé au phasage des données afin d'obtenir la carte de densité électronique. La détermination de phase a été effectuée par remplacement moléculaire en utilisant le programme PHASER MR (CCP4). La première solution de la structure du dsRBD3 résolue par RMN (174) (code PDB 2MDR) a été utilisée comme modèle pour le remplacement moléculaire. Le programme PHASER MR a permis d'obtenir une première solution dans le groupe d'espace $P3_121$ avec des statistiques telles que la solution obtenue est probable, mais non certaine ($TFZ = 7,4$).

2.4 Affinement et validation du modèle

J'ai ensuite effectué un premier affinement de type « jelly-body » avec le programme REFMAC5 (CCP4). Puis, à partir de ces premières cartes affinées, le modèle est reconstruit avec le programme ARP/WARP (classic) de la suite CCP4. Enfin, le modèle est reconstruit manuellement dans le logiciel COOT avec des cycles itératifs de correction-affinement jusqu'à des facteurs d'accord $R_{\text{facteur}} = 20,3 \%$ et $R_{\text{libre}} = 22,1 \%$ (Tableau 22). La validation des statistiques d'affinement a été réalisée avec le serveur Molprobité qui indique de bonnes statistiques, et le modèle a été validé avec le serveur PDB-Validation Server.

Affinement	ADAR1ΔNLS
Limite de résolution (Å)	19,75 - 1,57 (1,62 - 1,57)
No. réflexions	21384 (2020)
Rfacteur	20,3 (29,2)
Rlibre	22,1 (37,9)
No. atomes	1415
RMSD	
Longueur de liaison (Å)	0,006
Angle de liaison (°)	0,76
Ramachandran	
Favorisé (%)	98,78
Autorisé (%)	1,22
Non autorisé (%)	0
Score Molprobité	
Score	1,10
Clashscore, tous atomes	3,11

Tableau 22 : Tableau regroupant les statistiques (REFMAC et MOLPROBITY) d'affinement de la structure cristalline d'ADAR1ΔNLS.

Les valeurs entre parenthèses correspondent aux statistiques de la dernière tranche de résolution (1,62 – 1,57 Å).

2.5 Structure générale d'ADAR1ΔNLS

La structure cristalline d'ADAR1ΔNLS (dsRBD3 d'ADAR1) résolue à 1,57 Å a permis d'obtenir des cartes de densité électronique bien définie dont les chaînes latérales des acides aminés sont facile d'interprétation (Figure 87A). Deux molécules, nommées chaînes A et B sont présentes dans l'unité asymétrique (Figure 87B). Cette solution cristalline corrobore la solution RMN résolue par Barraud *et al* (174). Globalement, ADAR1ΔNLS présente la topologie classique des dsRBD avec les éléments de structure secondaire $\alpha 1$ - $\beta 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ ainsi que l'hélice supplémentaire en N-terminal, l'hélice αN .

La chaîne A présente les deux hélices $\alpha 1$ - $\alpha 2$ qui entourent le feuillet β (Figure 87C). Dans la chaîne A, l'hélice αN n'adopte cependant pas la même conformation que dans la structure RMN. Sa structuration est moins régulière, elle est plus courte et seuls les derniers tours d'hélices sont présents (Figure 87C).

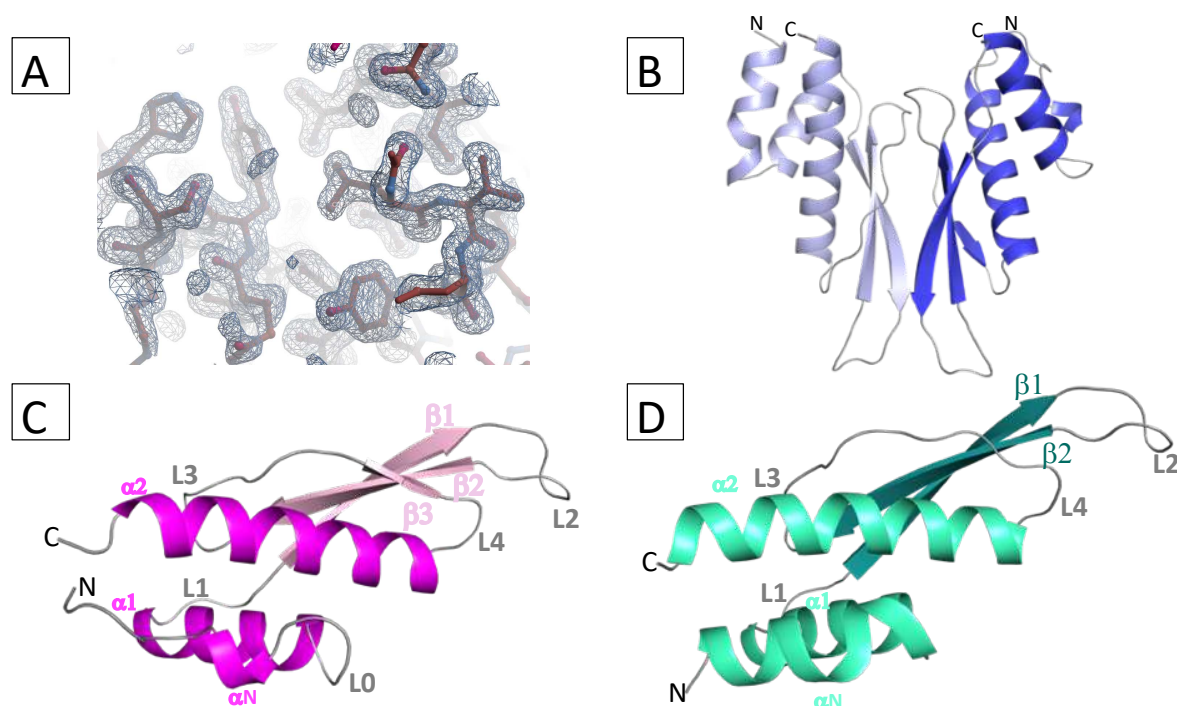


Figure 87 : Structure générale d'ADAR1ΔNLS.

(A) Image de la carte de densité électronique (carte 2Fo-Fc, contours à 1σ) d'ADAR1ΔNLS. (B) Représentation générale en cartoon de la structure cristalline d'ADAR1ΔNLS, les chaînes A et B sont respectivement colorées en bleu et bleu clair. (C) Représentation en cartoon de la chaîne A, les hélices et les brins sont respectivement colorés en magenta et rose. (D) Représentation en cartoon de la chaîne B, les hélices et les brins sont respectivement colorés en vert clair et vert foncé.

La chaîne B présente également les deux hélices $\alpha 1$ - $\alpha 2$ qui entourent le feuillet β (Figure 87 D). Dans cette chaîne B, l'hélice αN est plus longue et moins déformée

par rapport à ce qui est observé dans la chaîne A, et de taille identique à ce qui est observé dans la structure RMN. Ainsi, les deux chaînes, A et B présentent des différences (Figure 87C-D), le RMSD calculé sur les C α est de 0,60 Å pour ces deux chaînes. Les différences majeures sont observées au niveau de l'hélice α N. De très légères différences sont également notées au niveau de la boucle L3 qui connecte les brins β 2 et β 3. Les hélices α 1 et α 2, le feuillet β 1- β 2- β 3, ainsi que les autres boucles des deux chaînes sont très similaires (Figure 88A). Les petites différences observées au niveau de la boucle L3 sont probablement dues à la plasticité de la boucle qui est riche en glycines.

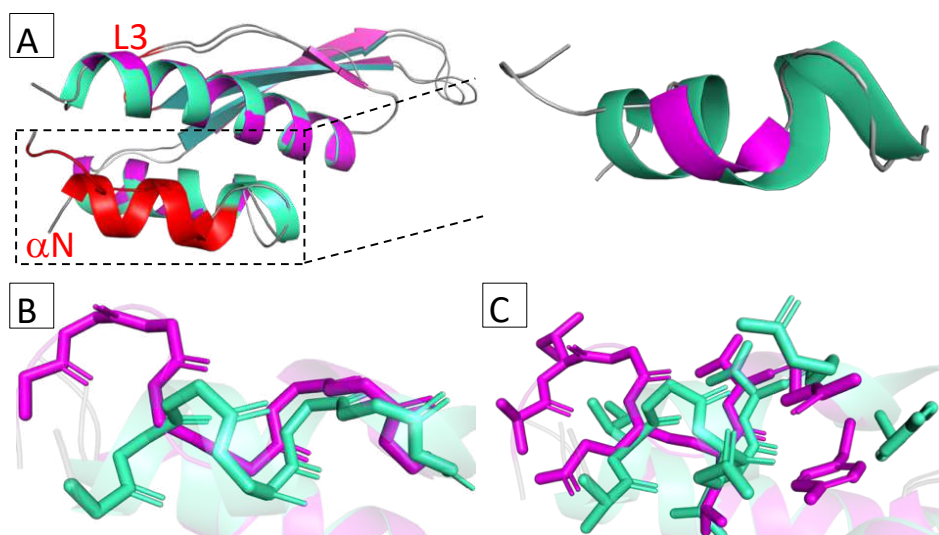


Figure 88 : Comparaison des chaînes de la structure cristalline d'ADAR1 Δ NLS.

Les différences entre les chaînes A (en magenta) et B (vert) sont essentiellement observées au niveau de l'hélice supplémentaire α N. De très légères modifications sont également observées au niveau de la boucle L3. (A) Superposition des chaînes A et B. Les régions où sont observées des différences sont en rouge. (B) Superposition des chaînes principales des chaînes A et B de l'hélice α N. (C) Superposition des chaînes latérales des chaînes A et B de l'hélice α N.

Les différences de structure au niveau de l'hélice α N sont observées au niveau du squelette de l'hélice (Figure 88B), ainsi qu'au niveau des chaînes latérales (Figure 88C). Ces différences peuvent être expliquées par la participation différentes des deux chaînes à l'empilement cristallin dans le cristal (Figure 89). La chaîne A est plus contrainte au niveau de l'hélice α N par rapport à la chaîne B (Figure 89). L'hélice α N de la chaîne A interagit avec deux boucles de la chaîne B. La boucle L0 qui relie les deux hélices α N et α 1, interagit au niveau du début de l'hélice α N, et la boucle L1 qui relie l'hélice α 1 et le brin β 1, interagit au niveau de la région centrale de l'hélice α N (Figure 89A). Quant à l'hélice α N de la chaîne B, elle interagit de manière moins extensive et uniquement avec la boucle L4 (qui relie le brin β 3 et l'hélice α 2) de la chaîne A (Figure 89B). Elle est ainsi plus « libre » et présente moins de déformation. Les

deux chaînes ne sont donc pas soumises aux mêmes contraintes au sein du cristal, et particulièrement l'hélice α N expliquant sa distorsion dans la chaîne A.

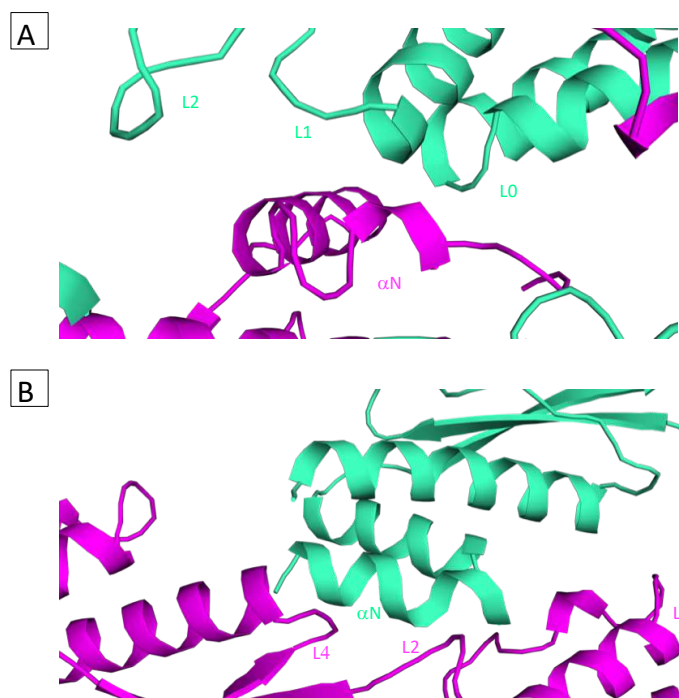


Figure 89 : Comparaison de l'empilement cristallin des chaînes de la structure cristalline d'ADAR1 Δ NLS.

La chaîne A en magenta (A) et B en vert (B) ne participent pas de la même manière à l'empilement au sein du cristal. L'hélice α N de la chaîne A interagit avec les boucles L0 et L1 tandis que celle de la chaîne B interagit brièvement avec la boucle L4. Ces interactions différentes des chaînes A et B dans le cristal entraînent une déformation au niveau de l'hélice α N de la chaîne A qui présente plus de contraintes dans cette région.

Les différences observées au niveau de l'hélice α N entre les chaînes A et B suggèrent que cette hélice est moins stable que les hélices α 1 et α 2 puisqu'elle est perturbée par les forces d'empilement du cristal, qui causent sa déformation dans la chaîne A.

La structure cristalline et la structure RMN en solution d'ADAR1 Δ NLS sont très semblables. La superposition des chaînes A et B de la structure cristalline avec la structure RMN permet d'obtenir un RMSD calculé sur les C α de 0,67 Å et 0,80 Å pour respectivement les chaînes A et B (Figure 90A-B). Les éléments de structures secondaires définissant les dsRBD sont présents dans les deux structures. La superposition de la chaîne A et de la structure en solution montre des différences entre les deux structures notamment au niveau de la boucle L2 reliant les brins β 1 et β 2 (Figure 90A). Cette différence a également été observée pour la superposition de la chaîne B avec la structure en solution (Figure 90B). Les différences au niveau de la boucle L2 sont principalement dues à sa flexibilité. Elle est impliquée dans la reconnaissance

canonique des ARNdb. Elle change de conformation lors de l'interaction de l'ARNdb pour former, avec les régions 1 et 3, la surface d'interaction canonique des dsRBD (163). Ainsi, la boucle L2 est plus dynamique dans la structure en solution que dans la structure cristalline où elle est soumise à différentes contraintes au sein du cristal. Par ailleurs, ces différences viennent principalement de l'orientation des chaînes latérales, plus que de la chaîne principale de la boucle (Figure 90C-D).

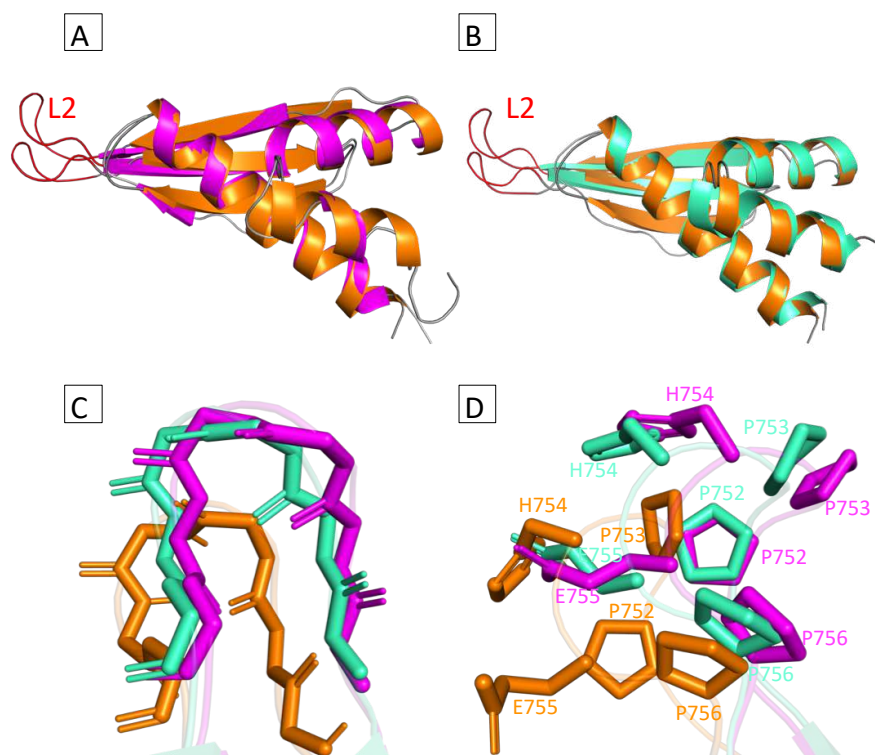


Figure 90 : Comparaison des chaînes A (en magenta) et chaîne B (en vert) de la structure cristalline avec la structure RMN en solution (en orange) (code PDB (2MDR)) d'ADAR1ΔNLS. (A et B) Superposition de la chaîne A / chaîne B de la structure cristalline et de la structure en solution. En rouge la région présentant des différences entre les deux structures. (C) Superposition du squelette des chaînes A et B de la structure cristalline et de la structure en solution. (C) Superposition des chaînes latérales des chaînes A et B de la structure cristalline et de la structure en solution (en orange).

Concernant l'hélice α_N , des différences existent entre les chaînes A et B de la structure cristalline à cause de la différence de l'empilement cristallin entre les deux chaînes. Des différences sont également remarquées au niveau de cette hélice entre les chaînes A/B de la structure cristalline et la structure en solution (Figure 91A-B). La superposition de la chaîne A/B et de la structure en solution ne montre pas de différences au niveau du squelette de cette hélice (Figure 91C). Les différences les plus marquées concernent la conformation de certaines chaînes latérales sur l'ensemble de la structure, et notamment pour le résidu Arg721 (Figure 91D). Les rotamères de ce résidu sont différents pour les chaînes A/B par rapport à

celui de la structure en solution. Chacune des h lices α N ne pr sentent pas les m mes interactions. Globalement, la structure de l'h lice α N de la chaine B est plus proche de celle de la structure en solution que celle de la chaine A.

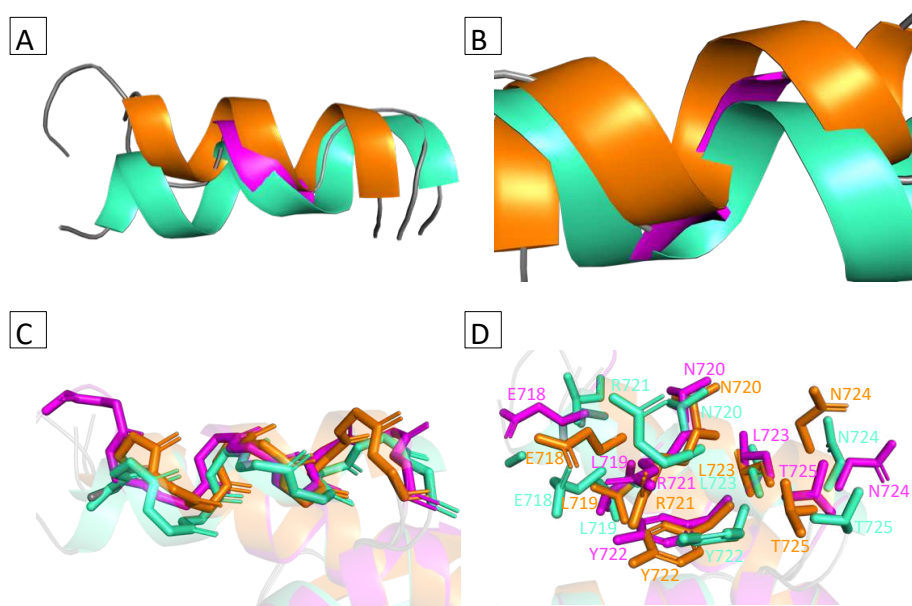


Figure 91 : Comparaison des h lices α N des chaines A (en magenta) et B (en vert) de la structure cristalline et de la structure en solution (en orange) (code PDB (2MDR)) d'ADAR1 Δ NLS.

(A et B) Superposition des h lices α N des chaines A et B de la structure cristalline et de la structure RMN en solution. (C) Superposition du squelette des h lices α N des chaines A et B de la structure cristalline et de la structure en solution. (D) Superposition des chaines lat rales des h lices α N des chaines A et B de la structure cristalline et de la structure en solution.

En r sum , les exp riences de cristallisation d'ADAR1 Δ NLS ont permis, apr s optimisation, de produire des cristaux reproductibles montrant que le degr  de puret  de la prot ine et sa flexibilit  en absence des extr mit s composant le NLS permettent sa cristallisation. Ceci m'a permis de me lancer en confiance dans des tests de co-cristallisation avec l'ARNdb. Ces cristaux ont permis de r soudre la structure cristallographique d'ADAR1 Δ NLS   1,57  . Cette structure vient corroborer la solution obtenue pr c demment par RMN avec une topologie typique des dsRBD ainsi que l'existence de l'h lice α N qui joue un r le primordial dans la construction du NLS d'ADAR1 (paragraphe 3.2.3 en page 51, Figure 23 en page 52). Toutefois, la structure en solution pr sente des diff rences par rapport   la structure au sein du cristal, notamment au niveau de certaines chaines lat rales de la boucle L2 dont la dynamique est mieux repr sent e dans la structure en solution. Par ailleurs, deux mol cules (cha ne A et cha ne B) sont pr sentes dans l'unit  asym trique. Ces deux mol cules diff rent par leur empilement au sein du cristal, ce qui affecte notamment la structure de l'h lice α N dont la structuration semble

moins stable que celles du cœur du dsRBD, car elle peut être déformée par les forces d'empilement au sein du cristal.

3 Structure cristallographique du complexe ADAR1ΔNLS/ARNdb

Après avoir obtenu des cristaux d'ADAR1ΔNLS et avoir résolu sa structure cristallographique, je me suis concentrée sur la co-cristallisation du complexe ADAR1ΔNLS/ARNdb. Pour cela, trois ARN, ARN-I, ARN-II et ARN-III (Tableau 9), de séquences différentes allant de onze à treize nucléotides ont été testés. Ces petits ARN ont une séquence qui leur permet d'une part de s'apparier avec eux-mêmes, et d'autre part de s'auto-assembler en solution pour former de longues hélices d'ARNdb lors de la cristallisation (Figure 34). Les expériences de co-cristallisation des deux partenaires ont été réalisées par la mise en présence de la protéine ADAR1ΔNLS et de l'ARN (ARN-I, ARN-II ou ARN-III) avant la réalisation des boîtes de cristallisation. En plus des gouttes réalisées avec le mélange protéine/ARN, une goutte est également réalisée avec la protéine seule pour servir de contrôle et orienter les conditions à optimiser en priorité.

3.1 Cristallisation du complexe ADAR1ΔNLS/ARNdb

Les premiers tests de co-cristallisation ont été réalisés avec l'ARN-I de séquence « CGAAGCUUCGCG » comptant 12 nucléotides puis avec l'ARN-II, de séquence « CGAACUUCGCG » comptant 11 nucléotides (Figure 34 et Tableau 9).

3.1.1 Tests de co-cristallisation du complexe ADAR1ΔNLS/ARN-I et ADAR1ΔNLS/ARN-II

Pour l'ARN-I, les différentes expériences de co-cristallisation ont permis d'obtenir des sphérolites dans la condition 0,10 M Hepes, pH 7,00, 4,00 M LiCl, 0,01 M MgCl₂. Ces sphérolites (Figure 92B) sont obtenus pour un mélange de deux protéines ADAR1ΔNLS pour un ARN-I, soit un ratio 2:1. Par ailleurs, un mélange de sphérolites et de micro-cristaux sont produits dans la même condition que précédemment pour les ratios 1:1 et 1:2. Les microcristaux sont moins présents avec un mélange 1:1 (Figure 92C) qu'avec un mélange 1:2 (Figure 92D). Ces premiers résultats ont permis de tester et d'observer l'influence du ratio protéine : ARN sur la cristallisation.

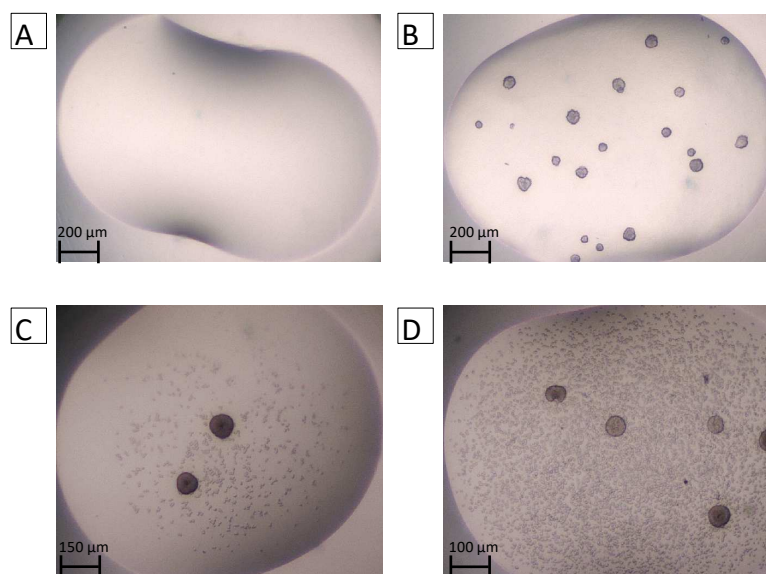


Figure 92 : Tests de co-cristallisation du complexe ADAR1 Δ NLS/ARN-I.

Les résultats des tests sont observés dans la condition 0,10 M Hepes, pH 7,00, 4,00 M LiCl, 0,01 M MgCl₂) avec ADAR1 Δ NLS concentrée à 0,76 mM. **(A)** La condition contrôle, ADAR1 Δ NLS seule. **(B)** Les sphérolites obtenus pour un ratio 2:1. **(C)** Le mélange sphérolites et microcristaux obtenus pour un ratio 1:1. **(D)** Le mélange sphérolites et microcristaux obtenus pour un ratio 1:2.

J'ai ensuite effectué l'optimisation du mélange sphérolites-microcristaux (Figure 93A) obtenus par criblage large. J'ai affiné les paramètres de la condition ayant permis la production de ce mélange, combiné à l'utilisation de différents additifs. Cette optimisation a permis de produire des cristaux irréguliers de petite taille (Figure 93B).

En résumé, les expériences de co-cristallisation d'ADAR1 Δ NLS avec l'ARN-I n'ont pas permis la production de cristaux réguliers pour une étude cristallographique, et ce malgré les différentes optimisations qui ont pu être réalisées. Les derniers cristaux obtenus n'ont pas été optimisés et les expériences de co-cristallisation n'ont pas été poursuivies.

Après les résultats négatifs des expériences de co-cristallisation avec l'utilisation de l'ARN-I de séquence « palindromique » paire, nous avons alors choisi de réaliser les tests de co-cristallisation avec des séquences « palindromiques » impaires de 11 et 13 nucléotides, pour respectivement les ARN-II et ARN-III (Tableau 9).

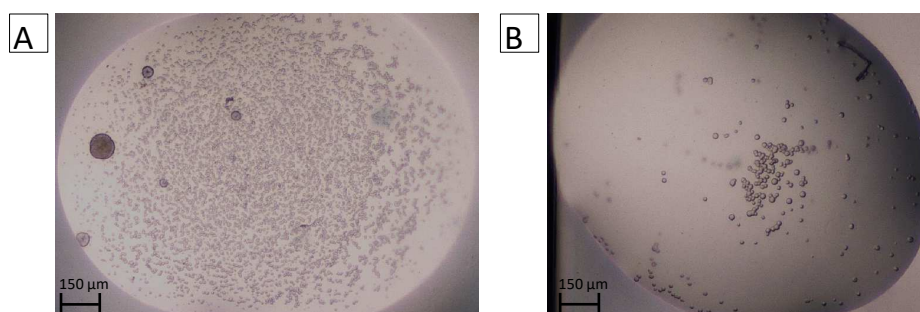


Figure 93 : Optimisation des tests de co-cristallisation du complexe ADAR1 NLS/ARN-I.

La prot ine est concentr e   0,76 mM et l'ARN est   deux  quivalents,   1,52 mM. (A) Le m lange sph rolites et microcristaux obtenus dans (0,10 M Hepes pH 6,50, 3,82 M LiCl, 0,02 M MgCl ). (B) Les cristaux irr guliers obtenus dans (0,10 M Hepes pH 6,50, 3,82 M LiCl, 0,02 M MgCl , 0,03 M glutamate).

Concernant, l'ARN-II, les diff rents tests ont conduit   l'obtention de petits cristaux (Figure 94A) dans la condition 0,05 M Na cacodylate pH 6,00, 10,00 % (w/v) PEG 4K, 0,01 M CaCl , 0,20 M KCl. Ces petits cristaux sont obtenus pour un m lange d'une prot ine ADAR1 NLS pour deux ARN-II, soit un ratio 1:2.

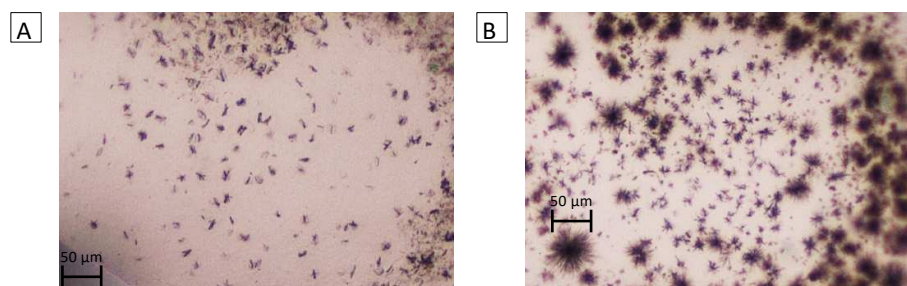


Figure 94 : Tests de co-cristallisation du complexe ADAR1 NLS/ARN-II.

La prot ine est concentr e   0,76 mM et l'ARN est   deux  quivalents,   1,52 mM. (A) Les petits cristaux obtenus dans (0,05 M Na cacodylate pH 6,00, 10,00 % (w/v) PEG 4K, 0,01 M CaCl , 0,20 M KCl). (B) Les  oursins produits dans (0,05 M Na cacodylate pH 6,00, 5,00 % (w/v) PEG 4K, 0,03 M CaCl , 0,23 M KCl, 4 % (v/v) 1,2,6-Hexanetriol).

J'ai ensuite proc d    l'optimisation de ces petits cristaux comme pr c demment, avec l'ARN- I. Cette optimisation a permis la production d' oursins (Figure 94B) dans la condition 0,05 M Na cacodylate pH 6,00, 5,00 % (w/v) PEG 4K, 0,03 M CaCl , 0,23 M KCl, 4 % (v/v) 1,2,6-Hexanetriol. Comme avec l'ARN-I, les tests de co-cristallisation d'ADAR1 NLS et l'ARN-II n'ont pas permis la production de cristaux compatibles avec une  tude cristallographique. Les  oursins obtenus n'ont pas  t  optimis s et les tests de co- cristallisation entre ADAR1 NLS et l'ARN-II n'ont pas  t  poursuivis.

3.1.2 Cristallisation du complexe ADAR1 Δ NLS/ARN-III

Suite à ces résultats, nous avons alors concentré nos efforts sur l'ARN de séquence « palindromique » impaire de plus grande taille. Comme pour les ARN-I et ARN-II, le criblage large d'ADAR1 Δ NLS avec l'ARN-III, de séquence « CGAAGCCUUCGCG » comptant 13 nucléotides (Tableau 9) a été réalisé.

Les tests de co-cristallisation d'ADAR1 Δ NLS avec l'ARN-III ont permis de produire des cristaux en « aiguilles » (Figure 95B) dans la condition 0,05 M MES pH 5,60, 10,00 % (w/v) PEG 400, 0,20 M KCl, 0,01 M MgSO₄. Ces petits cristaux sont obtenus pour un mélange d'une protéine ADAR1 Δ NLS pour un ARN-III, soit un ratio 1:1. Par ailleurs, des cristaux en « aiguilles » (Figure 95C) sont également produits dans la même condition que précédemment pour un ratio 1:2 soit pour un mélange d'une protéine ADAR1 Δ NLS pour deux ARN-III.

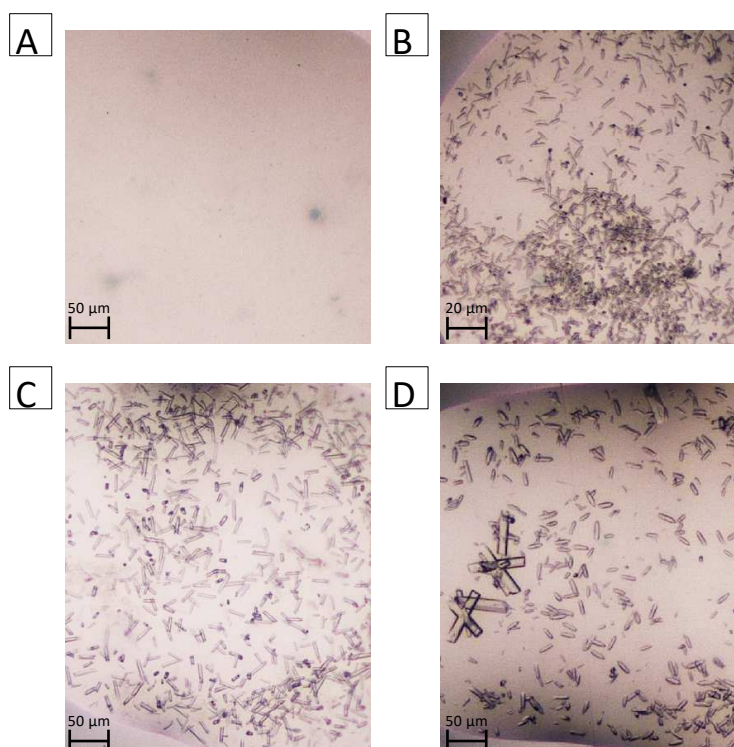


Figure 95 : Tests de co-cristallisation du complexe ADAR1 Δ NLS/ARN-III.

La protéine ADAR1 Δ NLS est concentrée à 0,76 mM (A) La condition contrôle, ADAR1 Δ NLS seule dans (0,05 M MES pH 5,60, 10,00 % (w/v) PEG 400, 0,20 M KCl, 0,01 M MgSO₄). (B) Les cristaux en aiguille obtenus dans (0,05 M MES pH 5,60, 10,00 % (w/v) PEG 400, 0,20 M KCl, 0,01 M MgSO₄) avec un ratio 1:1. (C) Les cristaux en aiguille obtenus dans (0,05 M Na cacodylate pH 6,00, 10,00 % (w/v) PEG 4K, 0,01 M CaCl₂, 0,20 M KCl) avec un ratio 1:2. (D) Les cristaux en baguette produits dans (0,05 M Na cacodylate, pH 6,00, 5,00 % (w/v) PEG 4K, 0,03 M CaCl₂, 0,23 M KCl) après optimisation, avec un ratio 1:2.

J'ai ensuite optimisé ces cristaux en « aiguilles » en affinant les paramètres de la condition qui ont permis leur production. Cette optimisation a permis l'obtention de cristaux en « baguette » (Figure 95D). Les cristaux obtenus dans la condition 0,05 M Na cacodylate pH 6,00, 5,00 % (w/v) PEG 4K, 0,03 M CaCl_2 , 0,23 M KCl ont été prélevés sur des boucles de nylon et cryoprotégés par incubation dans une goutte de solution de cristallisation contenant 20 % (v/v) de glycérol avant d'être congelés dans l'azote liquide.

3.1.2.1 Collecte et traitement des données de diffraction

La collecte du meilleur jeu de données obtenu à partir de ces cristaux a été réalisée à SOLEIL sur la ligne de lumière PX2. Le jeu de données a été indexé et intégré avec les scripts XDSME associés au logiciel XDS. Ce programme a proposé deux groupes d'espace possibles : $P6_122$ et $P6_522$. Le remplacement moléculaire effectué par la suite a montré que ce jeu de données appartient au groupe d'espace $P6_122$. Le jeu de données est assez fortement anisotrope, avec une diffraction allant jusqu'à $\sim 2,8$ Å dans une dimension et n'allant que jusqu'à $\sim 3,3 - 3,4$ Å dans les deux autres dimensions. Une coupure « classique » amènerait à couper ce jeu à $\sim 3,1 - 3,2$ Å, mais avec des statistiques médiocres à haute résolution. La mise à l'échelle des données a ainsi été réalisée avec le serveur STARANISO afin de corriger cette anisotropie. Les réflexions ont ainsi été réduites, mises à l'échelle et coupées de manière anisotrope jusqu'à 2,80 Å de résolution (Tableau 23). A cette limite de résolution, le coefficient $CC_{1/2}$ dans la dernière tranche de résolution est autour de 0,5 supérieure à la limite de 0,3.

Collecte des données	ADAR1ΔNLS/ARN-III
Beamline	PX2 (SOLEIL)
Longueur d'onde (Å)	0,9999
Limite de résolution (Å)	44,58 - 2,80 (2,99 - 2,80)
Groupe d'espace	P 6 ₁ 2 2
Paramètres de maille	
a b c (Å)	a = 91,3 b = 91,3 c = 207,5
α β γ (°)	$\alpha = \beta = 90$ $\gamma = 120$
Multiplicité	19,2 (18,3)
Réflexions mesurées	192634 (68448)
Réflexions uniques	10044 (503)
R _{merge} (%)	0,163 (2,67)
R _{pim} (%)	0,038 (0,633)
Complétude (%)	
sphérique	75,9 (22,1)
ellipsoïdale	94,9 (88,6)
I/s(I)	16,6 (1,2)
CC _{1/2}	1,00 (0,52)

Tableau 23 : Tableau regroupant les statistiques (XDS et STARANISO) du jeu de données d'ADAR1ΔNLS en complexe avec l'ARN-III.

Les valeurs entre parenthèses correspondent aux statistiques de la dernière tranche de résolution (2,99 – 2,80 Å).

3.1.2.2 Phasage par remplacement moléculaire

Une fois traité par XDS et STARANISO, la qualité du jeu de données a été évaluée avec le programme X-triage (PHENIX) indiquant une qualité acceptable pour le jeu de données. Nous avons ensuite procédé au phasage par remplacement moléculaire.

Au vu du faible poids moléculaire de chaque entité (ADAR1ΔNLS : MW = 9236 g/mol, ARN- III : MW= 4115 g/mol) nous ne nous sommes pas basés sur la méthode de Matthews pour la détermination de la composition de l'unité asymétrique. Nous avons plutôt utilisé différentes combinaisons des deux entités (différentes constructions d'ARN-III : simple brin, double brin) pour trouver la bonne solution.

Nous avons procédé au phasage des données par remplacement moléculaire en utilisant le programme PHASER MR (CCP4). J'ai utilisé comme modèle notre structure cristallographique d'ADAR1ΔNLS seule ainsi qu'un ARN modèle de structure en hélice A idéale construit dans COOT. J'ai essayé différentes combinaisons (1 molécule d'ADAR1ΔNLS et deux ARNdb de

13 nucléotides chacun, 2 molécules d'ADAR1 Δ NLS et deux ARNdb de 26 nucléotides chacun...). Dans un premier temps, j'ai pu obtenir une première solution en recherchant une molécule d'ADAR1 Δ NLS ainsi qu'une molécule d'ARN-III double brin. Les statistiques étaient satisfaisantes pour être considérées comme un modèle certain (LLG = 314 ; TFZ = 8,7). Puis, à partir de cette première solution, j'ai recherché une deuxième molécule d'ADAR1 Δ NLS. Ceci a permis l'obtention d'une solution partielle avec un score TFZ de 28,2. Après avoir examiné les cartes de densité, il apparaissait qu'il manquait une molécule d'ARN simple brin, et j'ai finalement trouvé le bon modèle avec une unité asymétrique complète (TFZ = 39,0) en recherchant à partir de la solution partielle précédente, une molécule d'ARN-III simple brin supplémentaire. L'unité asymétrique est ainsi composée de deux molécules d'ADAR1 Δ NLS, une molécule d'ARN-III double brin (de 26 nucléotides) ainsi qu'une molécule d'ARN-III simple brin (de 13 nucléotides) (Figure 96).

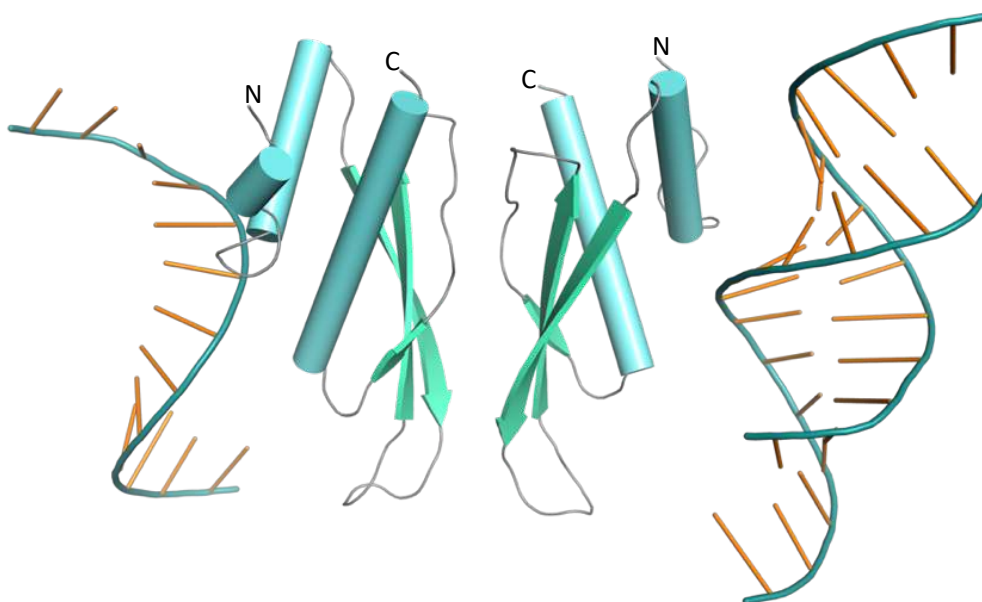


Figure 96 : Composition de l'unité asymétrique du modèle du complexe ADAR1 Δ NLS/ARN-III. L'unité asymétrique est composée de deux monomères d'ADAR1 Δ NLS (en cyan/vert), et une molécule d'ARN-III simple brin ainsi qu'une molécule d'ARN-III double brin (bleu ciel).

3.1.2.3 Affinement et validation du modèle

La solution de remplacement moléculaire correspondant au modèle final complet est premièrement affiné avec un affinement de type « jelly-body » avec le programme REFMAC5 (CCP4). Le modèle est ensuite globalement affiné avec le programme LOW RESOLUTION REFINEMENT (CCP4). L'affinement est poursuivi avec le programme REFMAC5.

Affinement	ADAR1ΔNLS/ARN-III
Limite de résolution (Å)	44,62 - 2,80 (2,99 - 2,80)
No. réflexions	10043 (494)
Rfacteur	25,5 (42,3)
Rlibre	28,6 (51,9)
No. atomes	2029
RMSD	
Longueur de liaison(Å)	0,0068
Angle de liaison (°)	1,33
Ramachandran	
Favorisé (%)	96,71
Autorisé (%)	3,29
Non autorisé (%)	0
Score Molprobit	
Score	2,11
Clashscore, tous atomes	12,32

Tableau 24 : Tableau regroupant les statistiques (REFMAC et MOLPROBITY) d’affinement de la structure cristallographique d’ADAR1ΔNLS en complexe avec l’ARN-III.

Les valeurs entre parenthèses correspondent aux statistiques de la dernière tranche de résolution (2,99 – 2,80 Å).

Enfin, le modèle est reconstruit manuellement dans le logiciel COOT avec des cycles itératifs de correction-affinement jusqu’à des facteurs d’accord Rfacteur = 25,5 % et Rlibre = 28,6 % (Tableau 24). La validation des statistiques d’affinement a été effectuée avec le serveur Molprobit qui indique des statistiques correctes, et le modèle a été validé par le serveur PDB- Validation Server.

3.2 Structure générale du complexe ADAR1ΔNLS/ARN-III

Contrairement à la structure cristalline d’ADAR1ΔNLS seule, la structure cristallographique du complexe ADAR1ΔNLS/ARN-III résolue à 2,80 Å ne permet pas toujours de placer les chaînes latérales de la protéine de façon non équivoque dans la carte de densité (Figure 97A). En revanche la densité électronique correspondant à l’ARN-III est très bien définie. La structure cristalline du complexe est composée de deux molécules d’ADAR1ΔNLS, d’une molécule double brin de l’ARN-III, et d’une molécule simple brin de l’ARN-III dont le complémentaire est obtenu par symétrie (Figure 97B).

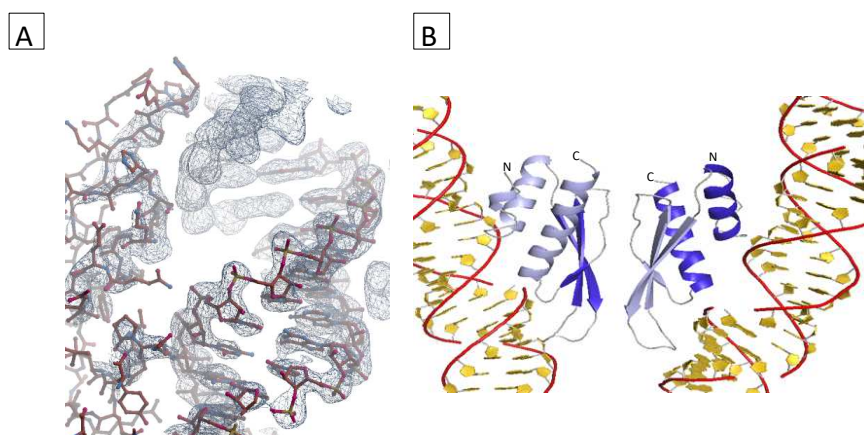


Figure 97 : Structure générale du complexe ADAR1ΔNLS/ARN-III.

(A) Image de la carte de densité électronique (carte 2Fo-Fc, contours à 1σ) d'ADAR1ΔNLS en complexe avec l'ARN-III. (B) Représentation générale en cartoon de la structure cristalline d'ADAR1ΔNLS en complexe avec l'ARN-III, les deux molécules d'ADAR1ΔNLS sont colorées en bleu.

La structure cristallographique entre la protéine ADAR1ΔNLS et l'ARN-III a révélé que suite à l'interaction entre ces deux partenaires, la protéine ADAR1ΔNLS ne subit aucun changement de conformation majeur par rapport à sa forme libre (Figure 98). La superposition de la chaîne A de la structure libre avec celle de la structure en complexe permet d'obtenir un RMSD calculé sur les Cα de 0,45 Å. Ceci montre que la reconnaissance de l'ARN par le dsRBD3 d'ADAR1 n'entraîne pas de changement majeur de conformation au niveau du dsRBD3 d'ADAR1.

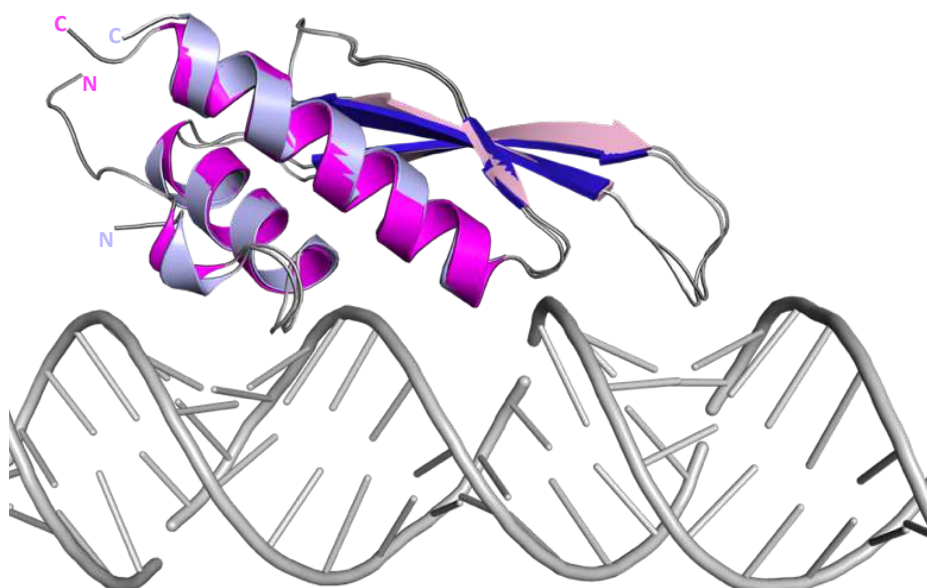


Figure 98 : Comparaison des chaînes A de la structure libre et en complexe d'ADAR1ΔNLS.

La superposition des chaînes A de la structure libre (en magenta/rose) et de la structure en complexe (en bleu/bleu ciel) montre qu'il n'y a pas de restructuration d'ADAR1ΔNLS suite à la liaison de l'ARNdb (en gris).

La structure cristallographique du complexe form  entre la prot ine ADAR1 NLS et l'ARN-III montre qu'ADAR1 NLS interagit de mani re canonique avec l'ARN-III (Figure 99). En effet, les trois grandes r gions d'interactions caract ristiques de la reconnaissance canonique de l'ARNdb par un dsRBD sont retrouv es dans l'interaction entre la prot ine ADAR1 NLS et l'ARN-III. L'h lice $\alpha 1$ et la boucle L2 reliant les brins $\beta 1$ et $\beta 2$ reconnaissent bien l'ARN-III au niveau des deux petits sillons successifs. Ces interactions impliquent des groupes fonctionnels des bases et du ribose de l'ARN. La boucle L4 reliant le brin $\beta 3$ et l'h lice $\alpha 2$ ainsi que la partie N-terminale de l'h lice $\alpha 2$ reconnaissent  galement bien le grand sillon situ  entre les deux petits sillons reconnus pr c demment. Ces interactions impliquent les phosphates du squelette sucre-phosphate de l'ARN. Ces diff rents  l ments s'associent pour former une surface de liaison qui va sp cifiquement reconna tre la forme A de l'h lice d'ARN double brin.

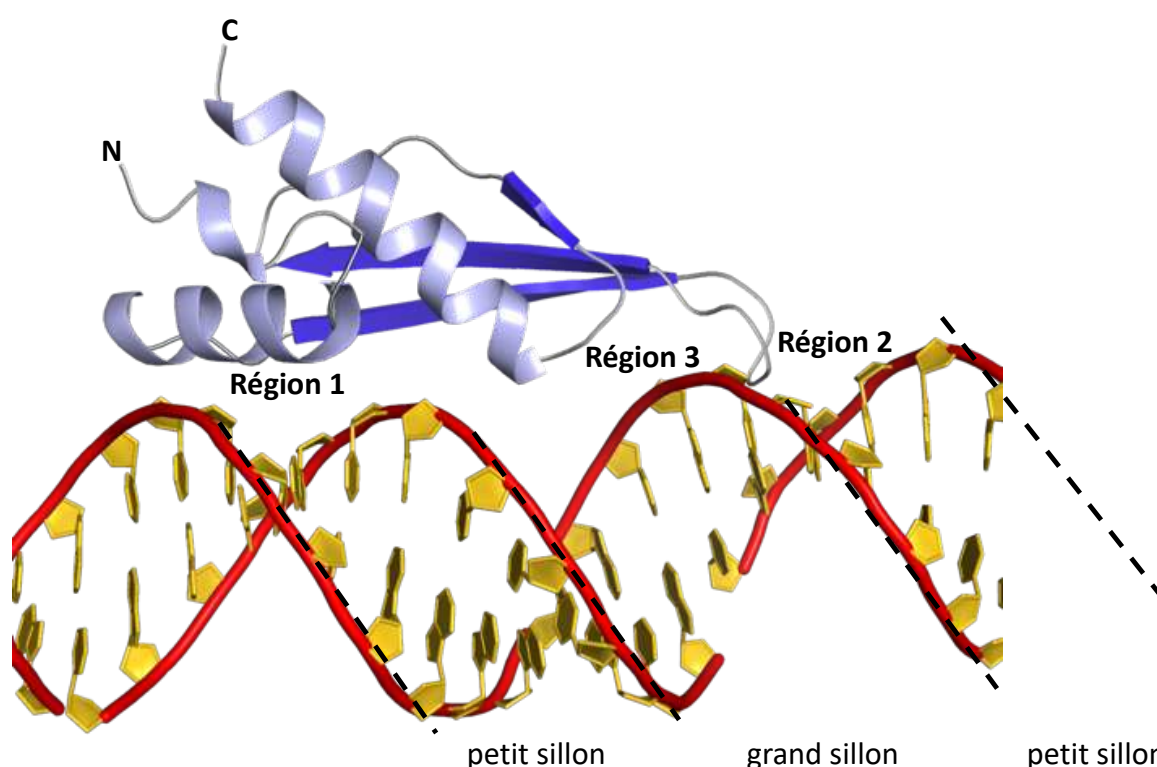


Figure 99 : Repr sentation de l'interaction canonique entre ADAR1 NLS et l'ARN-III.

La reconnaissance de l'ARN-III par la prot ine ADAR1 NLS implique des interactions au niveau de trois r gions d'ADAR1 NLS, caract ristiques de la reconnaissance canonique. La prot ine ADAR1 NLS (au-dessus) est repr sent e en cartoon et est color e en bleu, les h lices en bleu clair et le feuillet β en bleu fonc . L'ARN-III (en-dessous) est repr sent e en cartoon avec le squelette sucre-phosphate color  en rouge et les bases color es en jaune.

3.3 Détails de l'interaction ADAR1 Δ NLS/ARN-III

3.3.1 Détails de l'interaction canonique ADAR1 Δ NLS/ARN-III

Comme décrit plus haut, trois régions sont impliquées dans l'interaction canonique entre un dsRBD et un ARNdb. Des interactions impliquant ces trois régions sont révélées par la structure cristallographique du complexe entre la protéine ADAR1 Δ NLS et l'ARN-III.

Tout d'abord, la région 1 qui se situe au niveau de la première hélice, à l'extrémité N-terminale de la protéine ADAR1 Δ NLS constitue la première zone de liaison canonique à l'ARNdb (Figure 99). Cette région interagit avec l'ARNdb au niveau du petit sillon. Il s'agit d'une hélice amphipatique présentant une surface hydrophobe qui participe à la constitution du cœur hydrophobe du dsRBD, et une surface hydrophile qui participe à la reconnaissance de l'ARNdb en établissant des contacts directs avec le ribose de l'ARNdb.

L'analyse de la structure du complexe ADAR1 Δ NLS/ARN-III montre que la surface d'interaction de la région 1 avec l'ARN est constituée de deux acides aminés. Des contacts directs avec des groupes 2' hydroxyles (OH) du ribose de chacun des brins de l'ARN-III sont observés (Figure 100). Les résidus E733 (fonction carboxyle) et R736 (fonction guanidinium) forment des contacts directs avec les groupements 2' OH respectivement des bases U9 et A4 de l'ARN-III.

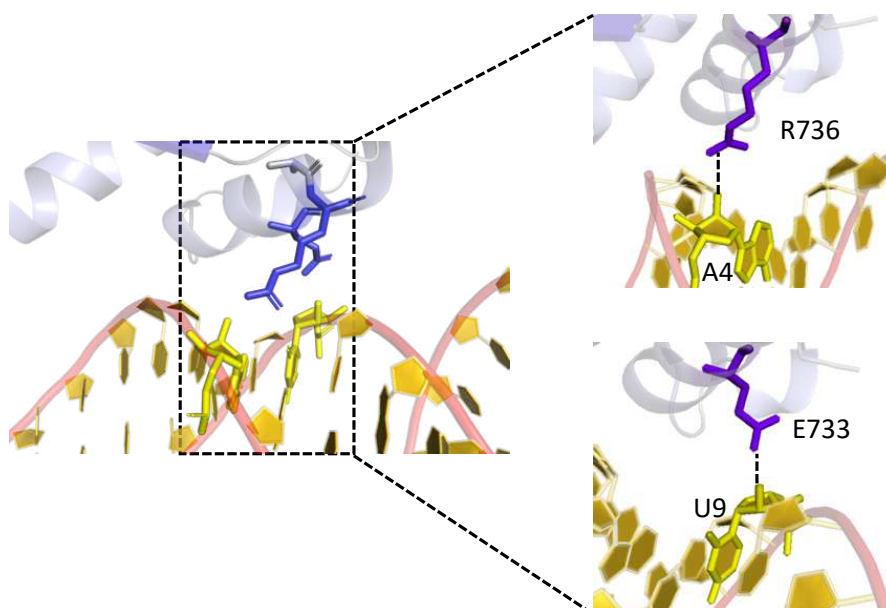


Figure 100 : Représentation de l'interaction canonique ADAR1 Δ NLS/ARN-III au niveau de la région 1.

L'ARN-III est reconnu par ADAR1 Δ NLS *via* des contacts directs impliquant des groupements 2'OH. Les résidus (en bâton) de la protéine impliqués dans l'interaction sont colorés en bleu, les résidus (en bâton) de l'ARN impliqués dans l'interaction sont colorés en jaune.

Ensuite, la boucle L2 reliant les deux premiers brins β du feuillet $\beta_1\beta_2\beta_3$ d'ADAR1 Δ NLS constitue la r gion 2 de la surface d'interaction canonique   l'ARNdb (Figure 99). Cette boucle interagit avec l'ARNdb au niveau du petit sillon adjacent   celui reconnu par la r gion 1. Cette interaction implique deux r sids de cette r gion qui interagissent au niveau du m me brin de l'ARNdb.

Comme au niveau de la r gion 1, un contact direct avec un groupe 2'OH de la base A3 de l'ARN-III est  galement observ  (Figure 101). Ce contact direct implique le r sidu H754 et le groupement 2'OH de la base A3 de l'ARN-III. En plus de ce contact impliquant un groupement 2'OH, un contact dit s quence sp cifique est observ  entre le r sidu P753 et la base G2. Ce contact implique le groupement NH₂ de la base G2 qui forme une liaison hydrog ne avec le groupement carbonyle de la proline P753. Cette interaction est qualifi e de contact s quence sp cifique car la guanine (base G2) est la seule base dans le petit sillon pr sentant un groupement amino NH₂, la formation d'une liaison hydrog ne avec ce groupement permettant de sp cifiquement reconna tre ce nucl oside (163).

Par ailleurs, compar e   la structure d'ADAR1 Δ NLS libre, la boucle L2 pr sente une l g re incurvation vers l'ARN (Figure 98), ce qui lui permet de constituer avec les deux autres r gions, la surface d'interaction avec l'ARNdb.

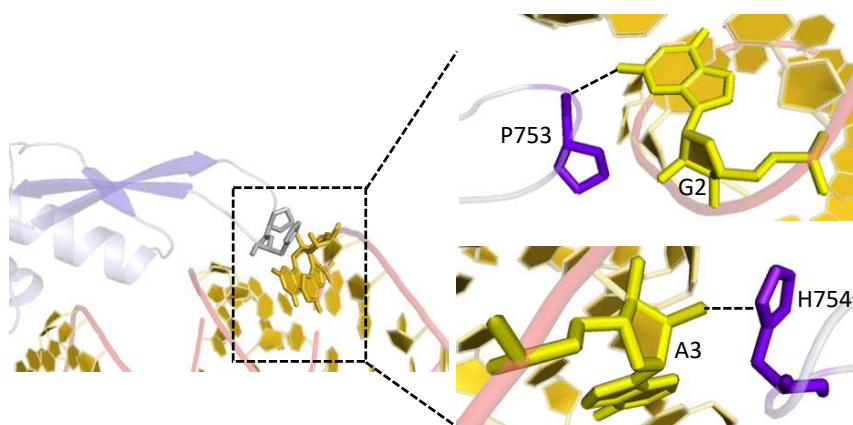


Figure 101 : Repr sentation de l'interaction canonique ADAR1 Δ NLS/ARN-III au niveau de la r gion 2.

L'ARN-III est reconnu par ADAR1 Δ NLS *via* un contact direct impliquant un groupement 2'OH et un contact s quence sp cifique impliquant un groupement NH₂. Les r sids (en  ton) de la prot ine impliqu s dans l'interaction sont color s en bleu, les r sids (en  ton) de l'ARN impliqu s dans l'interaction sont color s en jaune.

Enfin, la boucle L4 reliant le brin β_3   l'h lice α_2 en C-terminale de la prot ine ADAR1 Δ NLS constitue la r gion 3 (Figure 99). Cette r gion reconna tre le grand sillon situ  entre les deux petits sillons, reconnus par les r gions 1 et 2. La r gion 3 poss de un motif riche en lysines tr s

conserv es parmi les dsRBDs (KKQ GK), l'interaction canonique implique g n ralement la reconnaissance de l'ARNdb par les trois lysines de ce motif.

Contrairement aux contacts observ s avec les bases ou les sucres de l'ARN-III au niveau des r gions 1 et 2, les contacts au niveau de la r gion 3 impliquent le squelette sucre-phosphate de l'ARN-III (Figure 102). Le motif riche en lysines (K₇₇₇KQ GK₇₈₁) retrouv  dans cette r gion forme des interactions claires en ce qui concerne les r sidus K₇₇₇ et K₇₇₈ (Figure 102). En effet, ces deux lysines forment des contacts directs avec le squelette sucre-phosphate de chacun des brins de l'ARN-III. Par ailleurs, K₇₈₁ ne pr sente pas d'interaction directe avec le squelette sucre-phosphate. La chaine lat rale de cette lysine est situ e   plus de 9   du squelette de l'ARN-III (Figure 102). La structure semble indiquer que la derni re lysine du motif (K₇₇₇KQ GK₇₈₁) pourrait  tre moins importante pour la reconnaissance des ARNdb que les deux premi res lysines de ce motif. Ceci devra  tre valid  par des mesures d'affinit , afin de d terminer l'implication r elle de ces r sidus (K₇₇₇, K₇₇₈ et K₇₈₁) dans la reconnaissance de l'ARNdb.

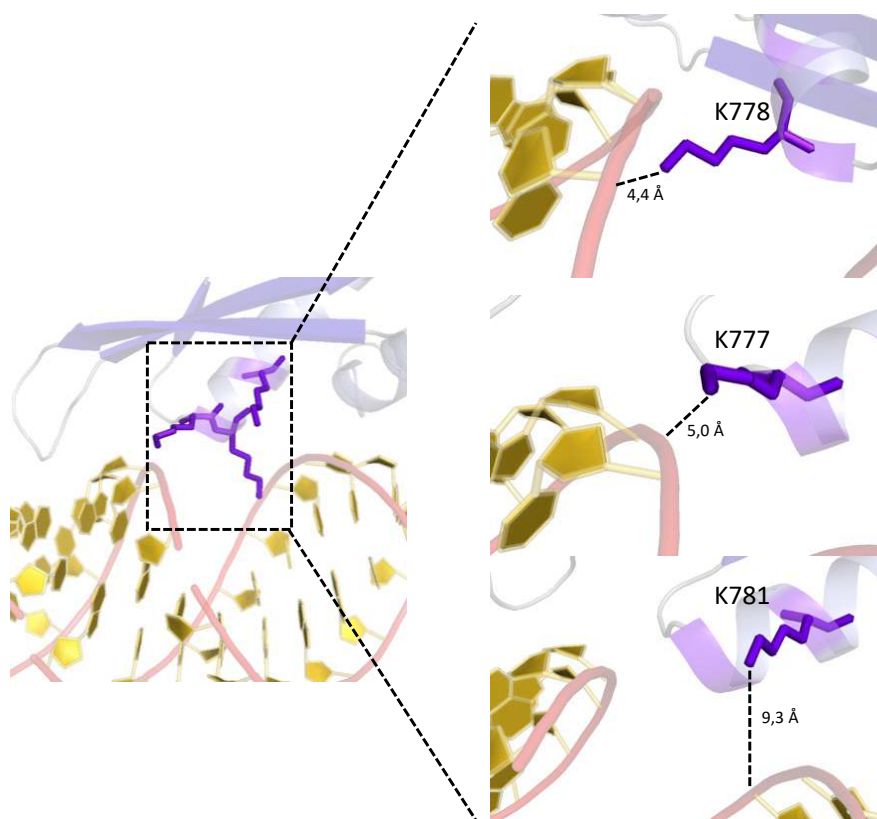


Figure 102 : Repr sentation de l'interaction canonique ADAR1ΔNLS/ARN-III au niveau de la r gion 3.

Le squelette sucre-phosphate de l'ARN-III est reconnu par deux lysines du motif KKQ GK d'ADAR1ΔNLS. Les r sidus (en b ton) de la prot ine impliqu s dans l'interaction sont color s en bleu, le squelette sucre-phosphate (en cartoon) de l'ARN-III est color  en rouge.

3.3.2 Implication de l'hélice α N dans la reconnaissance de l'ARN-III

La structure cristallographique a également montré l'interaction de l'hélice α N d'ADAR1 Δ NLS avec l'ARN-III. Des contacts directs avec des groupements 2' OH des sucres riboses de l'ARN-III sont observés (Figure 103). Le résidu R721 de l'hélice α N reconnaît le groupement 2' OH de U8. De plus, le résidu N726 forme un contact direct avec le groupement 2' OH de C6.

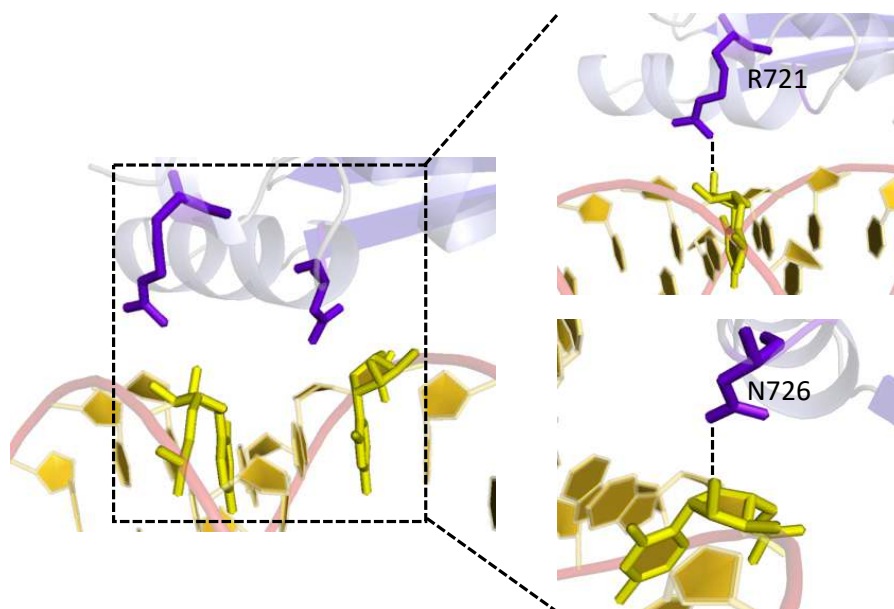


Figure 103 : Représentation de l'interaction ADAR1 Δ NLS/ARN-III impliquant l'hélice α N.

L'ARN-III est reconnu par ADAR1 Δ NLS *via* un contact direct impliquant le groupement 2'OH des sucres. Les résidus (en bâton) de la protéine impliqués dans l'interaction sont colorés en bleu, les résidus de l'ARN (en bâton) impliqués dans l'interaction sont colorés en jaune.

Il est intéressant de noter qu'au sein de l'unité asymétrique, les deux chaînes A et B établissent les mêmes interactions avec la double hélice d'ARN. Bien que les mêmes interactions soient formées par les chaînes A et B, elles présentent un empilement différent au sein du cristal et l'empilement de la chaîne B est plus dense que celui de la chaîne A (Figure 104B).

4 Conclusion

Cette partie portait sur l'étude de l'interaction entre la protéine ADAR1 Δ NLS (dsRBD3 d'ADAR1) et un ARN synthétique double brin. Il s'agissait de résoudre la structure cristallographique du complexe ADAR1 Δ NLS/ARNdb pour ainsi explorer la deuxième hypothèse sur laquelle est basée le modèle de l'inactivation du NLS en présence d'ARN (voir Introduction, paragraphe 3.2.3.1.2, page 55, et Figure 24).

Les premières expériences de cristallisation sur ADAR1 Δ NLS libre ont permis de résoudre la structure cristallographique de la protéine à 1,57 Å. Cette structure a permis de corroborer la topologie classique des dsRBD révélée par la structure d'ADAR1 Δ NLS résolue par RMN. Elle a confirmé de façon indépendante la présence de l'hélice additionnelle α N essentielle pour la construction du NLS bimodulaire d'ADAR1. Des expériences de co-cristallisation ont ensuite été réalisées avec des ARN de séquences « palindromiques ». Des cristaux ont été obtenus suite aux tests de co-cristallisation avec un ARN de 13 nucléotides, l'ARN-III, mais n'ont pas pu être obtenus avec des séquences plus courtes de 11 ou 12 nucléotides. Ceci s'explique probablement par le fait qu'il y aurait un chevauchement et donc une incompatibilité stérique au niveau des sites d'interaction entre l'hélice α 1, qui interagit avec A4 (Figure 100), et la boucle L2 (β 1– β 2) d'un dsRBD symétrique, qui interagit avec les nucléotides directement adjacents A3 et G2 (Figure 101). Les positions des dsRBD sur les hélices d'ARN sont représentées de façon schématisée sur la Figure 104B. C'est l'empilement très compact de la chaîne B qui permet d'expliquer pourquoi des cristaux n'ont probablement pas pu être obtenus avec les ARN plus courts.

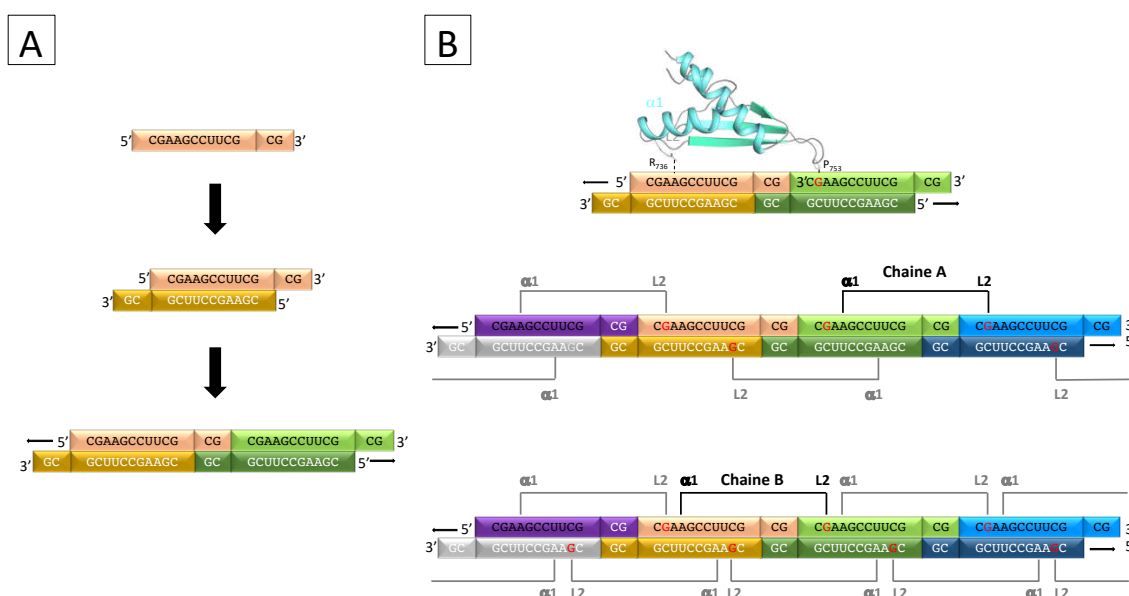


Figure 104 : Schéma de l'assemblage de l'ARN-III et de sa reconnaissance canonique par ADAR1 Δ NLS.

(A) Formation de la double hélice à partir de la séquence « palindromique » de l'ARN-III. La séquence de l'ARN-III s'apparie avec elle-même, et s'auto-assemble en solution pour former une longue hélice d'ARNdb dans le cristal. (B) Schéma de l'interaction d'ADAR1 Δ NLS avec la double hélice d'ARN formée à partir d'une séquence « palindromique » de 13 nucléotides. Ce schéma montre que l'interaction d'ADAR1 Δ NLS avec le nucléotide A4 au niveau de l'hélice α 1 et avec le nucléotide G2 au niveau de la boucle L2 ne permettent pas une reconnaissance d'un ARNdb de séquence plus courte par ADAR1 Δ NLS. En noir est représenté le dsRBD dans l'unité asymétrique, et en gris les symétriques. En rouge, l'interaction séquence spécifique avec la guanosine G2.

Les co-cristaux obtenus avec l'ARN-III ont permis de résoudre la structure cristallographique du complexe ADAR1 Δ NLS/ARN-III à 2,80 Å de résolution. Cette structure a permis de montrer que le dsRBD3 d'ADAR1 reconnaît l'ARNdb de façon canonique sans aucun changement de conformation majeur suite à l'interaction avec l'ARN. Ces résultats ont ainsi permis de valider la deuxième hypothèse sur laquelle est basée le modèle de l'inactivation du NLS en présence d'ARN. La reconnaissance canonique de l'ARNdb par ADAR1 Δ NLS valide l'hypothèse d'une surface d'interaction spécifique de l'ARNdb au niveau d'ADAR1 Δ NLS, et donc que l'interaction simultanée de Trn1 et l'ARNdb avec ADAR1 est impossible à cause d'encombres stériques.

De plus, et de façon tout à fait intéressante, la structure cristallographique de ce complexe suggère que l'hélice α N supplémentaire en N-terminale d'ADAR1 Δ NLS dont la fonction jusqu'ici avait été montrée n'être que structurale, semble être impliquée dans la reconnaissance de l'ARNdb. Ceci devra être validé par des mesures d'affinité entre des mutants d'ADAR1 Δ NLS et d'un ARNdb, afin de déterminer l'implication réelle des résidus R721 et N726 dans la reconnaissance de l'ARNdb.

Conclusion / Perspectives

Conclusion et perspectives

La Transportine 1 (Trn1) est probablement un des récepteurs d'import nucléaire le mieux caractérisé à ce jour après l'importine β . Trn1 est connue pour la reconnaissance d'une classe particulière de NLS, les PY-NLS dont la reconnaissance est bien caractérisée. Cependant, Trn1 reconnaît également des NLS ne faisant pas partie de la classe des PY-NLS. La reconnaissance de ces non-PY-NLS est moins étudiée et peu comprise. De plus, Trn1 est impliquée dans divers mécanismes de régulation, notamment dans la régulation de la reconnaissance de protéines cargos en présence d'ARN. Ceci est le cas pour l'enzyme d'édition d'ARN ADAR1.

L'enzyme ADAR1 est principalement localisée dans le noyau. Cette protéine fait toutefois des navettes entre le noyau et le cytoplasme. En effet, ADAR1 possède d'une part un NLS dépendant de Trn1, ce qui lui permet d'être importée dans le noyau, et d'autre part elle peut quitter le noyau, liée à des molécules d'ARN substrats grâce au récepteur d'export nucléaire Exportine 5. Le NLS porté par ADAR1 se situe dans la région de son 3^e domaine de liaison à l'ARN double brin (dsRBD3). Celui-ci est complètement atypique, car formé de deux modules éloignés dans la séquence de la protéine, mais rapprochés dans une position fonctionnelle par le repliement d'un domaine dsRBD. Le modèle qui existait avant ma thèse, quant à la compréhension du fonctionnement de ce NLS atypique, suggérait que le dsRBD3 d'ADAR1 ne peut être reconnu par Trn1 que lorsqu'il n'est pas lié à des ARN substrats (Figure 105A). Ceux-ci empêcheraient la liaison de Trn1 avec le dsRBD3 en raison d'incompatibilités stériques. Cependant, jusqu'à présent ce modèle ne reposait pas sur des données structurales expérimentales, mais uniquement sur des données de modélisation.

L'objectif de ma thèse était d'obtenir des données structurales concernant deux axes distincts.

Le premier axe de mes travaux de thèse a consisté à caractériser l'interaction entre Trn1 et ADAR1 en utilisant différentes constructions de Trn1 (Trn1 Δ H8 et Trn1FL) et d'ADAR1 (ADAR1court, ADAR1long). A défaut d'avoir pu résoudre la structure cristalline du complexe Trn1/ADAR1, j'ai pu apporter un meilleur éclairage sur le mode de reconnaissance du NLS atypique d'ADAR1 grâce à une combinaison de données de SAXS et de RMN qui nous ont permis de montrer qu'ADAR1 ne présente pas exactement le même mode de reconnaissance que les PY-NLS (Figure 105B). Le deuxième axe de ma thèse a consisté à étudier la régulation de la reconnaissance d'ADAR1 par Trn1 en présence d'ARN. La résolution de la structure

cristallographique du dsRBD3 d'ADAR1 en complexe avec un ARNdb a montré une reconnaissance canonique de l'ARN par le dsRBD3 d'ADAR1 (Figure 105C).

En parallèle de l'étude des mécanismes d'import d'ADAR1, j'ai étudié un ensemble de non-PY-NLS pour une caractérisation de leurs mécanismes de reconnaissance. Cette étude a permis une caractérisation physico-chimique de certains de ces non-PY-NLS.

I. La reconnaissance d'ADAR1 par Trn1

J'ai commencé l'étude de la reconnaissance d'ADAR1 par Trn1 en étudiant le complexe Trn1 Δ H8/ADAR1court. L'étude de ce complexe n'a pas permis de mettre en évidence une interaction entre ces deux partenaires. Cette étude a toutefois conduit à la résolution de la structure cristallographique de Trn1 Δ H8 à 1,81 Å. Cette structure a permis d'observer la flexibilité structurale de la structure en super-hélice composée de HEAT de Trn1 (44).

J'ai ensuite étudié le complexe Trn1 Δ H8/ADAR1long. J'ai pu mettre en évidence l'interaction entre Trn1 et ADAR1 *in vitro*. Cependant, cette étude n'a pas permis d'obtenir de cristaux du complexe et donc la résolution de la structure cristallographique de celui-ci. Des expériences réalisées sur ce complexe ont permis de réaliser l'impact du glycérol dans les différents tampons utilisés, ce qui nous a conduit à l'optimisation des conditions de mélange des protéines pour la préparation du complexe. En effet, une grande concentration (20 % (v/v)) de glycérol était utilisée dans les tampons d'analyse de Trn1 Δ H8 pour favoriser sa solubilité (207). Le glycérol semble avoir un impact négatif sur la conformation des protéines seules (Trn1 Δ H8 et ADAR1long) et donc sur la formation du complexe Trn1 Δ H8/ADAR1long. En supprimant le glycérol dans les tampons utilisés, nous avons non seulement pu mettre en évidence une interaction entre les deux partenaires par des expériences de SEC-MALS mais également améliorer la caractérisation des masses moléculaires des protéines seules.

Finalement, j'ai étudié le complexe Trn1FL/ADAR1long. J'ai pu mettre clairement en évidence l'interaction entre les deux partenaires par des expériences de SEC-MALS. Nous avons ainsi déterminé un K_D de l'ordre d'une centaine de nanomolaire pour cette interaction par des expériences d'ITC. Malheureusement, nous n'avons pas pu produire de cristaux du complexe bien qu'on ait réussi à former le complexe entre les deux partenaires en solution, et obtenir des échantillons compatibles avec une étude cristallographique. Cette dernière étude a toutefois permis, avec la combinaison des données structurales de SAXS et de RMN, d'apporter des

éclaircissements aux données de modélisation sur lesquelles s'appuie la première hypothèse du modèle d'inactivation du NLS en présence d'ARN.

Les données expérimentales obtenues avec les expériences de SAXS ont permis de construire un modèle généré par déformation par modes normaux à partir du modèle construit précédemment (de l'inactivation du NLS en présence d'ARN) (Figure 105B). Ce modèle généré permet de confirmer avec des données expérimentales, que le modèle de l'interaction entre le dsRBD3 d'ADAR1 de l'étude précédente (174) est un modèle plausible de l'interaction même si celui-ci ne rend pas parfaitement compte des données expérimentales obtenues. Néanmoins, ce modèle confirme la position d'ADAR1 au niveau de l'arche C-terminale de Trn1. Les données de RMN montrent bien une interaction entre Trn1FL et ADAR1long mais les signaux des méthyles de Trn1FL impactés par l'interaction n'étant pas encore attribués, ces données n'ont pas permis pour l'instant d'obtenir des informations plus précises sur l'interaction entre ADAR1 et Trn1. Cependant, cette étude a montré l'implication de la boucle H8 dans la reconnaissance d'ADAR1long par Trn1FL. Cette boucle était connue pour participer uniquement à la libération des PY-NLS de Trn1. Toutefois, l'implication de cette boucle a été récemment et pour la première fois mise en évidence dans la reconnaissance d'un autre non- PY-NLS (96). Il semblerait donc que la boucle H8 soit impliquée, au-delà de son rôle initial envisagé, dans la reconnaissance de certains NLS (probablement en majorité des non-PY-NLS). A noter que l'implication de la boucle H8 dans la reconnaissance de cargos est connue pour les karyophérines β . L'importine β reconnaît le facteur de transcription SNAI1 avec la participation de la boucle H8 (208).

L'étude du complexe Trn1FL/ADAR1long a également montré qu'ADAR1long apporte une plus grande affinité d'interaction que la construction ADAR1court, et Trn1 Δ H8 est moins affin que Trn1FL. Ces observations basées d'abord sur des observations biochimiques, ont été confirmées par une caractérisation de l'interaction de ces différents complexes par ITC mais également par des données de titrations en RMN qui montrent une implication de la boucle H8 dans l'interaction des deux partenaires.

Le modèle du complexe Trn1/dsRBD3 généré n'a pas permis de valider complètement la première partie du modèle d'inactivation du NLS en présence d'ARN. Pour construire un meilleur modèle en accord avec les données SAXS, il faudrait obtenir des contraintes précises afin de guider le processus de modélisation. Ces contraintes pourraient être obtenues d'une part par RMN et d'autre part par réticulation chimique associée à la spectrométrie de masse (chemical cross-linking coupled with mass spectrometry, XL-MS). En RMN, l'attribution est

actuellement en cours de réalisation dans l'équipe de nos collaborateurs (Tobias Madl, Medical University of Graz). En XL-MS, des expériences de réticulation intermoléculaire sont menées dans l'équipe de nos collaborateurs (Michael Jantsch, Medical University of Vienna). La réticulation intermoléculaire permettra de caractériser la surface d'interaction du complexe Trn1/ADAR1. La réaction de réticulation conduira à des produits de réticulation intermoléculaire qui permettront de caractériser l'interaction. Les agents réticulants utilisés serviront pour l'identification des résidus en contact lors de la formation du complexe. Une fois l'une ou l'autre de ces approches abouties, ou les deux, la modélisation du complexe Trn1FL/ADAR1long à l'aide de ces nouvelles contraintes et des données de SAXS pourra être de nouveau envisagée afin d'obtenir un modèle de meilleure qualité.

La résolution de la structure cristallographique du complexe Trn1/ADAR1 peut toujours être envisagée. Des expériences sur la formation du complexe Trn1FL/ADAR1long en solution ont permis d'obtenir des échantillons compatibles avec une étude cristallographique, par manque de temps cette étude n'a pas été poursuivie.

En conclusion, une combinaison de différentes techniques structurales a montré qu'ADAR1 ne présente pas le même mode de reconnaissance que les PY-NLS. En effet, ADAR1 interagit probablement au niveau de l'arche C-terminale de Trn1, comme les PY-NLS, mais également avec la boucle acide H8. La surface d'interaction du non-PY-NLS d'ADAR1 est plus grande que celle observée avec les PY-NLS. Ces données à basse résolution, ne permettent pas de valider complètement la première partie du modèle d'inactivation du NLS en présence d'ARN, mais donnent tout de même une indication sur la reconnaissance du NLS d'ADAR1 par Trn1.

II. La reconnaissance d'un ARN double brin par ADAR1

Pour le deuxième axe de mes recherches, j'ai commencé par mener une étude cristallographique du dsRBD3 d'ADAR1. Cette étude a permis de résoudre la structure cristallographique du dsRBD3 d'ADAR1 à 1,57 Å par remplacement moléculaire et de confirmer que ce dsRBD adopte la topologie classique des dsRBD. Elle a également permis de confirmer la présence de l'hélice supplémentaire α_N dont le rôle structural avait été identifié précédemment (174). Ce type d'extension est observé pour différents dsRBD. C'est le cas du dsRBD de Rnt1p (RNase III de *S. cerevisiae*) qui présente une longue hélice α_3 supplémentaire en C-terminal (122, 209), ou encore du dsRBD de la Dcr1 (Dicer de *K. polysporus*) qui présente une petite

hélice $\alpha 3$ supplémentaire en C-terminal (210, 211). Des extensions ont également été observées au niveau de l'extrémité N-terminale pour certains dsRBD, c'est le cas du dsRBD1 de la protéine TRBP (*HIV TAR RNA binding protein*) qui présente une petite hélice $\alpha 0$ en N-terminal (175), ou encore du dsRBD de l'enzyme Dus2 (*Dihydrouridine synthase 2*) qui présente une extension composée d'un tour d'hélice et d'un petit brin β séparés par une boucle (212).

J'ai ensuite effectué des expériences de co-cristallisation de ce dsRBD avec un ARN synthétique de séquence « palindromique » pouvant s'auto-assembler en solution pour former de longues hélices d'ARN double brin (ARNdb) dans le cristal. La résolution de la structure cristallographique du complexe formé à 2,80 Å, a montré une interaction canonique entre les deux partenaires, avec l'implication des trois régions caractéristiques de ce mode de reconnaissance (Figure 105C). Seule une lysine du motif caractéristique d'une de ces trois régions présente une interaction ambiguë du fait de la moyenne résolution de la structure. La structure de ce complexe a permis de montrer une interaction séquence spécifique impliquant la boucle L2 ($\beta 1$ – $\beta 2$) du dsRBD et une base du petit sillon de l'ARNdb. Des interactions séquence spécifiques sont observées dans différentes structures de dsRBD en complexe avec des ARNdb. Dans certains cas, elles impliquent l'hélice $\alpha 1$ du dsRBD et une base du petit sillon de l'ARNdb comme dans la structure du complexe du dsRBD2 de la protéine ADAR2 (168). Dans d'autres cas, ces interactions séquence spécifiques sont retrouvées entre le petit sillon de l'ARNdb et la boucle L2 ($\beta 1$ – $\beta 2$) du dsRBD, comme cela est le cas pour la reconnaissance de l'ARNdb par le dsRBD2 de la protéine XlbpA (166). Les interactions séquence spécifiques sont retrouvées dans des structures de dsRBD avec des ARNdb substrats naturels et non naturels (166, 168, 213).

Par ailleurs, la structure du complexe dsRBD/ARNdb montre que la reconnaissance de l'ARNdb se fait sans réarrangement structural majeur du dsRBD. Ces résultats viennent ainsi confirmer la deuxième hypothèse sur laquelle est basé le modèle d'inactivation du NLS en présence d'ARN, validant ainsi cette partie du modèle (Figure 105C). De manière intéressante, cette structure suggère l'implication de l'hélice αN dans la reconnaissance de l'ARNdb alors que son rôle a été uniquement décrit jusqu'ici comme étant structural. L'hélice αN interagit au niveau du petit sillon également reconnu par l'hélice $\alpha 1$. L'implication d'une extension en N-terminal des dsRBD dans la reconnaissance de l'ARNdb a été auparavant montrée notamment dans l'étude cristallographique du complexe entre le dsRBD de l'enzyme Dus2 et un ARNdb non naturel (197). Cette étude a montré l'implication du demi-tour de l'hélice 3₁₀

dans la reconnaissance de l'ARNdb, au niveau du petit sillon. En revanche, l'analyse structurale du dsRBD1 de la protéine TRBP n'a pas démontré l'implication de l'hélice $\alpha 0$ dans la reconnaissance de l'ARNdb naturel (175). Les extensions des dsRBD en C-terminal ne sont pas impliquées dans la reconnaissance des substrats ARNdb mais plutôt dans des interactions protéines-protéines comme cela est le cas pour l'hélice $\alpha 3$ de la protéine Rnt1p (214, 215) ou pour l'extension en C-terminal de Dcr1 (Dicer de *S. pombe*) (210).

Il sera ainsi nécessaire de réaliser des mutants des résidus impliqués dans l'interaction de l'hélice αN et des lysines impliquées dans l'interaction canonique pour consolider nos données structurales. Ces mutants serviront pour des expériences ITC afin d'analyser l'affinité de l'interaction de ces mutants à un ARNdb. Pour cela, nous envisageons d'utiliser un ARN avec une séquence « palindromique » qui peut s'apparier avec elle-même comme précédemment, mais qui ne s'auto-assemble pas, et qui devra donc être suffisamment long pour permettre la fixation d'un dsRBD sur cet ARNdb.

En conclusion, les études du complexe Trn1/ADAR1 suggèrent une interaction d'ADAR1 au niveau de l'arche C-terminal de Trn1 impliquant la boucle H8. Et, la structure cristalline du complexe dsRBD3/ARNdb montre une interaction canonique entre les deux partenaires et sans réarrangement structural majeur du dsRBD. Toutefois, un meilleur modèle que celui obtenu avec les expériences SAXS est nécessaire pour complètement valider le modèle de l'inactivation du NLS en présence d'ARN.

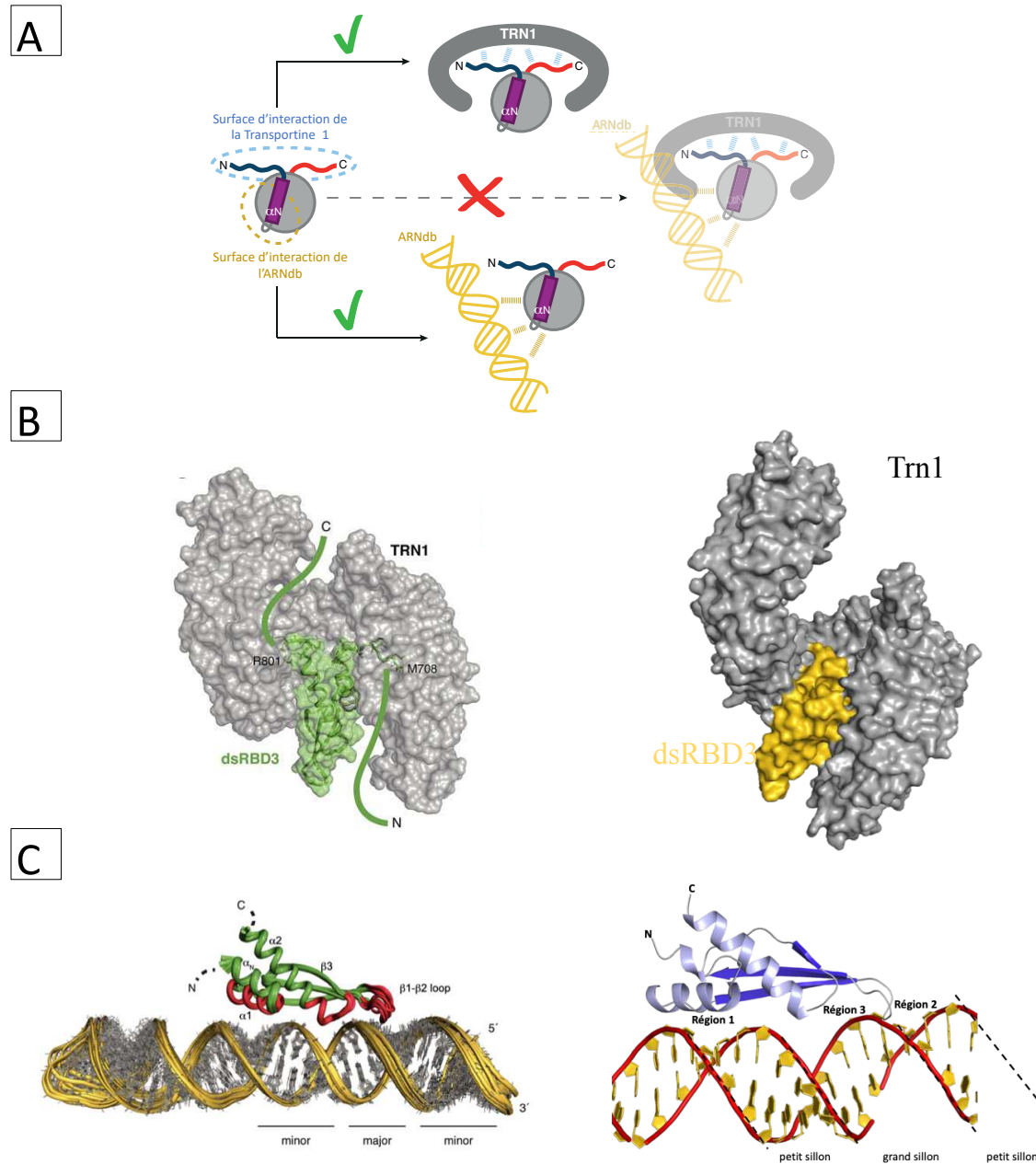


Figure 105 : Schéma récapitulatif des principaux axes étudiés.

(A) Schéma résumant le modèle d'inactivation du NLS d'ADAR1 en présence d'ARN. Ce schéma montre les deux sites d'interactions distincts d'ADAR1 pour Trn1 et l'ARNdb. Ces deux sites ne peuvent être occupés au même moment pour cause d'encombrement stérique. (B) (A gauche) modèle de l'interaction Trn1/ADAR1 proposé à partir de données d'alignement de divers PY-NLS avec le dsRBD3 d'ADAR1. (A droite) modèle de l'interaction Trn1/ADAR1 obtenu pendant ma thèse à partir de données expérimentales (SAXS). (C) (A gauche) modèle de l'interaction dsRBD3 d'ADAR1/ARNdb proposé à partir de la structure du dsRBD2 d'ADAR2 en complexe avec une tige-boucle d'ARN (code PDB 2L3J). (A droite) Structure cristalline du complexe dsRBD3 d'ADAR1/ARNdb obtenue pendant ma thèse.

III. La reconnaissance d'un set de non-PY-NLS par Trn1

J'ai aussi étudié les non-PY-NLS suivant : REV, HPV16, HPV18 et CD44 pour une étude élargie des non-PY-NLS reconnus par Trn1.

Cette étude n'a pas amené à la résolution de structure cristallographique des différents complexes étudiés. Les études biochimiques réalisées sur ces différents complexes ont toutefois mis en évidence une interaction *in vitro* entre les peptides REV et HPV16 avec Trn1ΔH8. L'interaction entre le peptide REV et Trn1ΔH8 a pu être quantifiée et un K_D de l'ordre d'une centaine de nanomolaire a été obtenu. Aucune interaction entre Trn1ΔH8 et le peptide CD44 n'a pu être mise en évidence par deux méthodes biochimiques différentes. Il se pourrait, comme c'est le cas pour le non-PY-NLS d'ADAR1, que le NLS de CD44 interagisse avec la boucle H8 de Trn1 et donc ait une meilleure affinité pour Trn1FL que pour Trn1ΔH8. Il serait intéressant de refaire les tests biochimiques précédemment réalisés avec Trn1FL. Cependant, l'utilisation de peptides synthétiques pour l'étude des différents complexes n'a pas été fructueuse, peut-être parce que les séquences choisies correspondaient aux non-PY-NLS minimaux tels que définis par leur fonctionnalité en tant que NLS *in vivo*. Or, comme on a pu le voir avec le non-PY-NLS d'ADAR1, la séquence minimale du NLS reconnue par Trn1 n'est pas suffisante pour observer une interaction *in vitro* avec Trn1.

Les différentes approches explorées pour résoudre les structures cristallographiques des différents complexes n'ont pas été un succès. Les expériences de co-cristallisation ou encore l'utilisation de protéines chimériques n'ont pas conduit à la résolution de structure cristallographique des différents complexes étudiés. Concernant le complexe Trn1ΔH8/REV, l'utilisation du peptide recombinant a permis de former le complexe en solution et d'obtenir un échantillon compatible avec une étude cristallographique. Ainsi, comme avec l'étude du complexe Trn1FL/ADAR1long, l'étude cristallographique pourrait être approfondie en diversifiant les méthodes de cristallogenèse.

A ce jour, seule la structure de Trn1 avec le non-PY-NLS présent au niveau de la queue de l'histone H3 a été résolue par cristallographie. Cette étude constitue ainsi la référence pour des études structurales concernant les non-PY-NLS (95). Cette structure a montré que ce non-PY-NLS est reconnu au niveau du même site de fixation que les PY-NLS, et cette reconnaissance n'implique pas la boucle H8. Deux des épitopes traditionnels des PY-NLS sont

impliqués dans l'interaction de Trn1. La poche de fixation du dernier épitope correspondant au motif PY des PY-NLS, qui n'est pas présent chez ce non-PY-NLS, reste donc vacante. Une étude récente par RMN, d'un non-PY-NLS avec Trn1 (96) a montré l'implication de la boucle H8 dans la reconnaissance de CIRBP dont le non-PY-NLS est riche en motif RG/RGG. Ces deux études sur les non-PY-NLS montrent déjà une grande variabilité de la reconnaissance de ces non-PY-NLS par Trn1. Il sera donc nécessaire d'obtenir plusieurs et différentes études structurales de Trn1 avec des non-PY-NLS pour pouvoir établir s'il y a, des règles communes de reconnaissance de ces non-PY-NLS par Trn1.

Bibliographie

Bibliographie

1. Görlich, D. and Kutay, U. (1999) Transport Between the Cell Nucleus and the Cytoplasm. *Annu. Rev. Cell Dev. Biol.*, **15**, 607–660.
2. Macara, I.G. (2001) Transport into and out of the Nucleus. *Microbiol Mol Biol Rev*, **65**, 570–594.
3. Conti, E. and Izaurralde, E. (2001) Nucleocytoplasmic transport enters the atomic age. *Curr Opin Cell Biol*, **13**, 310–319.
4. Fried, H. and Kutay, U. (2003) Nucleocytoplasmic transport: taking an inventory. *Cellular and Molecular Life Sciences (CMLS)*, **60**, 1659–1688.
5. Atlanta Cook, Fulvia Bono, Martin Jinek, Elena Conti (2007) Structural biology of nucleocytoplasmic transport. *Annual Review of Biochemistry*, **76**, 647–671.
6. Gall, J.G. (1967) Octogonal nuclear pores. *J Cell Biol*, **32**, 391–399.
7. Tran, E.J. and Wentz, S.R. (2006) Dynamic Nuclear Pore Complexes: Life on the Edge. *Cell*, **125**, 1041–1053.
8. Allegretti, M., Zimmerli, C.E., Rantos, V., Wilfling, F., Ronchi, P., Fung, H.K.H., Lee, C.-W., Hagen, W., Turoňová, B., Karius, K., *et al.* (2020) In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature*, **586**, 796–800.
9. Beck, M. and Hurt, E. (2017) The nuclear pore complex: understanding its function through structural insight. *Nat Rev Mol Cell Biol*, **18**, 73–89.
10. Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B.T., Sali, A. and Rout, M.P. (2004) Components of Coated Vesicles and Nuclear Pore Complexes Share a Common Molecular Architecture. *PLoS Biol*, **2**, e380.
11. Devos, D., Dokudovskaya, S., Williams, R., Alber, F., Eswar, N., Chait, B.T., Rout, M.P. and Sali, A. (2006) Simple fold composition and modular architecture of the nuclear pore complex. *PNAS*, **103**, 2172–2177.
12. Schwartz, T.U. (2005) Modularity within the architecture of the nuclear pore complex. *Curr. Opin. Struct. Biol.*, **15**, 221–226.
13. Maco, B., Fahrenkrog, B., Huang, N.-P. and Aeby, U. (2006) Nuclear pore complex structure and plasticity revealed by electron and atomic force microscopy. *Methods Mol. Biol.*, **322**, 273–288.
14. Strawn, L.A., Shen, T., Shulga, N., Goldfarb, D.S. and Wentz, S.R. (2004) Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat Cell Biol*, **6**, 197–206.
15. Beck, M., Förster, F., Ecker, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W. and Medalia, O. (2004) Nuclear Pore Complex Structure and Dynamics Revealed by Cryoelectron Tomography. *Science*, **306**, 1387–1390.

16. Hampoelz, B., Schwarz, A., Ronchi, P., Bragulat-Teixidor, H., Tischer, C., Gaspar, I., Ephrussi, A., Schwab, Y. and Beck, M. (2019) Nuclear Pores Assemble from Nucleoporin Condensates During Oogenesis. *Cell*, **179**, 671–686.
17. Fu, X., Liang, C., Li, F., Wang, L., Wu, X., Lu, A., Xiao, G. and Zhang, G. (2018) The Rules and Functions of Nucleocytoplasmic Shuttling Proteins. *IJMS*, **19**, 1445.
18. Hülsmann, B.B., Labokha, A.A. and Görlich, D. (2012) The Permeability of Reconstituted Nuclear Pores Provides Direct Evidence for the Selective Phase Model. *Cell*, **150**, 738–751.
19. Ribbeck, K. and Görlich, D. (2001) Kinetic analysis of translocation through nuclear pore complexes. *EMBO J*, **20**, 1320–1330.
20. Rexach, M. and Blobel, G. (1995) Protein import into nuclei: association and dissociation reactions involving transport substrate, transport factors, and nucleoporins. *Cell*, **83**, 683–692.
21. Weis, K. (2007) The Nuclear Pore Complex: Oily Spaghetti or Gummy Bear? *Cell*, **130**, 405–407.
22. Rout, M.P., Aitchison, J.D., Magnasco, M.O. and Chait, B.T. (2003) Virtual gating and nuclear transport: the hole picture. *Trends in Cell Biology*, **13**, 622–628.
23. Lim, R.Y.H., Fahrenkrog, B., Köser, J., Schwarz-Herion, K., Deng, J. and Aeby, U. (2007) Nanomechanical Basis of Selective Gating by the Nuclear Pore Complex. *Science*, **318**, 640–643.
24. Kapinos, L.E., Schoch, R.L., Wagner, R.S., Schleicher, K.D. and Lim, R.Y.H. (2014) Karyopherin-Centric Control of Nuclear Pores Based on Molecular Occupancy and Kinetic Analysis of Multivalent Binding with FG Nucleoporins. *Biophysical Journal*, **106**, 1751–1762.
25. Yamada, J., Phillips, J.L., Patel, S., Goldfien, G., Caestagne-Morelli, A., Huang, H., Reza, R., Acheson, J., Krishnan, V.V., Newsam, S., *et al.* (2010) A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins. *Molecular & Cellular Proteomics*, **9**, 2205–2224.
26. Solmaz, S.R., Blobel, G. and Melčák, I. (2013) Ring cycle for dilating and constricting the nuclear pore. *PNAS*, **110**, 5858–5863.
27. Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117–127.
28. Sorokin, A.V., Kim, E.R. and Ovchinnikov, L.P. (2007) Nucleocytoplasmic transport of proteins. *Biochemistry Mosc.*, **72**, 1439–1457.
29. Görlich, D., Panté, N., Kutay, U., Aeby, U. and Bischoff, F.R. (1996) Identification of different roles for RanGDP and RanGTP in nuclear protein import. *EMBO J*, **15**, 5584–5594.
30. Nakashima, N., Hayashi, N., Noguchi, E. and Nishimoto, T. (1996) Putative GTPase Gtr1p genetically interacts with the RanGTPase cycle in *Saccharomyces cerevisiae*. *J. Cell. Sci.*, **109** (Pt 9), 2311–2318.
31. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. and Wittinghofer, A. (1995) Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature*, **374**, 378–381.

32. Nilsson, J., Weis, K. and Kjems, J. (2002) The C-terminal extension of the small GTPase Ran is essential for defining the GDP-bound form. *J. Mol. Biol.*, **318**, 583–593.
33. Michael J. Matunis, Elias Coutavas, and Giinter Blobel (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J Cell Biol*, **135**, 1457–1470.
34. Mahajan, R., Delphin, C., Guan, T., Gerace, L. and Melchior, F. (1997) A Small Ubiquitin-Related Polypeptide Involved in Targeting RanGAP1 to Nuclear Pore Complex Protein RanBP2. *Cell*, **88**, 97–107.
35. Klebe, C., Prinz, H., Wittinghofer, A. and Goody, R.S. (1995) The kinetic mechanism of Ran--nucleotide exchange catalyzed by RCC1. *Biochemistry*, **34**, 12543–12552.
36. Izaurralde, E. (1997) The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *The EMBO Journal*, **16**, 6535–6547.
37. Dasso, M. (2002) The Ran GTPase: Theme and Variations. *Current Biology*, **12**, R502–R508.
38. Klebe, C., Bischoff, F.R., Ponstingl, H. and Wittinghofer, A. (1995) Interaction of the Nuclear GTP-Binding Protein Ran with Its Regulatory Proteins RCC1 and RanGAP1. *Biochemistry*, **34**, 639–647.
39. Koepp, D.M. and Silver, P.A. (1996) A GTPase controlling nuclear trafficking: running the right way or walking RANdomly? *Cell*, **87**, 1–4.
40. Bischoff, F.R., Klebe, C., Kretschmer, J., Wittinghofer, A. and Ponstingl, H. (1994) RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc Natl Acad Sci U S A*, **91**, 2587–2591.
41. Ribbeck, K., Lipowsky, G., Kent, H.M., Stewart, M. and Görlich, D. (1998) NTF2 mediates nuclear import of Ran. *EMBO J*, **17**, 6587–6598.
42. Mosammaparast, N. and Pemberton, L.F. (2004) Karyopherins: from nuclear-transport mediators to nuclear-function regulators. *Trends in Cell Biology*, **14**, 547–556.
43. Ström, A.-C. and Weis, K. (2001) Importin-beta-like nuclear transport receptors. *Genome Biology*, **2**, reviews3008.1-3008.9.
44. Christie, M., Chang, C.-W., Róna, G., Smith, K.M., Stewart, A.G., Takeda, A.A.S., Fontes, M.R.M., Stewart, M., Vértessy, B.G., Forwood, J.K., *et al.* (2016) Structural Biology and Regulation of Protein Import into the Nucleus. *Journal of Molecular Biology*, **428**, 2060–2090.
45. Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T. and Adam, S.A. (2004) Importin α : a multipurpose nuclear-transport receptor. *Trends in Cell Biology*, **14**, 505–514.
46. Cingolani, G., Petosa, C., Weis, K. and Müller, C.W. (1999) Structure of importin- β bound to the IBB domain of importin- α . *Nature*, **399**, 221–229.
47. Andrade, M.A., Petosa, C., O'Donoghue, S.I., Müller, C.W. and Bork, P. (2001) Comparison of ARM and HEAT protein repeats 11 Edited by P. E. Wright. *Journal of Molecular Biology*, **309**, 1–18.

48. Swale, C., Costa, B. da, Sedano, L., Garzoni, F., McCarthy, A., Berger, I., Bieniossek, C., Ruigrok, R.W.H., Delmas, B. and Crépin, T. (2020) X-ray Structure of the Human Karyopherin RanBP5, an Essential Factor for Influenza Polymerase Nuclear Trafficking. *Journal of Molecular Biology*, **432**, 3353–3359.
49. Yoon, J., Kim, S.J., An, S., Cho, S., Leitner, A., Jung, T., Aebersold, R., Hebert, H., Cho, U.-S. and Song, J.-J. (2018) Integrative Structural Investigation on the Architecture of Human Importin4_Histone H3/H4_Asf1a Complex and Its Histone H3 Tail Binding. *Journal of Molecular Biology*, **430**, 822–841.
50. Conti, E., Müller, C.W. and Stewart, M. (2006) Karyopherin flexibility in nucleocytoplasmic transport. *Current Opinion in Structural Biology*, **16**, 237–244.
51. Chook, Y. (2001) Karyopherins and nuclear import. *Current Opinion in Structural Biology*, **11**, 703–715.
52. O'Reilly, A.J., Dacks, J.B. and Field, M.C. (2011) Evolution of the Karyopherin- β Family of Nucleocytoplasmic Transport Factors; Ancient Origins and Continued Specialization. *PLoS ONE*, **6**, e19308.
53. Chook, Y.M. and Süel, K.E. (2011) Nuclear import by karyopherin- β s: Recognition and inhibition. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1813**, 1593–1606.
54. Kimura, M. and Imamoto, N. (2014) Biological Significance of the Importin- β Family-Dependent Nucleocytoplasmic Transport Pathways: Biological Significance of Importin- β Family. *Traffic*, **15**, 727–748.
55. Xu, D., Farmer, A. and Chook, Y.M. (2010) Recognition of nuclear targeting signals by Karyopherin- β proteins. *Current Opinion in Structural Biology*, **20**, 782–790.
56. Cook, A.G. and Conti, E. (2010) Nuclear export complexes in the frame. *Current Opinion in Structural Biology*, **20**, 247–252.
57. Putker, M., Madl, T., Vos, H.R., de Ruiter, H., Visscher, M., van den Berg, M.C.W., Kaplan, M., Korswagen, H.C., Boelens, R., Vermeulen, M., *et al.* (2013) Redox-Dependent Control of FOXO/DAF-16 by Transportin-1. *Molecular Cell*, **49**, 730–742.
58. Wiese, C., Wilde, A., Moore, M.S., Adam, S.A., Merdes, A. and Zheng, Y. (2001) Role of Importin- β in Coupling Ran to Downstream Targets in Microtubule Assembly. *Science*, **291**, 653–656.
59. Pemberton, L.F. and Paschal, B.M. (2005) Mechanisms of Receptor-Mediated Nuclear Import and Nuclear Export. *Traffic*, **6**, 187–198.
60. Dingwall, C. and Laskey, R.A. (1991) Nuclear targeting sequences--a consensus? *Trends Biochem. Sci.*, **16**, 478–481.
61. Lange, A., McLane, L.M., Mills, R.E., Devine, S.E. and Corbett, A.H. (2010) Expanding the definition of the classical bipartite nuclear localization signal. *Traffic*, **11**, 311–323.
62. Lee, B.J., Cansizoglu, A.E., Süel, K.E., Louis, T.H., Zhang, Z. and Chook, Y.M. (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell*, **126**, 543–558.

63. Wen, W., Meinkoth, J.L., Tsien, R.Y. and Taylor, S.S. (1995) Identification of a signal for rapid export of proteins from the nucleus. *Cell*, **82**, 463–473.
64. Neville, M., Stutz, F., Lee, L., Davis, L.I. and Rosbash, M. (1997) The importin-beta family member Crm1p bridges the interaction between Rev and the nuclear pore complex during nuclear export. *Current Biology*, **7**, 767–775.
65. Dong, X., Biswas, A. and Chook, Y.M. (2009) Structural basis for assembly and disassembly of the CRM1 nuclear export complex. *Nat. Struct. Mol. Biol.*, **16**, 558–560.
66. Lee, S.J., Matsuura, Y., Liu, S.M. and Stewart, M. (2005) Structural basis for nuclear import complex dissociation by RanGTP. *Nature*, **435**, 693–696.
67. Lott, K. and Cingolani, G. (2011) The importin β binding domain as a master regulator of nucleocytoplasmic transport. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1813**, 1578–1592.
68. Palacios, I., Hetzer, M., Adam, S.A. and Mattaj, I.W. (1997) Nuclear import of U snRNPs requires importin β . *The EMBO Journal*, **16**, 6783–6792.
69. Cingolani, G., Bednenko, J., Gillespie, M.T. and Gerace, L. (2002) Molecular basis for the recognition of a nonclassical nuclear localization signal by importin beta. *Mol. Cell*, **10**, 1345–1353.
70. Kutay, U., Bischoff, F.R., Kostka, S., Kraft, R. and Görlich, D. (1997) Export of Importin α from the Nucleus Is Mediated by a Specific Nuclear Transport Factor. *Cell*, **90**, 1061–1071.
71. Pollard, V.W., Michael, W.M., Nakielnny, S., Siomi, M.C., Wang, F. and Dreyfuss, G. (1996) A Novel Receptor-Mediated Nuclear Protein Import Pathway. *Cell*, **86**, 985–994.
72. Nakielnny, S., Siomi, M.C., Siomi, H., Michael, W.M., Pollard, V. and Dreyfuss, G. (1996) Transportin: Nuclear Transport Receptor of a Novel Nuclear Protein Import Pathway. *Experimental Cell Research*, **229**, 261–266.
73. Fridell, R.A., Truant, R., Thorne, L., Benson, R.E. and Cullen, B.R. (1997) Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta. *J. Cell. Sci.*, **110 (Pt 11)**, 1325–1331.
74. Bonifaci, N., Moroianu, J., Radu, A. and Blobel, G. (1997) Karyopherin β 2 mediates nuclear import of a mRNA binding protein. *Proc Natl Acad Sci U S A*, **94**, 5055–5060.
75. Michael, W.M., Choi, M. and Dreyfuss, G. (1995) A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell*, **83**, 415–422.
76. Siomi, M.C., Eder, P.S., Kataoka, N., Wan, L., Liu, Q. and Dreyfuss, G. (1997) Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. *J. Cell Biol.*, **138**, 1181–1192.
77. Imasaki, T., Shimizu, T., Hashimoto, H., Hidaka, Y., Kose, S., Imamoto, N., Yamada, M. and Sato, M. (2007) Structural Basis for Substrate Recognition and Dissociation by Human Transportin 1. *Molecular Cell*, **28**, 57–67.
78. Siomi, H. and Dreyfuss, G. (1995) A nuclear localization domain in the hnRNP A1 protein. *J Cell Biol*, **129**, 551–560.

79. Chook, Y.M. and Blobel, G. (1999) Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature*, **399**, 230–237.
80. Twyffels, L., Gueydan, C. and Kruys, V. (2014) Transportin-1 and Transportin-2: Protein nuclear import and beyond. *FEBS Letters*, **588**, 1857–1868.
81. Cansizoglu, A.E. and Chook, Y.M. (2007) Conformational heterogeneity of karyopherin beta2 is segmental. *Structure*, **15**, 1431–1441.
82. Quan, Y., Ji, Z.-L., Wang, X., Tartakoff, A.M. and Tao, T. (2008) Evolutionary and Transcriptional Analysis of Karyopherin β Superfamily Proteins. *Mol Cell Proteomics*, **7**, 1254–1269.
83. Wozniak, R.W., Rout, M.P. and Aitchison, J.D. (1998) Karyopherins and kissing cousins. *Trends Cell Biol.*, **8**, 184–188.
84. Wang, X., Xu, X., Zhu, S., Xiao, Z., Ma, Z., Li, Y. and Wang, Y. (2012) Molecular dynamics simulation of conformational heterogeneity in transportin 1. *Proteins*, **80**, 382–397.
85. Vetter, I.R., Nowak, C., Nishimoto, T., Kuhlmann, J. and Wittinghofer, A. (1999) Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature*, **398**, 39–46.
86. Chook, Y.M., Jung, A., Rosen, M.K. and Blobel, G. (2002) Uncoupling Kapbeta2 substrate dissociation and ran binding. *Biochemistry*, **41**, 6955–6966.
87. Cansizoglu, A.E., Lee, B.J., Zhang, Z.C., Fontoura, B.M.A. and Chook, Y.M. (2007) Structure-based design of a pathway-specific nuclear import inhibitor. *Nat. Struct. Mol. Biol.*, **14**, 452–454.
88. Soniat, M., Sampathkumar, P., Collett, G., Gizzi, A.S., Banu, R.N., Bhosle, R.C., Chamala, S., Chowdhury, S., Fiser, A., Glenn, A.S., *et al.* (2013) Crystal structure of human Karyopherin $\beta 2$ bound to the PY-NLS of *Saccharomyces cerevisiae* Nab2. *J. Struct. Funct. Genomics*, **14**, 31–35.
89. Zhang, Z.C. and Chook, Y.M. (2012) Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 12017–12021.
90. Yoshizawa, T., Ali, R., Jiou, J., Fung, H.Y.J., Burke, K.A., Kim, S.J., Lin, Y., Peeples, W.B., Saltzberg, D., Soniat, M., *et al.* (2018) Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. *Cell*, **173**, 693-705.e22.
91. Huber, F.M. and Hoelz, A. (2017) Molecular basis for protection of ribosomal protein L4 from cellular degradation. *Nat Commun*, **8**, 77–85.
92. Süel, K.E., Gu, H. and Chook, Y.M. (2008) Modular Organization and Combinatorial Energetics of Proline–Tyrosine Nuclear Localization Signals. *PLOS Biology*, **6**, e137.
93. Arnold, M., Nath, A., Hauber, J. and Kehlenbach, R.H. (2006) Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *J. Biol. Chem.*, **281**, 20883–20890.

94. Jäkel, S., Mingot, J.-M., Schwarzmaier, P., Hartmann, E. and Görlich, D. (2002) Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO J*, **21**, 377–386.
95. Soniat, M. and Chook, Y.M. (2016) Karyopherin- β 2 Recognition of a PY-NLS Variant that Lacks the Proline-Tyrosine Motif. *Structure*, **24**, 1802–1809.
96. Bourgeois, B., Hutten, S., Gottschalk, B., Hofweber, M., Richter, G., Sternat, J., Abou-Ajram, C., Göbl, C., Leitinger, G., Graier, W.F., *et al.* (2020) Nonclassical nuclear localization signals mediate nuclear import of CIRBP. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 8503–8514.
97. Kimura, M., Morinaka, Y., Imai, K., Kose, S., Horton, P. and Imamoto, N. (2017) Extensive cargo identification reveals distinct biological roles of the 12 importin pathways. *eLife*, **6**, e21184.
98. Guo, L., Kim, H.J., Wang, H., Monaghan, J., Freyermuth, F., Sung, J.C., O'Donovan, K., Fare, C.M., Diaz, Z., Singh, N., *et al.* (2018) Nuclear-Import Receptors Reverse Aberrant Phase Transitions of RNA-Binding Proteins with Prion-like Domains. *Cell*, **173**, 677-692.e20.
99. Hock, E.-M., Maniecka, Z., Hruska-Plochan, M., Reber, S., Laferrière, F., Sahadevan M.K., S., Ederle, H., Gittings, L., Pelkmans, L., Dupuis, L., *et al.* (2018) Hypertonic Stress Causes Cytoplasmic Translocation of Neuronal, but Not Astrocytic, FUS due to Impaired Transportin Function. *Cell Reports*, **24**, 987-1000.e7.
100. Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.-D., Simons, M., Niessing, D., Madl, T., *et al.* (2018) Phase Separation of FUS Is Suppressed by Its Nuclear Import Receptor and Arginine Methylation. *Cell*, **173**, 706-719.e13.
101. Qamar, S., Wang, G., Randle, S.J., Ruggeri, F.S., Varela, J.A., Lin, J.Q., Phillips, E.C., Miyashita, A., Williams, D., Ströhl, F., *et al.* (2018) FUS Phase Separation Is Modulated by a Molecular Chaperone and Methylation of Arginine Cation- π Interactions. *Cell*, **173**, 720-734.e15.
102. Eckmann, C.R., Neunteufl, A., Pfaffstetter, L. and Jantsch, M.F. (2001) The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein. *Mol. Biol. Cell*, **12**, 1911–1924.
103. Bass, B.L. (2002) RNA Editing by Adenosine Deaminases That Act on RNA. *Annual Review of Biochemistry*, **71**, 817–846.
104. Savva, Y.A., Rieder, L.E. and Reenan, R.A. (2012) The ADAR protein family. *Genome Biology*, **13**, 252.
105. Nishikura, K. (2010) Functions and Regulation of RNA Editing by ADAR Deaminases. *Annu Rev Biochem*, **79**, 321–349.
106. Brennicke, A., Marchfelder, A. and Binder, S. (1999) RNA editing. *FEMS Microbiology Reviews*, **23**, 297–316.
107. Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H. and Tromp, M.C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria

contains four nucleotides that are not encoded in the DNA. *Cell*, **46**, 819–826.

108. Nishikura, K. (2016) A-to-I editing of coding and non-coding RNAs by ADARs. *Nature Reviews Molecular Cell Biology*, **17**, 83–96.

109. Gott, J.M. and Emeson, R.B. (2000) Functions and mechanisms of RNA editing. *Annu. Rev. Genet.*, **34**, 499–531.

110. Bass, B.L. and Weintraub, H. (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell*, **55**, 1089–1098.

111. Wagner, R.W., Smith, J.E., Cooperman, B.S. and Nishikura, K. (1989) A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and *Xenopus* eggs. *PNAS*, **86**, 2647–2651.

112. Wang, Q., Khillan, J., Gadue, P. and Nishikura, K. (2000) Requirement of the RNA Editing Deaminase ADAR1 Gene for Embryonic Erythropoiesis. *Science*, **290**, 1765–1768.

113. Kim, U., Garner, T.L., Sanford, T., Speicher, D., Murray, J.M. and Nishikura, K. (1994) Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts. *J. Biol. Chem.*, **269**, 13480–13489.

114. Lai, F., Chen, C.X., Carter, K.C. and Nishikura, K. (1997) Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Molecular and Cellular Biology*, **17**, 2413–2424.

115. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H. and Higuchi, M. (1996) A mammalian RNA editing enzyme. *Nature*, **379**, 460–464.

116. Jin, Y., Zhang, W. and Li, Q. (2009) Origins and evolution of ADAR-mediated RNA editing. *IUBMB Life*, **61**, 572–578.

117. Slavov, D., Clark, M. and Gardiner, K. (2000) Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish. *Gene*, **250**, 41–51.

118. Slavov, D., Crnogorac-Jurčević, T., Clark, M. and Gardiner, K. (2000) Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish. *Gene*, **250**, 53–60.

119. Gerber, A.P. and Keller, W. (2001) RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends in Biochemical Sciences*, **26**, 376–384.

120. Keller, W., Wolf, J. and Gerber, A. (1999) Editing of messenger RNA precursors and of tRNAs by adenosine to inosine conversion. *FEBS Letters*, **452**, 71–76.

121. Wolf, J., Gerber, A.P. and Keller, W. (2002) tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *The EMBO Journal*, **21**, 3841–3851.

122. Barraud, P. and Allain, F.H.-T. (2012) ADAR Proteins: Double-stranded RNA and Z-DNA Binding Domains. *Curr Top Microbiol Immunol*, **353**, 35–60.

123. Chen, C.-X., Cho, D.-S.C., Wang, Q., Lai, F., Carter, K.C. and Nishikura, K. (2000) A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA*, **6**, 755–767.

124. Melcher, T., Maas, S., Herb, A., Sprengel, R., Higuchi, M. and Seeburg, P.H. (1996) RED2, a Brain-specific Member of the RNA-specific Adenosine Deaminase Family. *J. Biol. Chem.*, **271**, 31795–31798.
125. Maas, S. and Gommans, W.M. (2009) Identification of a selective nuclear import signal in adenosine deaminases acting on RNA. *Nucleic Acids Res*, **37**, 5822–5829.
126. Fritz, J., Strehblow, A., Taschner, A., Schopoff, S., Pasierbek, P. and Jantsch, M.F. (2009) RNA-Regulated Interaction of Transportin-1 and Exportin-5 with the Double-Stranded RNA-Binding Domain Regulates Nucleocytoplasmic Shuttling of ADAR1. *Molecular and Cellular Biology*, **29**, 1487–1497.
127. Wagner, R.W. and Nishikura, K. (1988) Cell cycle expression of RNA duplex unwindase activity in mammalian cells. *Mol. Cell. Biol.*, **8**, 770–777.
128. Cattaneo, R. (1994) Biased (A→I) hypermutation of animal RNA virus genomes. *Current Opinion in Genetics & Development*, **4**, 895–900.
129. Guidez, F. (2014) Rétrotransposons et cellules somatiques - Parasites ou acteurs actifs du contrôle épigénétique ? *Med Sci (Paris)*, **30**, 659–664.
130. Cordaux, R. and Batzer, M.A. (2009) The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.*, **10**, 691–703.
131. Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhataar, R. and Nishikura, K. (2006) Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol*, **13**, 13–21.
132. Kawahara, Y., Zinshteyn, B., Chendrimada, T.P., Shiekhataar, R. and Nishikura, K. (2007) RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer–TRBP complex. *EMBO Rep*, **8**, 763–769.
133. Nishikura, K., Yoo, C., Kim, U., Murray, J. m., Estes, P. a., Cash, F. e. and Liehaber, S. a. (1991) Substrate specificity of the dsRNA unwinding/modifying activity. *The EMBO Journal*, **10**, 3523–3532.
134. Lehmann, K.A. and Bass, B.L. (1999) The importance of internal loops within RNA substrates of ADAR111Edited by D. E. Draper. *Journal of Molecular Biology*, **291**, 1–13.
135. Polson, A.G. and Bass, B.L. (1994) Preferential selection of adenosines for modification by double-stranded RNA adenosine deaminase. *EMBO J*, **13**, 5701–5711.
136. Wong, S.K., Sato, S. and Lazinski, D.W. (2001) Substrate recognition by ADAR1 and ADAR2. *RNA*, **7**, 846–858.
137. Maas, S., Melcher, T., Herb, A., Seeburg, P.H., Keller, W., Krause, S., Higuchi, M. and O’Connell, M.A. (1996) Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J. Biol. Chem.*, **271**, 12221–12226.
138. Tan, M.H., Li, Q., Shanmugam, R., Piskol, R., Kohler, J., Young, A.N., Liu, K.I., Zhang, R., Ramaswami, G., Ariyoshi, K., *et al.* (2017) Dynamic landscape and regulation of RNA

editing in mammals. *Nature*, **550**, 249–254.

139. Stephens, O.M., Haudenschield, B.L. and Beal, P.A. (2004) The Binding Selectivity of ADAR2's dsRBMs Contributes to RNA-Editing Selectivity. *Chemistry & Biology*, **11**, 1239–1250.

140. Liu, Y., Lei, M. and Samuel, C.E. (2000) Chimeric double-stranded RNA-specific adenosine deaminase ADAR1 proteins reveal functional selectivity of double-stranded RNA-binding domains from ADAR1 and protein kinase PKR. *PNAS*, **97**, 12541–12546.

141. Wang, Y., Park, S. and Beal, P.A. (2018) Selective Recognition of RNA Substrates by ADAR Deaminase Domains. *Biochemistry*, **57**, 1640–1651.

142. Li, J.B., Levanon, E.Y., Yoon, J.-K., Aach, J., Xie, B., LeProust, E., Zhang, K., Gao, Y. and Church, G.M. (2009) Genome-Wide Identification of Human RNA Editing Sites by Parallel DNA Capturing and Sequencing. *Science*, **324**, 1210–1213.

143. Kim, U., Wang, Y., Sanford, T., Zeng, Y. and Nishikura, K. (1994) Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *PNAS*, **91**, 11457–11461.

144. Wang, Y. and Beal, P.A. (2016) Probing RNA recognition by human ADAR2 using a high-throughput mutagenesis method. *Nucleic Acids Res.*, **44**, 9872–9880.

145. Matthews, M.M., Thomas, J.M., Zheng, Y., Tran, K., Phelps, K.J., Scott, A.I., Havel, J., Fisher, A.J. and Beal, P.A. (2016) Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.*, **23**, 426–433.

146. Kawakubo, K. and Samuel, C.E. (2000) Human RNA-specific adenosine deaminase (ADAR1) gene specifies transcripts that initiate from a constitutively active alternative promoter. *Gene*, **258**, 165–172.

147. Yang, J.-H., Luo, X., Nie, Y., Su, Y., Zhao, Q., Kabir, K., Zhang, D. and Rabinovici, R. (2003) Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation. *Immunology*, **109**, 15–23.

148. Poulsen, H., Nilsson, J., Damgaard, C.K., Egebjerg, J. and Kjems, J. (2001) CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain. *Mol. Cell. Biol.*, **21**, 7862–7871.

149. Herbert, A., Alfken, J., Kim, Y.-G., Mian, I.S., Nishikura, K. and Rich, A. (1997) A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *PNAS*, **94**, 8421–8426.

150. Schwartz, T., Rould, M.A., Lowenhaupt, K., Herbert, A. and Rich, A. (1999) Crystal structure of the Z α domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA. *Science*, **284**, 1841–1845.

151. Placido, D., Brown, B.A., Lowenhaupt, K., Rich, A. and Athanasiadis, A. (2007) A Left-Handed RNA Double Helix Bound by the Z α Domain of the RNA-Editing Enzyme ADAR1. *Structure*, **15**, 395–404.

152. Athanasiadis, A., Placido, D., Maas, S., Brown, B.A., Lowenhaupt, K. and Rich, A. (2005) The crystal structure of the Zbeta domain of the RNA-editing enzyme ADAR1 reveals distinct conserved surfaces among Z-domains. *J. Mol. Biol.*, **351**, 496–507.
153. Schwartz, T., Lowenhaupt, K., Kim, Y.-G., Li, L., Brown, B.A., Herbert, A. and Rich, A. (1999) Proteolytic Dissection of Zab, the Z-DNA-binding Domain of Human ADAR1. *J. Biol. Chem.*, **274**, 2899–2906.
154. Koeris, M., Funke, L., Shrestha, J., Rich, A. and Maas, S. (2005) Modulation of ADAR1 editing activity by Z-RNA in vitro. *Nucleic Acids Res.*, **33**, 5362–5370.
155. Bahn, J.H., Ahn, J., Lin, X., Zhang, Q., Lee, J.-H., Civelek, M. and Xiao, X. (2015) Genomic analysis of ADAR1 binding and its involvement in multiple RNA processing pathways. *Nature Communications*, **6**, 6355.
156. Herbert, A. (2019) Z-DNA and Z-RNA in human disease. *Commun Biol*, **2**, 7.
157. Samuel, C.E. (2011) Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology*, **411**, 180–193.
158. Weiden, M.D., Hoshino, S., Levy, D.N., Li, Y., Kumar, R., Burke, S.A., Dawson, R., Hioe, C.E., Borkowsky, W., Rom, W.N., *et al.* (2014) Adenosine Deaminase Acting on RNA-1 (ADAR1) Inhibits HIV-1 Replication in Human Alveolar Macrophages. *PLoS One*, **9**, e108476.
159. Doria, M., Neri, F., Gallo, A., Farace, M.G. and Michienzi, A. (2009) Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Res*, **37**, 5848–5858.
160. Macbeth, M.R., Schubert, H.L., VanDemark, A.P., Lingam, A.T., Hill, C.P. and Bass, B.L. (2005) Inositol Hexakisphosphate Is Bound in the ADAR2 Core and Required for RNA Editing. *Science*, **309**, 1534–1539.
161. Montiel-Gonzalez, M.F., Diaz Quiroz, J.F. and Rosenthal, J.J.C. (2019) Current strategies for Site-Directed RNA Editing using ADARs. *Methods*, **156**, 16–24.
162. Phelps, K.J., Tran, K., Eifler, T., Erickson, A.I., Fisher, A.J. and Beal, P.A. (2015) Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res.*, **43**, 1123–1132.
163. Masliah, G., Barraud, P. and Allain, F.H.-T. (2012) RNA recognition by double-stranded RNA binding domains: a matter of shape and sequence. *Cell. Mol. Life Sci.*, **70**, 1875–1895.
164. Banerjee, S. and Barraud, P. (2014) Functions of double-stranded RNA-binding domains in nucleocytoplasmic transport. *RNA Biology*, **11**, 1226–1232.
165. Gleghorn, M.L. and Maquat, L.E. (2014) ‘Black sheep’ that don’t leave the double-stranded RNA-binding domain fold. *Trends Biochem. Sci.*, **39**, 328–340.
166. Ryter, J.M. and Schultz, S.C. (1998) Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J*, **17**, 7505–7513.
167. Johnston, D.S., Brown, N.H., Gall, J.G. and Jantsch, M. (1992) A conserved double-

stranded RNA-binding domain. *PNAS*, **89**, 10979–10983.

168. Stefl, R., Oberstrass, F.C., Hood, J.L., Jourdan, M., Zimmermann, M., Skrisovska, L., Maris, C., Peng, L., Hofr, C., Emeson, R.B., *et al.* (2010) The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell*, **143**, 225–237.

169. BOHNSACK, M.T., CZAPLINSKI, K. and GÖRLICH, D. (2004) Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA*, **10**, 185–191.

170. Yi, R., Qin, Y., Macara, I.G. and Cullen, B.R. (2003) Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev*, **17**, 3011–3016.

171. Brownawell, A.M. and Macara, I.G. (2002) Exportin-5, a novel karyopherin, mediates nuclear export of double-stranded RNA binding proteins. *J Cell Biol*, **156**, 53–64.

172. Wang, J., Lee, J.E., Riemondy, K., Yu, Y., Marquez, S.M., Lai, E.C. and Yi, R. (2020) XPO5 promotes primary miRNA processing independently of RanGTP. *Nat Commun*, **11**, 1845.

173. Strehlow, A., Hallegger, M. and Jantsch, M.F. (2002) Nucleocytoplasmic Distribution of Human RNA-editing Enzyme ADAR1 Is Modulated by Double-stranded RNA-binding Domains, a Leucine-rich Export Signal, and a Putative Dimerization Domain. *MBoC*, **13**, 3822–3835.

174. Barraud, P., Banerjee, S., Mohamed, W.I., Jantsch, M.F. and Allain, F.H.-T. (2014) A bimodular nuclear localization signal assembled via an extended double-stranded RNA-binding domain acts as an RNA-sensing signal for transportin 1. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, E1852–1861.

175. Masliah, G., Maris, C., König, S.L., Yulikov, M., Aeschmann, F., Malinowska, A.L., Mabillet, J., Weiler, J., Holla, A., Hunziker, J., *et al.* (2018) Structural basis of siRNA recognition by TRBP double-stranded RNA binding domains. *The EMBO Journal*, **37**, e97089.

176. Okada, C., Yamashita, E., Lee, S.J., Shibata, S., Katahira, J., Nakagawa, A., Yoneda, Y. and Tsukihara, T. (2009) A High-Resolution Structure of the Pre-microRNA Nuclear Export Machinery. *Science*, **326**, 1275–1279.

177. Liu, H. and Naismith, J.H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology*, **8**, 91.

178. Cimmerman, P., Baranauskienė, L., Jachimovičiūtė, S., Jachno, J., Torresan, J., Michailovienė, V., Matulienė, J., Sereikaitė, J., Bumelis, V. and Matulis, D. (2008) A Quantitative Model of Thermal Stabilization and Destabilization of Proteins by Ligands. *Biophysical Journal*, **95**, 3222–3231.

179. Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N. and Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.*, **357**, 289–298.

180. Wyatt, P.J. (1993) Light scattering and the absolute characterization of macromolecules.

Analytica Chimica Acta, **272**, 1–40.

181. McPherson, A., Malkin, A. and Kuznetsov, Y. (1995) The science of macromolecular crystallization. *Structure*, **3**, 759–768.

182. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, **26**, 795–800

183. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.

184. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.

185. Matthews, B.W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology*, **33**, 491–497.

186. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Cryst*, **40**, 658–674.

187. Heras, B. and Martin, J.L. (2005) Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. D Biol. Crystallogr.*, **61**, 1173–1180.

188. Heras, B., Edeling, M.A., Byriel, K.A., Jones, A., Raina, S. and Martin, J.L. (2003) Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure*, **11**, 139–145.

189. Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Cryst D*, **60**, 2126–2132.

190. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst D*, **66**, 12–21.

191. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J. and Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst*, **36**, 1277–1282.

192. Petoukhov, M.V. and Svergun, D.I. (2005) Global Rigid Body Modeling of Macromolecular Complexes against Small-Angle Scattering Data. *Biophys J*, **89**, 1237–1250.

193. Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M., *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr*, **50**, 1212–1225.

194. Konarev, P.V. and Svergun, D.I. (2015) A posteriori determination of the useful data range for small-angle scattering experiments on dilute monodisperse systems. *IUCrJ*, **2**, 352–360.

195. Gelis, I., Bonvin, A.M.J.J., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A. and Kalodimos, C.G. (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell*, **131**, 756–769.

196. Dormann, D., Madl, T., Valori, C.F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., Kremmer, E., Ansorge, O., Mackenzie, I.R.A., Neumann, M., *et al.* (2012) Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. *The EMBO Journal*, **31**, 4258–4275.
197. Bou-Nader, C., Barraud, P., Pecqueur, L., Pérez, J., Velours, C., Shepard, W., Fontecave, M., Tisné, C. and Hamdane, D. (2019) Molecular basis for transfer RNA recognition by the double-stranded RNA-binding domain of human dihydrouridine synthase 2. *Nucleic Acids Res*, **47**, 3117–3126.
198. Diederichs, K. and Karplus, P.A. (2013) Better models by discarding data? *Acta Crystallogr D Biol Crystallogr*, **69**, 1215–1222.
199. Suhre, K. and Sanejouand, Y.-H. (2004) Elnémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res*, **32**, W610–W614.
200. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F. and Vagin, A.A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst D*, **67**, 355–367.
201. Langer, G.G., Cohen, S.X., Lamzin, V.S. and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc*, **3**, 1171–1179.
202. Heyam, A., Coupland, C.E., Dégut, C., Haley, R.A., Baxter, N.J., Jakob, L., Aguiar, P.M., Meister, G., Williamson, M.P., Lagos, D., *et al.* (2017) Conserved asymmetry underpins homodimerization of Dicer-associated double-stranded RNA-binding proteins. *Nucleic Acids Res*, **45**, 12577–12584.
203. Svergun, D., Barberato, C. and Koch, M.H.J. (1995) CRY SOL— a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, **28**, 768–773.
204. Panjkovich, A. and Svergun, D.I. (2016) Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys. Chem. Chem. Phys.*, **18**, 5707–5719.
205. Franke, D., Jeffries, C.M. and Svergun, D.I. (2015) Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nature Methods*, **12**, 419–422.
206. Edgcomb, S.P., Aschrafi, A., Kompfner, E., Williamson, J.R., Gerace, L. and Hennig, M. (2008) Protein structure and oligomerization are important for the formation of export-competent HIV-1 Rev–RRE complexes. *Protein Sci*, **17**, 420–430.
207. Crow, M.K., Karasavvas, N. and Sarris, A.H. (2001) Protein aggregation mediated by cysteine oxidation during the stacking phase of discontinuous buffer SDS-PAGE. *Biotechniques*, **30**, 311–316.
208. Choi, S., Yamashita, E., Yasuhara, N., Song, J., Son, S.-Y., Won, Y.H., Hong, H.R., Shin, Y.S., Sekimoto, T., Park, I.Y., *et al.* (2014) Structural basis for the selective nuclear import of

the C2H2 zinc-finger protein Snail by importin β . *Acta Cryst D*, **70**, 1050–1060.

209. Leulliot, N., Quevillon-Cheruel, S., Graille, M., van Tilbeurgh, H., Leeper, T.C., Godin, K.S., Edwards, T.E., Sigurdsson, S.T., Rozenkrants, N., Nagel, R.J., *et al.* (2004) A new α -helical extension promotes RNA binding by the dsRBD of Rnt1p RNase III. *The EMBO Journal*, **23**, 2468–2477.

210. Barraud, P., Emmerth, S., Shimada, Y., Hotz, H.-R., Allain, F.H.-T. and Bühler, M. (2011) An extended dsRBD with a novel zinc-binding motif mediates nuclear retention of fission yeast Dicer. *The EMBO Journal*, **30**, 4223–4235.

211. Weinberg, D.E., Nakanishi, K., Patel, D.J. and Bartel, D.P. (2011) The inside-out mechanism of Dicers from budding yeasts. *Cell*, **146**, 262–276.

212. Bou-Nader, C., Pecqueur, L., Bregeon, D., Kamah, A., Guérineau, V., Golinelli-Pimpaneau, B., Guimarães, B.G., Fontecave, M. and Hamdane, D. (2015) An extended dsRBD is required for post-transcriptional modification in human tRNAs. *Nucleic Acids Research*, **43**, 9446–9456.

213. Ramos, A., Grünert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D. and Varani, G. (2000) RNA recognition by a Staufen double-stranded RNA-binding domain. *The EMBO Journal*, **19**, 997–1009.

214. Wu, H., Henras, A., Chanfreau, G. and Feigon, J. (2004) Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *PNAS*, **101**, 8307–8312.

215. Wang, Z., Hartman, E., Roy, K., Chanfreau, G. and Feigon, J. (2011) Structure of a yeast RNase III dsRBD complex with a noncanonical RNA substrate provides new insights into binding specificity of dsRBDs. *Structure*, **19**, 999–1010.

Annexes

Article



Transportin-1: A Nuclear Import Receptor with Moonlighting Functions

Allegra Mboukou¹, Vinod Rajendra², Renata Kleinova², Carine Tisné¹, Michael F. Jantsch² and Pierre Barraud^{1*}

¹Expression Génétique Microbienne, Institut de Biologie Physico-Chimique (IBPC), UMR 8261, CNRS, Université de Paris, Paris, France, ²Department of Cell and Developmental Biology, Center for Anatomy and Cell Biology, Medical University of Vienna, Vienna, Austria

OPEN ACCESS

Edited by:

Maria Rosaria Conte,
King's College London,
United Kingdom

Reviewed by:

Tobias Madl,
Medical University of Graz, Austria
Roberta Spadaccini,
University of Sannio, Italy

*Correspondence:

Pierre Barraud
pierre.barraud@cnrs.fr

Specialty section:

This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: 05 December 2020

Accepted: 13 January 2021

Published: 18 February 2021

Citation:

Mboukou A, Rajendra V, Kleinova R,
Tisné C, Jantsch MF and Barraud P
(2021) Transportin-1: A Nuclear Import
Receptor with Moonlighting Functions.
Front. Mol. Biosci. 8:638149.
doi: 10.3389/fmolb.2021.638149

Transportin-1 (Trn1), also known as karyopherin- $\beta 2$ (Kap $\beta 2$), is probably the best-characterized nuclear import receptor of the karyopherin- β family after Importin- β , but certain aspects of its functions in cells are still puzzling or are just recently emerging. Since the initial identification of Trn1 as the nuclear import receptor of hnRNP A1 ~25 years ago, several molecular and structural studies have unveiled and refined our understanding of Trn1-mediated nuclear import. In particular, the understanding at a molecular level of the NLS recognition by Trn1 made a decisive step forward with the identification of a new class of NLSs called PY-NLSs, which constitute the best-characterized substrates of Trn1. Besides PY-NLSs, many Trn1 cargoes harbour NLSs that do not resemble the archetypical PY-NLS, which complicates the global understanding of cargo recognition by Trn1. Although PY-NLS recognition is well established and supported by several structures, the recognition of non-PY-NLSs by Trn1 is far less understood, but recent reports have started to shed light on the recognition of this type of NLSs. Aside from its principal and long-established activity as a nuclear import receptor, Trn1 was shown more recently to moonlight outside nuclear import. Trn1 has for instance been caught in participating in virus uncoating, ciliary transport and in modulating the phase separation properties of aggregation-prone proteins. Here, we focus on the structural and functional aspects of Trn1-mediated nuclear import, as well as on the moonlighting activities of Trn1.

Keywords: karyopherin, karyopherin- $\beta 2$, nucleo-cytoplasmic transport, NLS, proline-tyrosine nuclear localization signal, phase-separation, FUS, TNPO1

1 INTRODUCTION

In eukaryotic cells, the presence of a physical separation between the nucleus and the cytoplasm in the form of a double membrane assures a physical and temporal separation of the transcription and translation processes but creates the need for a selective and efficient transport of thousands of macromolecules across the nuclear envelope (Görllich and Kutay, 1999; Conti and Izaurralde, 2001; Fried and Kutay, 2003; Cook et al., 2007). This continual ballet, with controlled bidirectional flows, is orchestrated by nuclear transport receptors (NTRs) that carry their cargoes from one compartment to the other by crossing the nuclear envelope at the level of the nuclear pore complexes (NPCs) (Tran and Wente, 2006; Wente and Rout, 2010; Allegretti et al., 2020). Transport receptors of the karyopherin- β (Kap β) family account for the vast majority of the cargo flow through the NPC. Karyopherins interact selectively with proteins of the NPC, namely the phenylalanine-glycine

nucleoporins (FG-Nups), which surround and line the NPC central channel (Hampoez et al., 2019). Within the NPC, these FG-Nups built and establish the proper operation of the permeability barrier (Fu et al., 2018). In general, large macromolecules (>40 kDa) are indeed excluded from the NPC channel, whereas karyopherins can cross the NPC barrier on account of their common property to selectively interact with FG-Nups (Rout et al., 2003; Weis, 2007; Hülsmann et al., 2012; Beck and Hurt, 2017).

Karyopherins that mediate nuclear import are also known as importins, whereas those mediating nuclear export are known as exportins. Besides the binding to FG-Nups, importins and exportins associate with their cargoes via signals, namely nuclear localization signals (NLSs) and nuclear export signals (NESs), which determine whether the cargo is imported in or exported out of the nucleus (Chook and Süel, 2011; Xu et al., 2010; Cook and Conti, 2010). Importins bind to their cargoes in the cytoplasm, reach the NPCs and translocate to the other side of the nuclear envelope, where they release their cargoes upon binding to the small GTPase Ran in its GTP-bound form (RanGTP). In a reciprocal manner, exportins associated with RanGTP bind to their cargoes in the nucleus, reach the NPCs and translocate to the other side of the nuclear envelope, where they dissociate from their cargoes upon GTP hydrolysis and RanGDP release (Conti and Izaurralde, 2001; Fried and Kutay, 2003; Weis, 2003). Transport directionality is thus primarily driven by the RanGTPase nucleotide cycle, which produces an asymmetric distribution of RanGTP and RanGDP on both sides of the nuclear envelope, with RanGTP being present at high concentrations in the nucleus and with RanGDP being mainly present in the cytoplasm (Izaurralde et al., 1997; Dasso, 2002). Transport selectivity, on the other hand, relies on the selective binding of NTRs to their cargoes via the selective recognition of their signals.

Within the Kap β family, Transportin-1 (Trn1), also known as karyopherin- β 2 (Kap β 2), is probably the best-characterized nuclear import receptor after Importin- β [also known as karyopherin- β 1 (Kap β 1)]. Despite nearly identical structures, Trn1 shows only 24% sequence similarity to Importin- β (Cingolani et al., 1999; Chook and Blobel, 1999). They share a higher similarity at the level of their N-terminal half where the binding of RanGTP takes place, and differ deeply at the level of their C-terminal half, the site of binding to their respective cargoes. Whereas Importin- β binds via the adaptor protein Importin- α to cargoes harbouring the 'classical' NLS (cNLS), Trn1 binds directly to its cargoes without adaptor (Chook and Süel, 2011; Cook et al., 2007; Twyffels et al., 2014). Trn1 has been initially identified as the nuclear import receptor of the heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Pollard et al., 1996; Fridell et al., 1997; Bonifaci et al., 1997). The NLS sequence of hnRNP A1 [designated the 'M9' sequence (Siomi and Dreyfuss, 1995; Weighardt et al., 1995)] can also mediate its nuclear export (Michael et al., 1995). Therefore, considering that the newly identified NTR could both transport hnRNP A1 in and out of the nucleus, this karyopherin was named 'transportin' in opposition to the names 'importins' and 'exportins' in use for unidirectional

transport receptors. Trn1 was in fact later shown to only function in nuclear import, with the hnRNP A1 cargo being released from Trn1 upon RanGTP binding (Siomi et al., 1997; Izaurralde et al., 1997), a common trait of nuclear import receptors, but the terminology has remained.

In the Kap β family, Trn1 shares the highest sequence identity (83%) with Transportin-2 (Trn2), also known as karyopherin- β 2B (Kap β 2B) (Twyffels et al., 2014). Trn2 exists in two isoforms called Trn2A or Kap β 2B(A) (Siomi et al., 1997) and Trn2B or Kap β 2B(B) (Shamsher et al., 2002), the expression of which results from an alternative splicing event (Rebane et al., 2004). This high sequence similarity between Trn1 and Trn2 correlates with a highly similar cargo spectrum for these two NTRs. Among different karyopherins, Trn1 indeed presents the highest degree of functional redundancy with Trn2 in terms of cargo specificity (Mackmull et al., 2017; Kimura et al., 2017). Their cargoes include many RNA-binding proteins involved in mRNA processing, such as proteins of the hnRNP family (e.g., hnRNP A1, hnRNP A2/B1, hnRNP D, hnRNP F, hnRNP H, hnRNP M), but also other protein families implicated in nucleic-acid-related functions in the nucleus. Although the exclusive import of specific cargoes by either Trn1 or Trn2 have not been experimentally confirmed, a high-throughput study suggested a slight specialization of these NTRs. Cargoes with efficient NLSs would be imported by any of the two Transportins, whereas cargoes with less efficient NLSs might be exclusively carried by either Trn1 or Trn2 (Kimura et al., 2017). For instance, actin and actin-related proteins implicated in chromatin remodeling and transcription, and proteins related to nuclear division, would preferentially be imported by Trn1, whereas proteins related to DNA repair would preferentially be Trn2 cargoes (Kimura et al., 2017). These recent studies aiming at defining the cargo spectrum of individual NTRs clearly provide insightful information on NTR-cargo recognition. However, these studies must be analyzed with caution, since among the few hundreds of potential Trn1 cargoes identified in two recent reports, only ~20 are common to the two datasets (Mackmull et al., 2017; Kimura et al., 2017; Baade and Kehlenbach, 2019). The need for experimental validation would therefore be essential to understand in detail the cargo-specificity of Trn1 and to establish the potential specialization of Trn1 and Trn2.

Aside from its principal and long-established activity as a nuclear import receptor, Trn1 was shown more recently to moonlight outside nuclear import. Trn1 has been for instance caught in participating in virus uncoating, cellular cilia formation and in modulating the phase separation properties of aggregation-prone proteins. In this review, we will focus on the structural and functional aspects of Trn1-mediated nuclear import, as well as on the moonlighting activities of Trn1.

2 TRANSPORTIN-1: A NUCLEAR IMPORT RECEPTOR

2.1 Structural Organization of Trn1

Like other Kap β , Trn1 is composed of HEAT-repeats and adopts an overall superhelical architecture (**Figure 1**). HEAT-repeats

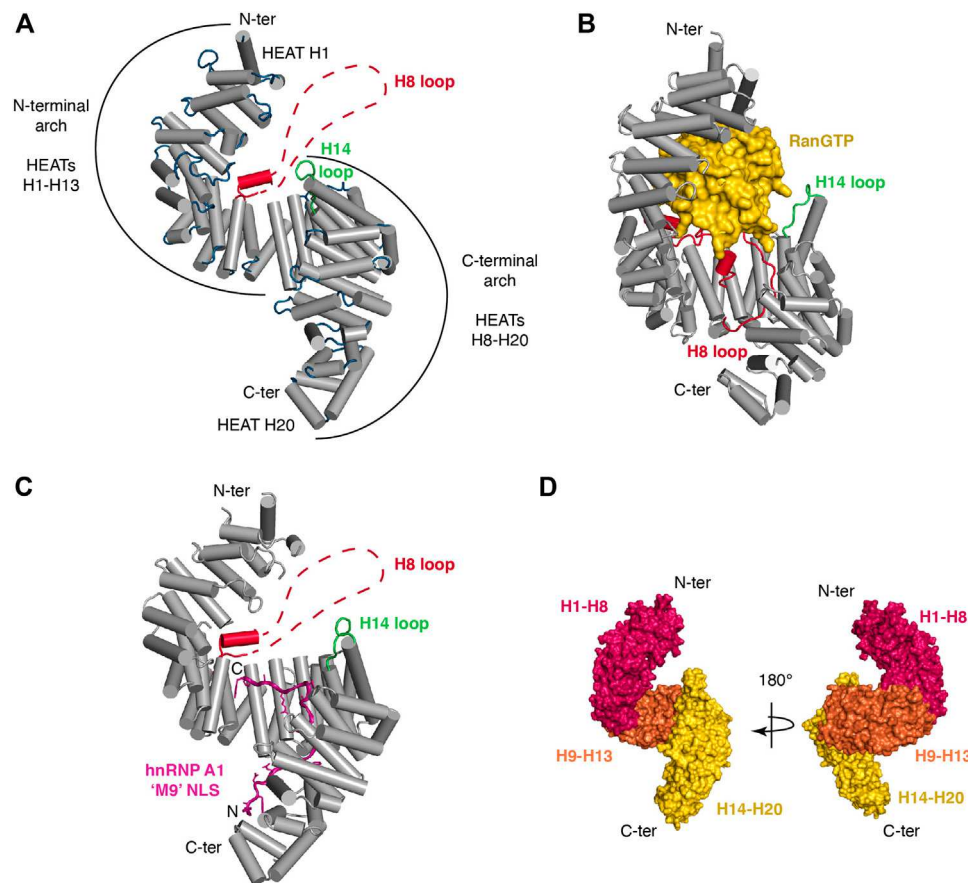


FIGURE 1 | Structural organization of Transportin-1. **(A)** Trn1 is composed of 20 HEAT repeats and adopts a superhelical architecture [free Trn1: PDB code 2QMR (Cansizoglu and Chook, 2007)]. Helices of the HEAT repeats are displayed as cylinders in gray. The loops connecting the different helices are in blue, except for the loop of HEAT H14 that is in green and for the long acidic loop of HEAT H8 (H8 loop) that is in red. Most of the H8 loop is unstructured in the free Trn1 structure, and this long loop is schematically represented with a dashed line. Trn1 structure is overall formed by two consecutive and overlapping arches. The N-terminal arch consists of HEATs H1-H13, and the C-terminal arch is formed by HEATs H8-H20. **(B)** RanGTP (in yellow) associates with Trn1 (in gray) and fits snugly into the N-terminal arch [Trn1-RanGTP complex: PDB code 1QBK (Chook and Blobel, 1999)]. Upon binding, the H8 loop (in red) is reorganized and becomes almost entirely structured. **(C)** The hnRNP A1 'M9' NLS (in purple) associates with Trn1 (in gray) in an extended conformation that covers almost entirely the C-terminal arch [Trn1-NLS complex: PDB code 2H4M (Lee et al., 2006)]. As in free Trn1, most of the H8 loop is unstructured in the NLS-bound structure of Trn1. **(D)** The three major segments identified to describe the conformational flexibility of Trn1 are depicted with three distinct colors, namely HEATs H1-H8 in red, H9-H13 in orange, and H14-H20 in yellow. These segments are displayed on Trn1 surface using two 180°-views to help better grasp the superhelical character of this HEAT-repeat protein.

consist of tandem repeats of a ~40-residue motif (Andrade et al., 2001) that was first identified in Huntingtin, Elongation factor 3, the regulatory A subunit of protein phosphatase 2A and the TOR lipid kinase (Andrade and Bork, 1995). Each HEAT motif folds into a pair of α -helices (known as the A and B helices). Consecutive HEAT motifs pile together in a nearly parallel fashion, which lead to an overall superhelical architecture. The inner concave surface of Trn1 is formed by the B helices, whereas the A helices are exposed at the outer convex surface. The different structures of Trn1, either free or in complex with its different partners, namely RanGTP or various NLSs (see Table 1), all showed that Trn1 is composed of 20 HEAT repeats.

As in other Kap β proteins, the HEAT motifs in Trn1 are connected to each other via small loops or small helices. In addition, within each HEAT motif, helix A is connected to

helix B via a small loop of ~2–4 residues, except at the level of HEAT H14, where this loop is slightly longer and consists of 9 residues, and notably also at the level of HEAT H8, where the two helices of the motif are connected via a long loop of about sixty residues (Figure 1A). This long loop, thereafter called the H8 loop, is intrinsically disordered and has an overall negative charge (27 Asp/Glu over 65 residues). A longer H8 loop is also present in other Kap β , such as Importin- β , but it is significantly longer in Trn1 (65 vs. 12 residues in Importin- β) (Cingolani et al., 1999; Chook and Blobel, 1999; Cook et al., 2007).

Structural and functional considerations have led to define Trn1 as formed by two consecutive and overlapping arches. The N-terminal arch consists of HEATs H1-H13, and the C-terminal arch is formed by HEATs H8-H20 (Figure 1A). The N-terminal arch is the site of interaction with RanGTP (Chook and Blobel, 1999), whereas the C-terminal arch is the site of NLS recognition,

TABLE 1 | Crystal structures of Transportin-1.

Transportin-1	Partner	Type of NLS	PDB code	References
Trn1-FL	RanGTP		1QBK	(Chook and Blobel, 1999)
Trn1-ΔH8	hnRNP A1 NLS ('M9' NLS)	hPY-NLS	2H4M	(Lee et al., 2006)
Trn1-ΔH8	hnRNP M NLS	bPY-NLS	2OT8	(Cansizoglu et al., 2007)
Trn1-FL	–		2QMR	(Cansizoglu and Chook, 2007)
Trn1-FL	–		2Z5J	(Imasaki et al., 2007)
Trn1-FL	TAP/NXF1 NLS	PY-NLS	2Z5K, 2Z5M	(Imasaki et al., 2007)
Trn1-FL	hnRNP D NLS	bPY-NLS	2Z5N	(Imasaki et al., 2007)
Trn1-FL	hnRNP DL/JKTBP NLS	PY-NLS	2Z5O	(Imasaki et al., 2007)
Trn1-ΔH8	FUS NLS	hPY-NLS	4FDD	(Zhang and Chook, 2012)
Trn1-FL	FUS NLS	hPY-NLS	4FQ3	(Niu et al., 2012)
Trn1-ΔH8	ScNab2 NLS	~PY-NLS	4JLQ	(Soniati et al., 2013)
Trn1-ΔH8	HCC1 NLS	PY-NLS	4O06	–
Trn1-ΔH8	histone H3 NLS	non-PY-NLS	5J3V	(Soniati and Chook, 2016)
Trn1-ΔH8	RPL4	PY-NLS	5TQC	(Huber and Hoelz, 2017)
Trn1-ΔH8	FUS NLS	hPY-NLS	5YVG, 5YVH, 5YVI	(Yoshizawa et al., 2018)

Trn1-FL: construct of Trn1 full-length. Trn1-ΔH8: construct of Trn1 with a shortened H8 loop. hPY-NLS: hydrophobic PY-NLS. bPY-NLS: basic PY-NLS. PY-NLS: PY-NLS without a basic or hydrophobic stretch seen to interact with Trn1 in the structure. ~PY-NLS: variation to the PY-NLS motif.

as seen for the 'M9' sequence of hnRNP A1 (Lee et al., 2006) (**Figures 1B,C**).

Conformational flexibility is a hallmark of the Kap β family of proteins (Conti et al., 2006). In Trn1, this intrinsic structural flexibility, which presumably allows Trn1 to adapt its conformation for binding its different partners, namely FG-Nups, RanGTP and various NLSs, has been initially investigated at low-resolution using small-angle X-ray scattering (SAXS) (Fukuhara et al., 2004). Later, with a growing number of Trn1 structures (**Table 1**), either free (Cansizoglu and Chook, 2007), or in complex with RanGTP (Chook and Blobel, 1999) and the 'M9' NLS of hnRNP A1 (Lee et al., 2006), the conformational flexibility of Trn1 has been analyzed at atomic-resolution using structure coordinates and some atomic-refinement parameters (namely the crystallographic B-factors) of free Trn1 (Cansizoglu and Chook, 2007). In addition, the conformational flexibility of Trn1 has been analyzed *in silico* using molecular dynamics simulations (Wang et al., 2012). Together, these studies have shown that the behaviour in terms of conformational flexibility is not uniform along the Trn1 structure, and that substructures or segments can be distinguished for describing the overall conformational flexibility of Trn1. The conformational flexibility of Trn1 has therefore been reported as 'segmental', with three major segments composed of HEATs H1-H8, H9-H13, and H14-H20 (**Figure 1D**). The two latest segments are relatively rigid substructures that rotate around a flexible hinge at the level of H13-H14 helices. The first segment is intrinsically more heterogeneous and can be subdivided into two smaller segments, the segment H1-H4, which is highly prone to conformation changes, and the segment H5-H8, the conformation of which is mostly sensitive to RanGTP binding (Cansizoglu and Chook, 2007; Wang et al., 2012). Within the segment H14-H20, repeats H19-H20 are the most prone to conformational changes.

The conformational flexibility of Trn1 is conferred by the repetition of the HEAT motifs, the packing of which produces a

spring-like effect. In the different structures (**Table 1**), Trn1 is more stretched or more compact, and this occurs mostly at the level of its N- and C-terminal flexible repeats (H1-H4 and H19-H20). This gives an overall structure that can be more or less extended. In the structures and the dynamic simulation, free Trn1 is the one adopting the most elongated conformation, whereas the RanGTP- and the 'M9'-NLS-bound forms are more compact (Cansizoglu and Chook, 2007; Wang et al., 2012). It is worth noting that longer loops connecting helices A and B are found at the level of the three main segment bounds, namely the H8 loop and the H14 loop. These two loops and the surrounding regions constitute dynamic 'hotspots' important for Trn1 function (Wang et al., 2012), namely cargo binding and its release upon binding to RanGTP (see below).

2.2 Trn1-Mediated Import Pathway

The first step of the Trn1 import cycle consists in the recognition and binding of the NLS of a Trn1 cargo in the cytoplasm (**Figure 2**, step 1). NLS recognition by Trn1 will be described in detail in the following subsections, but overall, binding occurs in the C-terminal arch of Trn1, as seen for the 'M9' sequence of hnRNP A1 (Lee et al., 2006) (**Figure 1C**). In the cytoplasm, RanGTP is scarce and binding of Trn1 to its cargoes occurs in its free form.

Then, Trn1 associated with its cargo via the NLS transits towards the NPCs, where its biochemical properties and ability to interact with FG-Nups allows it to pass through the NPC selectivity barrier (**Figure 2**, step 2).

In the nucleus, RanGTP is abundant, and cargo-associated Trn1 binds to RanGTP, which triggers cargo release (**Figure 2**, step 3). RanGTP binds Trn1 at the concave surface of the N-terminal arch, and more precisely at the level of the segments H1-H4 and H7-H8 but also at the level of the long and acidic H8 loop (Chook and Blobel, 1999) (**Figure 1B**). Parts of the 'switch I' and 'switch II' regions of Ran (switch I: residues 30–47; switch II: residues 65–80), which experience large conformational changes during the interconversion between

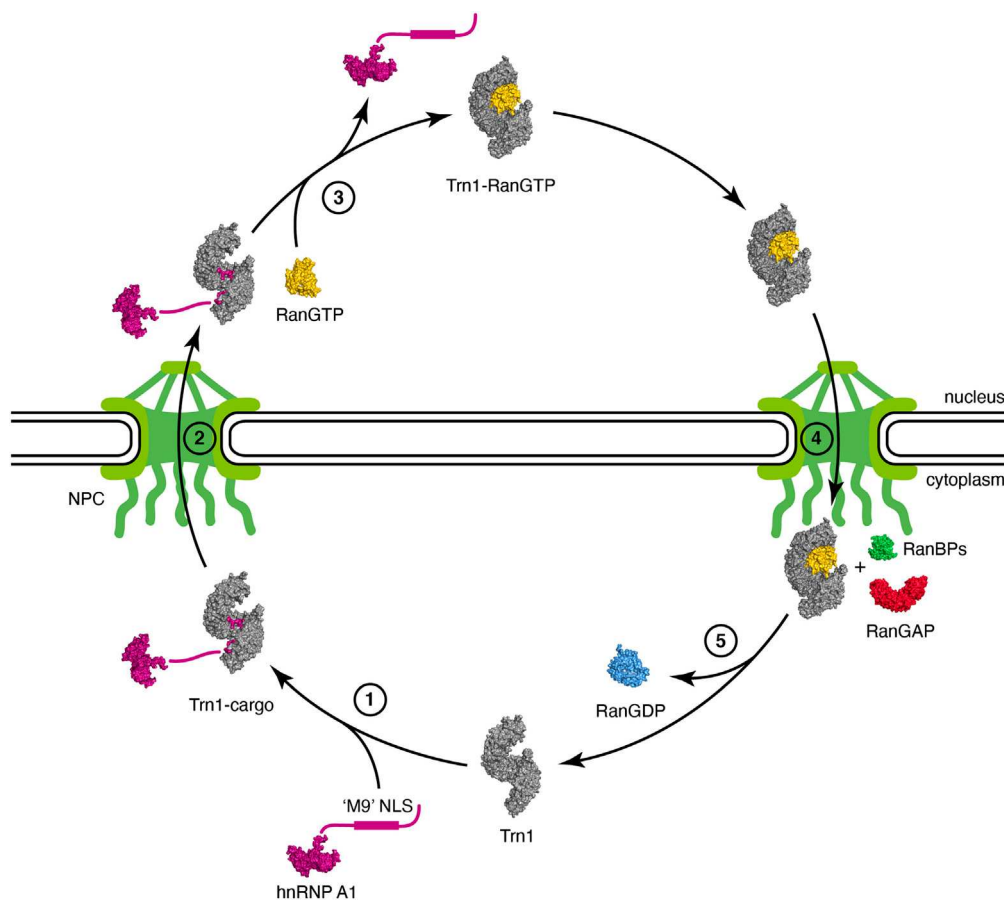


FIGURE 2 | Nuclear import cycle of Transportin-1. (step 1) Trn1 associates with its cargo via the NLS in the cytoplasm [free Trn1 in gray: PDB code 2QMR (Cansizoglu and Chook, 2007); hnRNP A1 cargo with its unstructured 'M9' NLS in purple: PDB code 2LYV (Barraud and Allain, 2013); Trn1-cargo in gray and purple: PDB code 2H4M (Lee et al., 2006)]. (step 2) Trn1 associated with its cargo transits towards and passes through the NPC. The NPC, with the nuclear basket and the cytoplasmic filaments, is schematically represented in green. (step 3) Trn1 associated with its cargo binds to RanGTP in the nucleus, which triggers cargo release [RanGTP in yellow: PDB code 1RRP (Vetter et al., 1999); Trn1-RanGTP in gray and yellow: PDB code 1QBK (Chook and Blobel, 1999)]. (step 4) Trn1 associated with RanGTP transits towards and passes through the NPC. (step 5) At the cytoplasmic face of the NPC, RanGTP is hydrolyzed and released from Trn1 upon the combined action of RanBPs and RanGAP (RanBP1 in green and RanGAP in red: PDB code 1K5D (Seewald et al., 2002); RanGDP in blue: PDB code 1BYU (Stewart et al., 1998)). Trn1 is free again in the cytoplasm and ready for another import cycle.

the different nucleotide states of Ran (Vetter et al., 1999; Scheffzek et al., 1995; Cook et al., 2007), are recognized at the level of the H1-H3 segment of Trn1. Similarly to what is observed in Imp β , the basic patch of Ran (residues 139–142), located opposite to the switch regions of Ran, is involved in electrostatic interactions with HEAT repeat H7 in Trn1, but not directly with the acidic H8 loop (Lee et al., 2005). Although the RanGTP-bound structure of Trn1 is less elongated than the free Trn1 structure, the associated structural change is only minor. In contrast, binding of RanGTP to Trn1 has pronounced and extensive effects on the conformation of the H8 loop (Figures 1A,B). The long H8 loop, which is almost entirely disordered in the free Trn1 structure, becomes indeed well-structured upon interacting with RanGTP. About one third of the H8 loop residues make direct contacts with RanGTP. These contacts include both hydrophobic interactions with apolar side-chains and aromatics, and electrostatic interactions with polar and charged side-chains (Chook and

Blobel, 1999). The part of the H8 loop that does not make direct contact with RanGTP adopts an extended conformation and interacts extensively with the segment H12-H18 in the C-terminal arch (Figure 1B). The H8 loop thereby occupies the NLS interaction sites as observed for the 'M9' sequence of hnRNP A1 (Figure 1C). The interaction of the loop with the NLS-binding site depends on RanGTP, and the displacement and structuring of the H8 loop at the NLS-binding site provides a molecular mechanism to the cargo-dissociation upon RanGTP binding (Lee et al., 2006; Chook and Blobel, 1999). The involvement of the H8 loop in cargo dissociation is also supported by the fact that its removal allows Trn1 to bind the 'M9' NLS and RanGTP simultaneously (Chook et al., 2002). By investigating the binding and dissociation of several NLSs, the mechanism of substrate release from Trn1 has been refined and was reported to occur in a stepwise manner. Upon RanGTP binding, NLS dissociation would first occur at the level of the segment H14-H18, also called 'site B', and complete release of the

cargo would then occur upon dissociation of the NLS from the segment H8-H13, also called 'site A' (Imasaki et al., 2007).

Then, Trn1 associated with RanGTP transits towards the NPCs, and interacts with FG-Nups to pass through the NPC selectivity barrier, similarly to its entry into the nucleus, but in the opposite direction (**Figure 2**, step 4).

When Trn1 bound to RanGTP reaches the cytoplasmic face of the NPC, it associates with several cytoplasmic factors involved in RanGTP hydrolysis and dissociation from the karyopherin, namely the Ran GTPase-activating protein (RanGAP) and the Ran-binding proteins RanBP1 and RanBP2. RanBP2, also known as nucleoporin 358 (Nup358), harbors four Ran-binding domains (RanBD) and is a component of the cytoplasmic filaments of the NPC (Delphin et al., 1997). The intrinsic GTP-hydrolysis by Ran is extremely slow, and RanGAP is needed in the cytoplasm to stimulate the hydrolysis rate by several orders of magnitude (Bischoff et al., 1994; Klebe et al., 1995; Cook et al., 2007). However this enhancement is drastically dampened when RanGTP is associated with karyopherins (Floer and Blobel, 1996; Bischoff and Görlich, 1997; Fried and Kutay, 2003), and RanGAP-mediated stimulation of GTP-hydrolysis necessitates the implication of RanBPs, which destabilize the RanGTP/karyopherin complex. RanGTP hydrolysis therefore occurs at the cytoplasmic face of the NPC and involves the combined action of RanBPs and RanGAP (**Figure 2**, step 5). Upon GTP hydrolysis, Ran switches conformation, both at the level of the 'switch I' and 'switch II' regions, but also at the level of its C-terminal extension that folds as an α -helix and packs against the Ran structural core (Vetter et al., 1999; Scheffzek et al., 1995). Altogether, these structural rearrangements provide a steric barrier that prevents RanGDP from binding back to the karyopherin. After RanGTP release from Trn1 and GTP hydrolysis, Trn1 is again free in the cytoplasm and ready for another import cycle (**Figure 2**, step 1).

3 NUCLEAR LOCALIZATION SIGNAL RECOGNITION BY TRANSPORTIN-1

In the years following the identification of Trn1 as the nuclear import receptor of hnRNP A1 (Pollard et al., 1996; Fridell et al., 1997; Bonifaci et al., 1997), several other proteins (including hnRNP D, hnRNP F, hnRNP M, HuR, Y-box binding protein 1, TAP, and histones H2A, H2B, H3 and H4) have been identified as Trn1 cargoes (Siomi et al., 1997; Fan and Steitz, 1998; Truant et al., 1999; Mühlhäusser et al., 2001; Rebane et al., 2004; Güttinger et al., 2004; Bader and Vogt, 2005; Suzuki et al., 2005). However, the lack of sequence similarity between the different NLSs has for a long time impeded the identification of the important elements defining NLSs recognized by Trn1. The understanding at a molecular level of the NLS recognition by Trn1 made a decisive step forward with the resolution of the first structure of Trn1 in complex with an NLS, namely the 'M9' NLS of hnRNP A1 (**Figure 1C**) (Lee et al., 2006). This work indeed provided the means to identify common patterns among seemingly contrasting NLSs and to group and classify them into a new class of NLSs called PY-NLSs. Following this

pioneering work, several studies have refined the understanding of PY-NLS recognition by Trn1 (see **subsection 3.1**). However, many Trn1 cargoes harbour NLSs that do not resemble PY-NLSs (Chook and Süel, 2011; Twyffels et al., 2014). The recognition of these non-PY-NLSs by Trn1 is until now far less understood than that of PY-NLSs. Whether non-PY-NLSs can be regarded as variations around the PY-NLS paradigm, or whether they are too divergent and use completely different recognition rules, has not been clearly established yet, mostly by lack of studies, including structural ones, focused on non-PY-NLSs (see **subsection 3.2**). Finally, over the years, a salient aspect of Trn1-mediated nuclear import has emerged and consists in the many reported mechanisms that modulate the karyopherin-cargo interactions, thereby enabling nuclear import regulation (see **subsection 3.3**).

3.1 PY-NLS Recognition

Unlike the classical-NLSs recognized by the Imp α /Imp β system, PY-NLSs cannot be described solely by a traditional consensus sequence. They are instead described by a collection of loose principles that can be summarized as follows: a peptide segment of 15–30 residues with intrinsic structural disorder, an overall basic character, and some weakly conserved sequence motifs, including a relatively conserved proline–tyrosine (PY) dipeptide, which gave the name PY-NLS to this class of signals (Lee et al., 2006). More precisely, the sequence motifs of PY-NLSs are composed of an N-terminal basic or hydrophobic motif followed by a C-terminal R/K/H-X₂₋₅-P-Y motif (Lee et al., 2006). The nature of the N-terminal motif further divides PY-NLSs into basic and hydrophobic subclasses (bPY-NLSs and hPY-NLSs, respectively). In bPY-NLSs, the N-terminal motif consists of a patch of basic residues, whereas in hPY-NLSs, the equivalent N-terminal motif conforms to the loose Φ -G/A/S- Φ consensus (where Φ is a hydrophobic residue).

To date, ~10 distinct crystal structures of Trn1 have been solved in complex with PY-NLS substrates, including both bPY-NLSs and hPY-NLSs (**Table 1**). Collectively, these structures show that all PY-NLSs interact with the same region of Trn1, namely an extended zone of the concave face of the C-terminal arch (**Figure 3A**). However, considering the poor resemblance of the NLSs in term of sequence, it is not surprising to see that structural superposition of the different PY-NLSs is really far from perfect (**Figure 3B**). PY-NLSs vary both in length and sequence, and their binding mode to Trn1 may at first sight look very disparate. However, structural convergence clearly appears at two sites in the C-terminal part of the NLSs, and to a lesser extent at one site in the N-terminal part. These three sites of structural convergence form three binding epitopes, which have been initially identified by the structural comparison of the binding mode of hnRNP A1 and hnRNP M PY-NLSs (Cansizoglu et al., 2007), and were later shown to provide a significant energetic contribution to the binding (Süel et al., 2008). Epitope 1, which shows the least structural convergence and the highest sequence divergence, corresponds to the N-terminal motif presented above and can be formed by a patch of either basic or hydrophobic residues. Epitope 2 and 3 correspond to the positively charged residue and to the PY

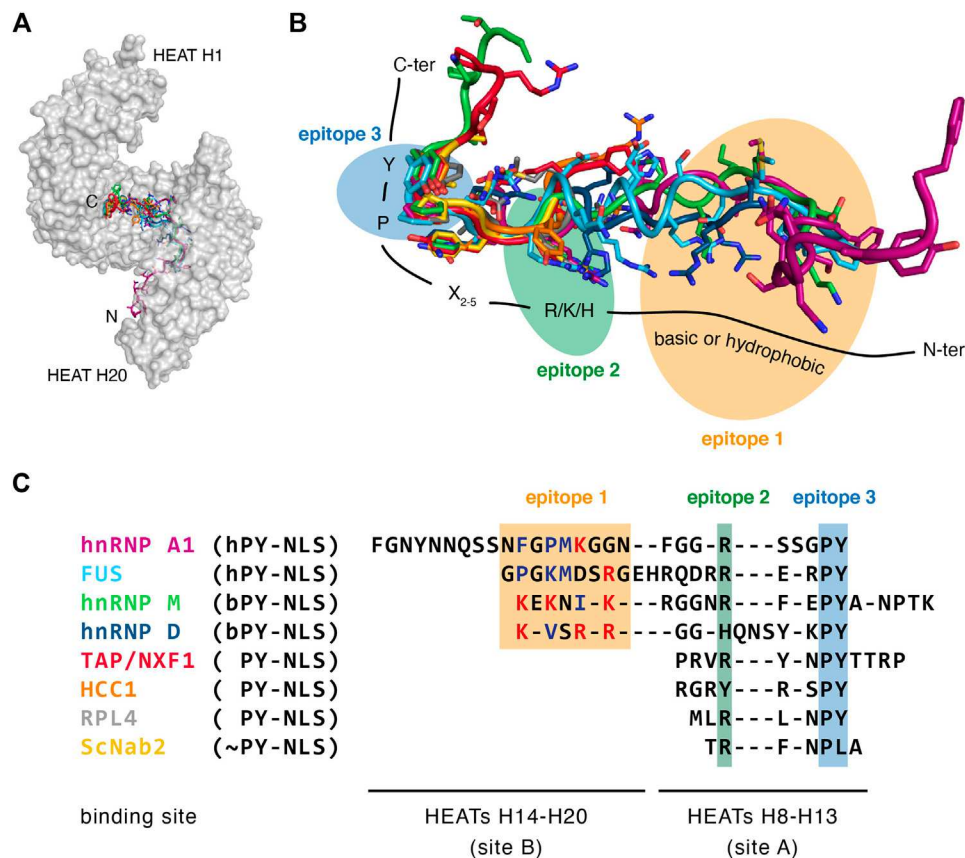


FIGURE 3 | PY-NLS recognition by Transportin-1. **(A)** A common binding site for PY-NLSs with the C-terminal arch of Trn1. Trn1 is shown in gray as a transparent surface to facilitate PY-NLS observation. The PY-NLSs are displayed with different colors (see below), and all share the same N-ter to C-ter orientation. **(B)** Structural superposition of PY-NLSs from the following proteins: hnRNP A1 in purple [PDB code 2H4M (Lee et al., 2006)], FUS in light blue [PDB code 4FDD (Zhang and Chook, 2012)], hnRNP M in green [PDB code 2OT8 (Cansizoglu et al., 2007)], hnRNP D in dark blue [PDB code 2Z5N (Imasaki et al., 2007)], TAP/NXF1 in red [PDB code 2Z5K (Imasaki et al., 2007)], HCC1 in orange [PDB code 4OO6], RPL4 in gray [PDB code 5TQC (Huber and Hoelz, 2017)], ScNab2 in yellow [PDB code 4JLQ (Soniat et al., 2013)]. The structures were superimposed on the Trn1 proteins, but for clarity they are not shown, and only the PY-NLSs are displayed as sticks. The three recognition epitopes are shown with colored ellipses. Epitope 1, which corresponds to a stretch of basic or hydrophobic residues, is highlighted in yellow. Epitope 2, which corresponds to a positively charged residue (R/K/H), is highlighted in green. Epitope 3, which corresponds to a proline-tyrosine dipeptide (PY) is highlighted in blue. **(C)** Structure-based alignment of the same PY-NLSs displayed on panel B. The PY-NLS sequences were aligned according to their relative position in the structures. Only residues actually observed in the structures are displayed. The three epitopes as described in panel B are highlighted with the same color code. Within epitope 1, hydrophobic residues are colored in blue, and positively charged residues are colored in red. hPY-NLS: hydrophobic PY-NLS. bPY-NLS: basic PY-NLS. PY-NLS: PY-NLS without a basic or hydrophobic stretch seen to interact with Trn1 in the structure. ~PY-NLS: variation to the PY-NLS motif.

dipeptide of the C-terminal R/K/H-X₂₋₅-P-Y motif, respectively (Figure 3B,C).

What emerges from such structural superposition, is that there are clearly two types of PY-NLSs: the ones that are seen to interact with Trn1 all along the three epitopes (e.g., hnRNP A1, FUS, hnRNP M and hnRNP D) and those that are only interacting at epitopes 2 and 3 (e.g., TAP/NXF1, HCC1, RPL4 and ScNab2). To help describe these two categories of PY-NLSs, and in connection with the structural heterogeneity and the segmental nature of Trn1 (Figure 1D), two main NLS-binding sites have been defined within the Trn1 C-terminal arch. The first one, called 'site A', corresponds to HEATs H8-H13, and represents the site of interaction of epitopes 2 and 3 (Figure 3C). The second one, called 'site B', corresponds to HEATs H14-H20, and is the site of interaction of epitope 1 (Figure 3C). PY-NLSs are thus

interacting at both site A and site B, or uniquely at site A for the ones only interacting at the level of epitopes 2 and 3 (Imasaki et al., 2007). Since the conformation of site A is independent of NLS binding, whereas that of site B is intrinsically more adaptable, it has been proposed that the C-terminal R/K/H-X₂₋₅-P-Y motif would bind first to site A, then a conformational change of site B would occur owing to a rearrangement at the level of the H13-H14 hinge, which with a sort of induced fit mechanism would adapt site B to epitope 1 binding (Imasaki et al., 2007). As previously mentioned, substrate dissociation upon RanGTP binding is likely occurring in the reverse order, with PY-NLSs dissociating first from site B, and then from site A (Imasaki et al., 2007). Molecular mechanisms governing proper functioning of Trn1 (e.g., substrate binding and release) are therefore deeply linked to its structural adaptability.

It is important to mention that although the structures of PY-NLSs in complex with Trn1 do not always show an interaction at the level of site B (**Figures 3B,C**), residues upstream of the R/K/H-X₂₋₅-P-Y motif seen to interact at site A can significantly contribute to the overall binding (Imasaki et al., 2007; Süel et al., 2008). For that reason, PY-NLSs shown to only interact at site A may still be classified as bPY-NLSs or hPY-NLSs depending on their amino-acid composition (Lee et al., 2006; Twyffels et al., 2014). Most importantly, the binding site accommodating epitope 1 seems to be highly tolerant, since it can bind to both basic and hydrophobic patches, and a combination of both basic and hydrophobic residues is often observed in the N-terminal motif (**Figure 3C**) (Lee et al., 2006). Overall, the distribution of the binding energy along the three binding epitopes highly depends on the PY-NLS considered. In some cases, the contribution of epitope 3 is pre-eminent, whereas in others main binding contributions are provided by epitopes 1 and 2 (Süel et al., 2008; Lee et al., 2006; Imasaki et al., 2007; Cansizoglu et al., 2007). This property, together with the fact that the epitopes are energetically quasi-independent, had been proposed as responsible for the high adaptability of the epitopes, which can accommodate large sequence diversity on condition that the others are energetically strong. A combinatorial mixing of energetically weak and strong motifs indeed maintains an overall affinity compatible with nuclear import (Süel et al., 2008).

3.2 Non-PY-NLS Recognition

In addition to the import of PY-NLS-containing cargoes, Trn1 is implicated in the import of various cargoes obviously lacking a PY-NLS (Twyffels et al., 2014; Chook and Süel, 2011). These cargoes are imported by Trn1 via non-PY-NLSs, and include diverse proteins such as the core histones (H2A, H2B, H3 and H4) (Mühlhäusser et al., 2001; Baake et al., 2001; Mosammamarast et al., 2002; Blackwell et al., 2007; Soniat et al., 2016), ribosomal proteins (RPL23A, RPL5, RPL7 and RPS7) (Jäkel and Görlich, 1998; Tai et al., 2013), viral proteins (Arnold et al., 2006; Le Roux and Moroianu, 2003; Klucsevsek et al., 2006), the RNA-editing enzyme ADAR1 (Fritz et al., 2009; Barraud et al., 2014; Banerjee and Barraud, 2014), the transcription factor FOXO4 (Putker et al., 2013), and the cold-inducible RNA-binding protein CIRBP (Bourgeois et al., 2020). With some exceptions, PY-NLS seem to be specific to Trn1, while non-PY-NLS can be imported by Trn1 but can frequently be recognized and imported by multiple karyopherins such as the heterodimer Impα/Impβ (Kimura et al., 2013; Twyffels et al., 2014).

Structural understanding of non-PY-NLS recognition by Trn1 is quite limited, since only a single structure of Trn1 in complex with a non-PY-NLS peptide has been solved to date (Soniat and Chook, 2016). In this structure, the non-PY-NLS of histone H3 spans HEATs H11-H18 in the concave site of the Trn1 C-terminal arch. It occupies similar positions as the ones of PY-NLSs bound to Trn1, with the exception that there is no PY dipeptide or other residues occupying the epitope 3 recognition site of PY-NLSs (**Figure 3**). A positively charged stretch binds the site for PY-NLS epitope 1, and an arginine residue occupies the PY-NLS epitope 2 position (Soniat and Chook, 2016). The binding mode is therefore somewhat similar to the one of PY-

NLSs, but uses a strong epitope 1 that compensates for the loss of the PY motif. The non-PY-NLS of histone H3 can thus be seen as an ultimate variation of the PY-NLS, lacking the PY motif, but retaining the other hallmarks of this class of NLSs. Whether this aspect is unique to histone H3, or is shared by other non-PY-NLSs of Trn1, is a likely possibility, but remains to be determined by further studies, including structural works dealing with other non-PY-NLSs.

Although no other crystal structure of non-PY-NLSs with Trn1 has been obtained, molecular information on non-PY-NLS interactions with Trn1 are available for several systems. These include the FOXO4 (forkhead box O4) protein, which binds Trn1 using a very unusual mechanism (Putker et al., 2013). Upon accumulation of reactive oxidative species (ROS), a disulfide bond is formed between a cysteine residue of FOXO4 and a cysteine residue in Trn1. The formation of a covalent complex ensures a strong interaction with Trn1 and an efficient nuclear import. In the nucleus, the more reducing environment facilitates the reduction of the intermolecular disulfide bond and the release of the FOXO4 transcription factor, which eventually activates transcription of ROS-detoxifying enzymes. Neither the interaction site in Trn1, nor the implicated cysteine, have so far been identified. Residues surrounding the implicated cysteine in FOXO4 have been shown to participate in the binding to Trn1 (Putker et al., 2013), but it remains to be determined whether this interaction occurs specifically at a single site on Trn1 surface or at several positions. Among other things, it would be interesting to know whether the present binding site is completely different from the one used by PY-NLSs or if they have some parts in common.

A non-PY-NLS for which molecular details of its recognition by Trn1 is available consists of the atypical NLS of the RNA editing enzyme ADAR1 (Eckmann et al., 2001; Strehlow et al., 2002; Fritz et al., 2009). This NLS overlaps the third double-stranded RNA-binding domain (dsRBD) of the protein (Barraud and Allain, 2012), and shows no similarity to PY-NLSs. The molecular basis for the dsRBD-mediated nuclear import of ADAR1 was investigated at a molecular level. The solution structure of the ADAR1-dsRBD3 revealed an extended dsRBD fold with an additional α -helix in the N-terminus. This extension radically changes the relative position of the flexible fragments flanking the dsRBD and brings the N- and C-terminal flanking regions in close proximity (Barraud et al., 2014). The two fragments flanking the folded dsRBD were shown to constitute two essential modules involved in the interaction with Trn1 and the non-PY-NLS of ADAR1 was thus called 'bimodular NLS'. The intervening dsRBD was shown to only act as a scaffolding domain, which properly positions the N- and C-terminal modules for an effective interaction with Trn1. Functional bimodular NLSs could indeed be designed by replacing the ADAR1-dsRBD3 with an unrelated dsRBD, or even with a small peptide linker, which clearly indicates that the dsRBD only helps bring together the two NLS-modules that are otherwise distantly spaced in the protein sequence (Barraud et al., 2014). Molecular modelling and functional assays involving Trn1 mutants affected in the regular PY-NLS binding sites (**Figure 3**), namely 'site A' and 'site B', suggested

that the bimodular NLS of ADAR1 may interact with Trn1 at the same interaction sites as the ones of PY-NLSs. The folded dsRBD would be small enough to insert into the C-terminal arch of Trn1, and the N- and C-terminal modules of the non-PY-NLS of ADAR1 could adopt an extended conformation that allows interaction with Trn1 at the position of epitopes 1–3, even though they do not contain a PY dipeptide (Barraud et al., 2014). However, the atomic details of the interaction still remain to be uncovered, which would expand our understanding of the repertoire of non-PY-NLS recognition by Trn1.

In another recent report, a non-PY-NLS has been identified in the cold-inducible RNA-binding protein CIRBP and was shown to participate in the Trn1-mediated nuclear import of CIRBP (Bourgeois et al., 2020). This non-PY-NLS corresponds to a ~40-residue region rich in RG and RGG motifs, called the RG/RGG region. This RG/RGG region binds to Trn1 in a RanGTP-competitive manner and does not contain PY or PΦ motifs (Φ: hydrophobic residue) that may be reminiscent of PY-NLSs. The most striking point regarding the RG/RGG region interaction with Trn1 consists in the involvement of the Trn1 H8 loop in the interaction (Bourgeois et al., 2020). The RG/RGG region of CIRBP was indeed shown to contact Trn1 at two key sites, a site competing with the binding of the PY-NLS of FUS, and a non-overlapping site within the unstructured H8 loop of Trn1. The importance of this additional site is reflected in the fact that deletion of the H8 loop reduces the binding affinity of the RG/RGG region to Trn1 by ~5 fold (Bourgeois et al., 2020). As presented above, although the H8 loop has been shown to be essential for cargo release upon RanGTP binding, it was reported as dispensable for NLS binding, and was indeed deleted in most structures of Trn1 in complex with PY-NLSs (Table 1). This study concerning the RG/RGG NLS of CIRBP raises the possibility that the Trn1 H8 loop might be important for NLS binding, at least for certain non-PY-NLSs (Bourgeois et al., 2020). To determine whether the disordered H8 loop is important for the recognition of other non-PY-NLSs is definitely a critical question that would necessitate further studies.

3.3 Regulation of Trn1-Cargo Interactions

Several mechanisms modulating Trn1-cargo interactions, thereby enabling nuclear import regulation, have been reported in the literature. This mostly includes the modulation of Trn1-cargo interactions via post-translational modifications (e.g., arginine methylation, lysine acetylation, and serine or tyrosine phosphorylation), but also the enhancement or inhibition of the interaction by more elaborated mechanisms.

3.1.1 Post-Translational Modifications Regulating Trn1-Cargo Interactions

Arginine methylation has been reported as a modulator of Trn1-cargo interactions in several contexts. First, asymmetric arginine dimethylation within the PY-NLS of the nuclear Polyadenylate-binding protein 2 (PABP-2) reduces its affinity for Trn1 (Fronz et al., 2011). Six arginine residues that are likely involved in Trn1-binding as part of PY-NLS epitopes 1 and 2, are subjected to post-translational modifications, which would explain the reduction of

interaction with Trn1. It is worth noting that arginine methylation is not the only mechanism that modulates interaction of PABP-2 with Trn1, since binding of PABP-2 to poly-A RNA competes with binding to Trn1. The two mechanisms are interconnected since arginine methylation slightly enhances binding of PABP-2 to RNA, which altogether reduces its interaction with Trn1 (Fronz et al., 2011). In cells, whether this competition occurs in the cytoplasm or in the nucleus has not yet been firmly established. Similarly, asymmetric arginine dimethylation of FUS was reported to reduce interaction with Trn1 thereby affecting nuclear import (Dormann et al., 2012). In contrast to the PABP-2 situation, arginine methylation does not occur directly within the PY-NLS of FUS but next to it, in the so-called RGG3 region present just upstream of the C-terminal PY-NLS. Interestingly, although arginine dimethylation modulates Trn1 binding to wild-type FUS with a fully functional PY-NLS, the effect of arginine methylation in the RGG3 region is more pronounced in the case of FUS mutants with defective PY-NLSs, and inhibition of arginine methylation in these mutant proteins restores an efficient nuclear import (Dormann et al., 2012). Strikingly, it was later reported that in contrast to arginine dimethylation that drastically reduces the interaction of the RGG3 region of FUS with Trn1, arginine monomethylation only slightly affects this interaction and behaves similarly to unmethylated arginine in FUS (Suárez-Calvet et al., 2016). Owing to their almost identical domain organization, the FET proteins, namely FUS, EWS and TAF15, are similarly affected by arginine methylation in their interaction properties with Trn1 (Dormann et al., 2012; Suárez-Calvet et al., 2016). Along the same lines, arginine methylation of the non-PY-NLS of CIRBP, also referred to as the RG/RGG region, reduces its binding affinity to Trn1, thereby regulating nuclear import of CIRBP (Bourgeois et al., 2020).

Lysine acetylation has been reported to regulate the interaction of FUS with Trn1. A lysine acetylation within the PY-NLS of FUS was shown to disrupt the interaction between FUS and Trn1, resulting in the reduction of FUS nuclear import and hence its mislocalization in the cytoplasm (Arenas et al., 2020). The implicated lysine is situated in epitope 1 of FUS PY-NLS (Figure 3), and is directly in contact with a glutamate and an aspartate residue of Trn1 HEATs H14 and H15 (Zhang and Chook, 2012; Niu et al., 2012). The reduction of interaction with Trn1 upon acetylation of this specific lysine is thereby clearly in accordance with structural data.

Phosphorylations have been reported as modulator of Trn1-cargo interactions in several contexts. First, the Trn1-mediated import of Sam68 was reported as regulated through phosphorylation (Lukong et al., 2005). The BRK kinase indeed phosphorylates Sam68 on three tyrosine residues in its PY-NLS, including Y440 that forms the PY motif of PY-NLS epitope 3 (Figure 3). While unphosphorylated Sam68 is predominantly nuclear, Sam68 phosphorylated on Y440 relocates to cytoplasmic perinuclear structures (Lukong et al., 2005), most probably as a consequence of an impaired interaction with Trn1. Similarly, the FET proteins FUS and EWS have been shown to be phosphorylated on their C-terminal tyrosine (i.e., Y526 and Y656, respectively), which forms the PY motif of their PY-NLSs (Leemann-Zakaryan et al., 2011; Darovic et al., 2015). In the

case of FUS, this phosphorylation completely abolishes the PY-NLS interaction with Trn1 (Darovic et al., 2015). The tyrosine residue that is phosphorylated forms epitope 3 of FUS PY-NLS (Figure 3), and contacts several Trn1 residues, including an aspartate residue in HEAT H8 (Zhang and Chook, 2012; Niu et al., 2012). The reduction of interaction with Trn1 upon phosphorylation is thereby clearly in accordance with structural data, since phosphorylation would cause steric hindrance and electrostatic repulsion between the PY motif and its binding site on Trn1 (Darovic et al., 2015). Finally, modulation of Trn1-cargo interaction upon phosphorylation has also been reported in the case of hnRNP A1 (Allemand et al., 2005). Upon osmotic stress, hnRNP A1 is relocalized from the nucleus to the cytoplasm. This cytoplasmic relocalization is dependent on the phosphorylation of several serine residues present at the C-terminus next to the 'M9' PY-NLS of hnRNP A1 (Allemand et al., 2005). Importantly, the hyperphosphorylation of hnRNP A1 C-terminal region reduces its ability to interact with Trn1. The mechanism by which hyperphosphorylation of serines near the PY-NLS of hnRNP A1 leads to impaired binding to Trn1 is unclear, but it has been proposed that these phosphorylations might modulate the accessibility of the PY-NLS (Allemand et al., 2005).

Interestingly, in a completely different context, osmotic stress also caused the cytoplasmic translocation of FUS in neurons (Hock et al., 2018). In this case, relocalization of FUS upon osmotic stress does not depend on phosphorylations of the cargo, but is caused by a global impairment of Trn1-mediated nuclear import. As a consequence, several Trn1 cargoes (e.g., EWS, TAF15, and hnRNP A1) also displayed a marked cytoplasmic shift in neurons (Hock et al., 2018).

3.1.2 Reactive Oxidative Species Regulating Trn1-Cargo Interactions

As already introduced above with the example of FOXO4, reactive oxidative species (ROS) can modulate Trn1-cargo interaction, and thereby regulate nuclear import (Putker et al., 2013). In this example, accumulation of ROS leads to the formation of a covalent complex between FOXO4 NLS and Trn1, which ensures a strong Trn1-cargo interaction and an efficient nuclear import (Putker et al., 2013). In another example, ROS also stimulates the interaction of Trn1 with the Parkinson protein 7 (PARK7/DJ-1), but the mechanism seems very different (Björklom et al., 2014). In this system, H₂O₂ treatment disrupts the DJ-1 dimeric complex, which renders the otherwise masked PY-NLS accessible for Trn1 binding, thereby activating nuclear import of monomeric DJ-1. Here ROS enhances the binding of the DJ-1 PY-NLS to Trn1, but the effect is likely indirect and triggered by DJ-1 monomerization (Björklom et al., 2014). In a last example, it was also shown that ROS stimulates the interaction of Trn1 with the circadian clock protein PERIOD 1 (PER1) (Korge et al., 2018). Here, H₂O₂ treatment resulted in a dose-dependent increase of Trn1-PER1 interaction. It is worth mentioning that Trn1-binding to PER1 is abolished under reducing conditions, which suggests that a disulfide bond might be formed between a cysteine residue of PER1 and a cysteine residue in Trn1, as reported for FOXO4 (Putker

et al., 2013). Unexpectedly, increased binding upon H₂O₂ treatment did not lead to an increased nuclear import of PER1, but was rather associated with a specific slow-down of PER1 nuclear import (Korge et al., 2018). This intriguing observation might be related to the fact that PER1 is translocated into the nucleus by several karyopherins, but would definitely deserve further examination.

3.1.3 Masking/Unmasking Functional NLS By Folding/Unfolding Transitions

In an intriguing example, modulation of Trn1-cargo interaction would involve the folding/unfolding of a zinc-finger (ZnF). In the TIS11 family of proteins, a cryptic PY-NLS is masked within a small ZnF domain, the folding of which competes with Trn1-mediated nuclear import (Twyffels et al., 2013). It is proposed that TIS11 proteins could be in equilibrium between two states, one in which the ZnF is firmly folded, and one in which it is partially unfolded. This latter state would allow the cryptic PY-NLS to become accessible and thus to mediate nuclear translocation of the protein (Twyffels et al., 2013).

3.1.4 RNA-Mediated Regulation of Trn1-Cargo Interactions

Another example of regulation of nuclear import by Trn1 consists in the RNA-mediated regulation of ADAR1 import. As presented above, the non-PY-NLS of ADAR1 is bimodular and formed by the N- and C-terminal extensions flanking a dsRBD. This molecular organization was shown to act as an RNA-sensing NLS that can be switched on and off, depending on the presence of dsRNA associated with the dsRBD (Fritz et al., 2009; Barraud et al., 2014). The non-PY-NLS of ADAR1 presents two functional and non-overlapping interaction surfaces, namely a functional dsRNA-binding interface involving the central dsRBD, and a functional Trn1-binding interface that consists of the N- and C-terminal modules flanking the folded domain. Even if these two interfaces are distinct, the non-PY-NLS of ADAR1 cannot bind to dsRNA and Trn1 simultaneously, most probably for steric reasons, which explains how this non-PY-NLS functions as an RNA-sensing NLS (Barraud et al., 2014). Interestingly, RNA-binding enhances nuclear export of ADAR1 (Fritz et al., 2009), which suggests that ADAR1 might leave the nucleus bound to substrate RNAs. The binding of the non-PY-NLS to Trn1 in the cytoplasm could then help the dissociation of ADAR1-bound RNAs and the RNA-sensing NLS would then prevent ADAR1 from carrying RNAs back into the nucleus (Barraud et al., 2014).

In another system, RNA has been shown not to impair, but to stimulate Trn1-cargo interaction. A circular RNA named circAnks1a, was indeed shown to enhance the interaction between the transcription factor YB-1 and Trn1, thus promoting the Trn1-mediated translocation of YB-1 into the nucleus (Zhang et al., 2019). The mechanism enabling the PY-NLS of YB-1 (Mordovkina et al., 2016) to interact more strongly with Trn1 in presence of RNA has not yet been investigated and would deserve further studies. Altogether, these examples illustrate the multiple layers and the different mechanisms that can affect Trn1-cargo interaction and regulate nuclear import.

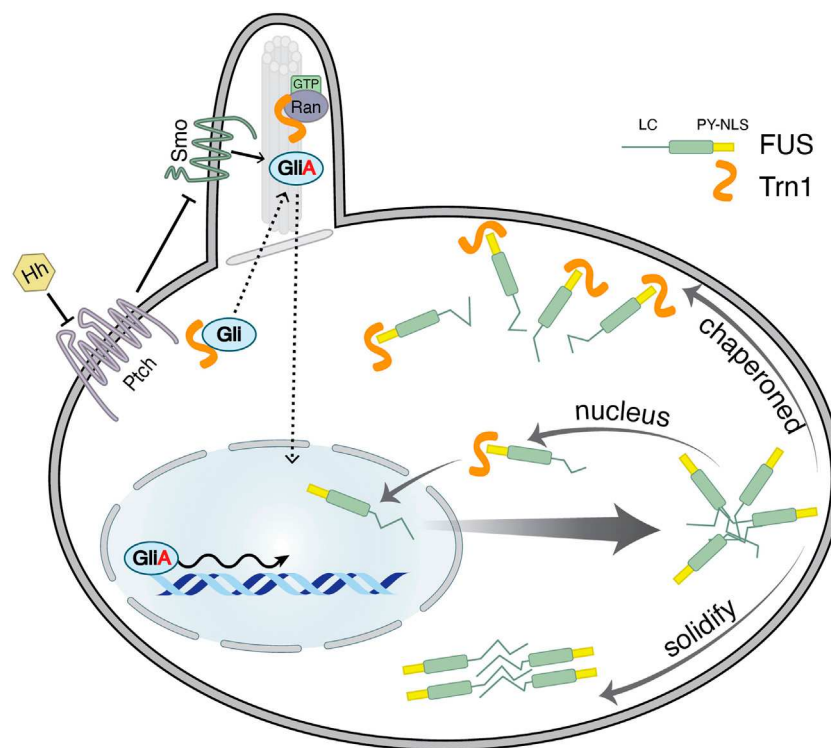


FIGURE 4 | Moonlighting functions of Transportin-1. On one hand, Trn1 has been caught in trafficking proteins destined to cilia (Dishinger et al., 2010; Hurd et al., 2011). For instance, Gli, a transcription factor associated with Hedgehog (Hh) signalling is shown to interact with Trn1 via its ciliary localization signal (CLS) (Han et al., 2017). Ran-GTP in the cilia appears to release Gli from Trn1. Upon Hh signalling Gli is converted to an activated form Gli^A, a transcriptional activator that then translocates to the nucleus to drive transcription (Kim et al., 2009; Han et al., 2017). On the other hand, Trn1 regulates the aggregation of mislocalized RNA-binding proteins (RBPs) in the cytoplasm (Guo et al., 2018; Hock et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018). Upon stress or when mutated, FUS translocates into the cytoplasm, where it can aggregate and further solidify into pathologic foci. Trn1 recognizes PY-NLS of FUS and directly drives it back into the nucleus (Dormann et al., 2010). Moreover, Trn1 can additionally bind to the low-complexity (LC) and folded region of FUS. This binding competes with FUS–FUS interaction and protein-assembly. Thus, PY-NLS serves as a cytoplasmic signal such that mislocalized nuclear RBPs are specifically chaperoned by Trn1 (Guo et al., 2018; Hock et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018).

4 MOONLIGHTING FUNCTIONS OF TRANSPORTIN-1

Besides nuclear trafficking, Trn1 has been shown to conduct a larger symphony of various cellular processes. These include ciliary transport, stress granule formation, and virus uncoating (Figure 4). In this section, we briefly describe the role of Trn1 in these processes.

4.1 Ciliary Transport

Cilia are protrusions on the cell surface, which help to bridge cells to the external environment. Primary cilia are frequently used for cell motility but can also play a key role in both chemo- and mechano-sensing. Further, several signalling pathways depend on proper ciliary function. Disruption of ciliary function leads to ciliopathies, an emerging disease condition (Reiter and Leroux, 2017). Until recently, a clear picture of proteins destined to ciliary targeting remained obscure. Meanwhile, the similarity between nuclear and ciliary transport has been demonstrated for several cases where cargo proteins interact with nuclear receptors to be

transported to cilia. In addition to a classical NLS-like ciliary localization signal (CLS), which depends upon Importin- β , CLSs have been identified that bind Trn1 to mediate their ciliary targeting (Fan et al., 2007; Dishinger et al., 2010; Hurd et al., 2011).

Retinitis pigmentosa is an X-linked disorder causing night blindness that eventually leads to permanent blindness (Patil et al., 2011). Mutations in the protein Retinitis pigmentosa 2 (RP2) are responsible for the disorder. Trn1-mediated transport of RP2 to cilia has been demonstrated (Hurd et al., 2011). The interaction between RP2 and Trn1 is mediated by two distinct sites that bind Trn1 independently. The first binding site is located at the N-terminus of RP2 and is similar to an NLS. The other binding site overlaps with the tubulin folding cofactor C (TBCC) domain of RP2 and shows slight similarity to the 'M9' sequence of hnRNP A1. Although both of these sites interact with Trn1 independently, only the M9-like sequence is essential for RP2 ciliary targeting (Hurd et al., 2011). Besides, mutations of the M9-like sequence were shown to abolish ciliary targeting of RP2 and were found mutated in several human diseases (Patil et al., 2011). Localization of RP2 to the plasma membrane was shown to

be a prerequisite for RP2 and Trn1 interaction. The plasma membrane association of RP2 is a direct consequence of myristoylation and palmitoylation of residues Gly2 and Cys3.

Primary cilia harbour many factors required for Hedgehog (Hh) signalling (Briscoe and Thérond, 2013). Hh signalling plays a central role in development and tissue homeostasis. Misregulation of the Hh signalling pathway results in developmental disorders and cancers. Trn1 has been shown to regulate Hh signalling by trafficking Gli transcription factor into cilia, where upon Hh signalling Gli is converted to activate the Gli^A transcriptional activator that then translocates to the nucleus to drive transcription (**Figure 4**) (Kim et al., 2009; Han et al., 2017). Interestingly, Trn1 itself is transcriptionally activated by Gli^A (Niewiadomski et al., 2014). Gli proteins exhibit both classical NLS and PY-NLS. The PY-like NLS along with the C-terminal residues (aa 874–1,080) is vital for the ciliary localization while the classical NLS is responsible for nuclear localization (Han et al., 2017). A knockdown of Trn1 diminishes ciliary localization of Gli both in the presence and in the absence of Hg signalling. Further, ciliary localization of Gli is reverted on transfecting Trn1 highlighting that Trn1 is necessary for Gli ciliary localization (Han et al., 2017). In addition, it was shown that Trn1 is required for Hh signalling and is therefore critical for zebrafish embryonic development.

Intra-flagellar transport (IFT) is mainly propelled by Kinesin-2 family members (Dishinger et al., 2010; Verhey et al., 2011). KIF17 is a member of the Kinesin-2 family. The similarity between Trn1-mediated nuclear and ciliary trafficking had been initially suggested from the conservation of common features (Devos et al., 2004), and direct evidence were reported few years later (Dishinger et al., 2010). It has been shown that Trn1 and RanGTP govern ciliary entry of KIF17 (Dishinger et al., 2010). Indeed, the authors have identified NLS-like sequences in KIF17: amino acids 767–777 (KRRKR) and 1,016–1,019 (KRKK). Only mutations in 1,016–1,019 (CLS) of the KIF17 tail domain abolished ciliary targeting and interaction with Trn1, whereas mutations in aa 767–772 did not affect ciliary entry. These findings indicate that Trn1 indeed governs the trafficking of KIF17 to cilia. Moreover, the C-terminal tail domain (amino acids 801–1,028) when fused to a non-ciliary kinesin promotes localization to cilia, indicating that this CLS is necessary and sufficient for ciliary transport. Consistently, a mutated CLS fused to non-ciliary kinesin did not promote transport into cilia. Further, the presence of RanGTP in cilia has been demonstrated using RanGTP specific antibodies (Fan et al., 2011). Also mass spectrometric analyses have documented the presence of Trn1 and RanGTP in the cilium (Liu et al., 2007; Ishikawa et al., 2012; Narita et al., 2012).

4.2 Virus Uncoating

Understanding the molecular mechanisms of host-virus interactions is of great medical interest. Virus uncoating is one of the critical steps in virus infection which involves the timely release of the viral genome from its shell/capsid (Yamauchi and Greber, 2016; Helenius, 2018). The stimulus and mechanistic details of virus uncoating remain largely concealed. Virus infection in eukaryotic cells often depends on the nuclear

transport machinery (Cohen et al., 2011; Matreyek and Engelman, 2013; Tessier et al., 2019). Recent studies have shown a direct role of Trn1 in virus uncoating (Fernandez et al., 2019; Miyake et al., 2019; Carlon-Andres et al., 2020; Yamauchi, 2020).

Uncoating of Influenza virus A (IVA) happens in three critical stages: priming, M1 shell dissociation and virus ribonucleoprotein untangling. Trn1 was found among the novel hits in short interfering RNA (siRNA)-based infection screenings to identify host proteins associated with viral cytosolic uncoating and nuclear import (Miyake et al., 2019). Upon Trn1 depletion by siRNA or short hairpin RNA (shRNA) knockdown, a drastic reduction in the infection rate (66–79%) in different cell types was observed. Infection was reversed upon expression of GFP-Trn1 that was insensitive to si or sh RNA. Immunofluorescence performed 4 hours post infection demonstrated the majority of the M1-ribonucleoprotein (vRNP) to be present in the cytosol in Trn1 depleted cells while M1-vRNP was present in the nucleus in MOCK depleted cells (Miyake et al., 2019). Glycine 18 and adjacent residues of M1 was identified to be essential for an interaction with Trn1. Further, a G18A mutation in M1 was shown to affect both viral assembly and uncoating, ultimately resulting in a drastic reduction in infection (Yamauchi, 2020). The atomic structure of G18A showed no major rearrangement when compared to wild-type M1, indicating that the mutation most likely affects the interaction with Trn1.

Capsid protein (CA) multimers encase the RNA genome of human immunodeficiency virus-1 (HIV-1) (Campbell and Hope, 2015). The capsid is believed to provide a secured micro-environment for the viral reverse transcriptase to transform RNA genome to double-stranded DNA and to evade the cell's innate immune system (Lahaye et al., 2013; Rasaiyaah et al., 2013). It is largely accepted that timely dissolution of CA is mandatory for efficient infection but the exact mechanism remains unclear. Since HIV-1 uncoating is linked to nuclear entry, Trn1 seems to play a critical role in HIV-1 infection (Fernandez et al., 2019). To test this, the authors knocked down Trn1, which led to a significant reduction in HIV-1 infection. While Trn1 transfection restored HIV-1 infection, implying a direct role of Trn1 in HIV-1 infection. Although Trn1 and Trn2 share 83% identity, Trn2 knockdown did not affect HIV-1 infection, suggesting a specific function of Trn1. Moreover, a significant reduction in infection was seen in Trn1 depleted CD4⁺ cells, substantiating that Trn1 is a prerequisite for early steps of infection (Fernandez et al., 2019). CA and Trn1 could be co-immunoprecipitated in HIV-1 infected cells, indicating interaction of Trn1 with CA. Further, Trn1 colocalization with CA in the cytosol and at the nuclear pore was shown. CA exhibits a hydrophobic patch containing a glycine residue at position 89 (Fernandez et al., 2019). Such a glycine residue was shown to be critical for interaction of hnRNP A1 PY-NLS with Trn1 (Lee et al., 2006). Consistently, the glycine at position 89 (G89) is conserved throughout HIV-1 subtypes (Kuiken et al., 2003) and G89 in conjunction with the proline at position 90 (P90) was shown to bind the peptidylprolyl isomerase CypA (Gamble et al., 1996; Yoo et al., 1997). A

G89V mutation drastically reduced the infection rate when compared to a P90A mutation, suggesting a dependency of G89 for the interaction with Trn1. Further, the dependency of Trn1-CA interaction on G89 was shown *in vitro* by surface plasmon resonance (Fernandez et al., 2019). The authors demonstrated the HIV-1 CA uncoating dependency on Trn1 by an elegant fate-of-capsid assay on sedimentation gradients. Soluble CA, representing fully uncoated capsid was only present in control HeLa cytosolic fractions and not in Trn1 knockdown lysates. Purified recombinant Trn1 was also shown to cause structural damage to the assembled capsid/nucleocapsid (CANC) by atomic force microscopy. Additionally, W730 of Trn1 was shown to be critical for the interaction with CA and uncoating (Fernandez et al., 2019). Molecular docking simulations revealed that Trn1 may have other points of contact in addition to G89 with CA hexamers and it may induce strong hindrance leading to uncoating. Moreover, Trn1 depletion significantly reduced nuclear accumulation of both CA and viral DNA. While CA and viral pre-integration complex DNA (PIC DNA) accumulated in the nucleus on Trn1 expression, suggesting that Trn1 traffics both capsid and PIC DNA.

4.3 Regulation of Stress Granule Formation

As mentioned previously, nuclear import of numerous RNA-binding proteins (RBPs) is controlled by Trn1, including the FET protein family (FUS, EWS, TAF15) and hnRNP A1/A2 (Lee et al., 2006). Upon stress response or when mutated, these predominantly nuclear proteins can translocate to the cytoplasm and accumulate in stress granules (SGs). SGs can further mature into pathogenic inclusions that are typical for a group of fatal neurodegenerative disorders including amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and multisystem proteinopathy (MSP) (Dormann et al., 2010; Harrison and Shorter, 2017; Neumann et al., 2011). Recent studies show, that Trn1 not only acts as a transport receptor but presumably serves as a powerful cytoplasmic chaperone for mislocalized RBPs. Thus, Trn1 regulates SGs formation and their further fibrillization (Figure 4) (Guo et al., 2018; Hock et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018; Sun et al., 2020).

4.1.1 Trn1 is a Chaperone of Mislocalized FET-Proteins

FET proteins are involved in many steps of RNA metabolism such as transcription and splicing (Lagier-Tourenne et al., 2010). Moreover, they possess additional cytoplasmic function e.g., FUS modulates the axonal mRNA transport and local translation (Fujii and Takumi, 2005; López-Erauskin et al., 2020). FET proteins share a common domain-architecture: An N-terminal Prion-like (PrLD) or low complexity domain (LC), followed by several arginine-glycine-glycine (RGG) regions, an RNA recognition motif (RRM) and a C-terminal PY-NLS (Iko et al., 2004; Springhower et al., 2020).

ALS-associated FUS mutations are usually found in the PY-NLS and directly disturb nuclear import mediated by Trn1.

This leads to FUS cytoplasmic retention and subsequent disease manifestation (Dormann et al., 2010). However, recent studies showed an additional chaperoning function of Trn1 that shapes phase separation of FUS and other FET proteins. The PY-NLS anchors Trn1 to mislocalized cargos and promotes additional weak binding to low complexity and folded regions which compete with self-aggregation of FET proteins (Figure 4) (Guo et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018). This chaperoning function seems to be independent of Trn1-mediated nuclear import, as import mutants of Trn1 (Chook et al., 2002) still prevent and reverse FUS fibrillization. In addition, Trn1 can specifically extract its cargos from SGs without affecting SG-formation per se (Guo et al., 2018). Thus, in this case, the NLS might serve as a signal that ensures that nuclear cargos are chaperoned and disaggregated when they are trapped in the cytoplasm (Figure 4) (Guo et al., 2018; Hofweber et al., 2018; Qamar et al., 2018; Yoshizawa et al., 2018). Similar mechanisms have been proposed for the Trn1-mediated disaggregation of the amyloid fibrils formed by hnRNP A1 LC domain, thus conferring a protective activity against hnRNP A1-driven ALS and MSP (Sun et al., 2020).

Chaperoning is presumably a common feature of importins, given that Imp α /Imp β could prevent TDP43 fibrillization *in vitro* (Guo et al., 2018). Moreover, Imp α /Imp β inhibits phase separation of FUS if fused to a classical NLS (Yoshizawa et al., 2018). Importin-mediated disaggregation might be similar to the transition of the nuclear pore complex, where importins break weak hydrophobic interactions between FG-Nups (Schmidt and Görlich, 2016). Corresponding to the interaction with nucleoporins, Trn1 interacts with tyrosines in the FUS low-complexity domain. Moreover, Trn1 contains highly acidic surfaces and loops that possibly interact with the basic RGG regions of FUS (Yoshizawa et al., 2018). Thus, Trn1 and presumably importins in general, might serve as cytoplasmic chaperones, that regulate the dynamics of phase separation and thereby the content of stress granules.

4.1.2 Chaperoning Modulation By Arginine Methylation and RNA-Binding

Arginine methylation can modulate not only direct recognition and binding to FUS-NLS as mentioned before (Dormann et al., 2012), but also chaperoning of FUS by Trn1. FUS-FTD is usually not linked to FUS mutations although its manifestation is analogous to ALS. Immunohistochemistry, however, revealed that the FUS-FTD inclusions contain hypomethylated arginines, hence suggesting a different pathomolecular mechanism that leads to fibril formation in FUS-FTDs (Dormann et al., 2012). Cation- π interaction between hypomethylated arginines in RGG and tyrosines in LC promotes FUS condensation into stable inter-molecular β -sheet-rich hydrogels, that cannot be disassociated by physiological Trn1-level (Qamar et al., 2018). Thus, methylation of arginines in RGG regions tunes the strength of FUS intermolecular interactions that promote phase separation (Hofweber et al., 2018; Qamar et al., 2018).

In the nucleus, high concentration of RNA saturates binding properties of FUS molecules and inhibits phase separation. In contrary, the low RNA concentration in the cytoplasm may promote FUS-RNP assembly (Maharana et al., 2018). RGG regions bind RNA and Trn1 in a mutually exclusive manner. Thus, RNA replacement by Trn1 might also contribute to FUS chaperoning (Hofweber et al., 2018; Yoshizawa et al., 2018). ALS-associated arginine-mutations lead to aberrant RNA binding that is static and, on the contrary to wild-type FUS, does not switch into dynamic binding stage and promotes larger droplets formation (Niaki et al., 2020).

5 CONCLUDING REMARKS

The recognition at a molecular level of NLSs by Trn1 is now relatively well understood at least for the PY-NLS family. The recognition of non-PY-NLSs by Trn1 is however far less understood and would definitely deserve comprehensive biochemical, structural and cellular studies. These studies focussed on non-PY-NLS recognition by Trn1 would be pivotal to achieve a broad understanding of the mode of action of Trn1 and to shed light on this rather unexplored area of Trn1-cargo recognition. It is worth noting that in the study of the CIRBP non-PY-NLS, an unexpected role for the H8 loop in NLS binding has been unveiled (Bourgeois et al., 2020). The role of the H8 loop has been traditionally restricted to cargo dissociation upon RanGTP binding, and was reported as dispensable for NLS binding in early studies (Lee et al., 2006; Imasaki et al., 2007). This new study thereby raises the possibility that the Trn1 H8 loop might also be important for binding to particular NLSs.

Although the implication of Trn1 in ciliary trafficking and virus uncoating has been clearly established, the binding of Trn1

in these processes and the related molecular mechanisms remain largely unknown. As for the recognition of non-PY-NLSs, detailed biochemical and structural studies would for instance greatly help understand how Trn1 interacts with substrates destined to ciliary trafficking.

Nowadays, Trn1 attracts a lot of attention for its clear protective activity in RNA-binding protein phase separation and maturation of their inclusions in cells. Overexpression of Trn1 reverses aberrant phase separation *in vitro* and in cells (Guo et al., 2018; Hofweber et al., 2018; Yoshizawa et al., 2018). Trn1 restores FUS-mRNA targets expression in ALS-derived fibroblast and rescues impaired protein synthesis in axon terminals (Guo et al., 2018; Qamar et al., 2018). Studies in fly models showed an increase in lifespan, when mutated FUS was co-expressed with Trn1 in motor neurons, and complete rescue of MSP-linked hnRNP A2 mutation (Guo et al., 2018). Together, these recent studies indicate that Trn1 might have a direct therapeutic utility in fatal neurodegenerative diseases.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

PB and MJ are supported by a joint ANR-FWF Grant (Nos. ANR-16-CE91-0003 and FWF-I2893). Research at the IBPC is supported by the CNRS, and the LABEX Dynamo (ANR-11-LABX-0011). Research in the lab of MJ is supported by FWF Grant numbers F80-07 and P32678.

REFERENCES

- Allegretti, M., Zimmerli, C. E., Rantos, V., Wilfling, F., Ronchi, P., Fung, H. K. H., et al. (2020). In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* 14, 33–39. doi:10.1038/s41586-020-2670-5
- Allemand, E., Guil, S., Myers, M., Moscat, J., Cáceres, J. F., and Krainer, A. R. (2005). Regulation of heterogenous nuclear ribonucleoprotein A1 transport by phosphorylation in cells stressed by osmotic shock. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3605–3610. doi:10.1073/pnas.0409889102
- Andrade, M. A., and Bork, P. (1995). HEAT repeats in the Huntington's disease protein. *Nat. Genet.* 11, 115–116. doi:10.1038/ng1095-115
- Andrade, M. A., Petosa, C., O'Donoghue, S. I., Müller, C. W., and Bork, P. (2001). Comparison of ARM and HEAT protein repeats. *J. Mol. Biol.* 309, 1–18. doi:10.1006/jmbi.2001.4624
- Arenas, A., Chen, J., Kuang, L., Barnett, K. R., Kasarskis, E. J., Gal, J., et al. (2020). Lysine acetylation regulates the RNA binding, subcellular localization and inclusion formation of FUS. *Hum. Mol. Genet.* 29, 2684–2697. doi:10.1093/hmg/ddaa159
- Arnold, M., Nath, A., Hauber, J., and Kehlenbach, R. H. (2006). Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *J. Biol. Chem.* 281, 20883–20990. doi:10.1074/jbc.M602189200
- Baade, I., and Kehlenbach, R. H. (2019). The cargo spectrum of nuclear transport receptors. *Curr. Opin. Cell Biol.* 58, 1–7. doi:10.1016/j.ceb.2018.11.004
- Baake, M., Bäuerle, M., Doenecke, D., and Albig, W. (2001). Core histones and linker histones are imported into the nucleus by different pathways. *Eur. J. Cell Biol.* 80, 669–677. doi:10.1078/0171-9335-00208
- Bader, A. G., and Vogt, P. K. (2005). Inhibition of protein synthesis by Y box-binding protein 1 blocks oncogenic cell transformation. *Mol. Cell Biol.* 25, 2095–2106. doi:10.1128/MCB.25.6.2095-2106.2005
- Banerjee, S., and Barraud, P. (2014). Functions of double-stranded RNA-binding domains in nucleocytoplasmic transport. *RNA Biol.* 11, 1226–1232. doi:10.4161/15476286.2014.972856
- Barraud, P., and Allain, F. H. (2012). ADAR proteins: double-stranded RNA and Z-DNA binding domains. *Curr. Top. Microbiol. Immunol.* 353, 35–60. doi:10.1007/82_2011_145
- Barraud, P., and Allain, F. H. (2013). Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM influences the nucleic acid binding topology. *J. Biomol. NMR* 55, 119–138. doi:10.1007/s10858-012-9696-4
- Barraud, P., Banerjee, S., Mohamed, W. I., Jantsch, M. F., and Allain, F. H. (2014). A bimodular nuclear localization signal assembled via an extended double-stranded RNA-binding domain acts as an RNA-sensing signal for transportin 1. *Proc. Natl. Acad. Sci. U.S.A.* 111, E1852–E1861. doi:10.1073/pnas.1323698111
- Beck, M., and Hurt, E. (2017). The nuclear pore complex: understanding its function through structural insight. *Nat. Rev. Mol. Cell Biol.* 18, 73–89. doi:10.1038/nrm.2016.147

- Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A., and Ponstingl, H. (1994). RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc. Natl. Acad. Sci. U.S.A.* 91, 2587–2591. doi:10.1073/pnas.91.7.2587
- Bischoff, F. R., and Görlich, D. (1997). RanBP1 is crucial for the release of RanGTP from importin beta-related nuclear transport factors. *FEBS Lett.* 419, 249–254. doi:10.1016/s0014-5793(97)01467-1
- Björkblom, B., Maple-Grødem, J., Puno, M. R., Odell, M., Larsen, J. P., and Møller, S. G. (2014). Reactive oxygen species-mediated DJ-1 monomerization modulates intracellular trafficking involving karyopherin β 2. *Mol. Cell Biol.* 34, 3024–3040. doi:10.1128/MCB.00286-14
- Blackwell, J. S., Jr, Wilkinson, S. T., Mosammaparast, N., and Pemberton, L. F. (2007). Mutational analysis of H3 and H4 N termini reveals distinct roles in nuclear import. *J. Biol. Chem.* 282, 20142–20150. doi:10.1074/jbc.M701989200
- Bonifaci, N., Moroiu, J., Radu, A., and Blobel, G. (1997). Karyopherin beta2 mediates nuclear import of a mRNA binding protein. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5055–5060. doi:10.1073/pnas.94.10.5055
- Bourgeois, B., Hutten, S., Gottschalk, B., Hofweber, M., Richter, G., Sternat, J., et al. (2020). Nonclassical nuclear localization signals mediate nuclear import of CIRBP. *Proc Natl Acad Sci USA* 117, 8503–8514. doi:10.1073/pnas.1918944117
- Briscoe, J., and Thérond, P. P. (2013). The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* 14, 416–429. doi:10.1038/nrm3598
- Campbell, E. M., and Hope, T. J. (2015). HIV-1 capsid: the multifaceted key player in HIV-1 infection. *Nat. Rev. Microbiol.* 13, 471–483. doi:10.1038/nrmicro3503
- Cansizoglu, A. E., and Chook, Y. M. (2007). Conformational heterogeneity of karyopherin beta2 is segmental. *Structure* 15, 1431–1441. doi:10.1016/j.str.2007.09.009
- Cansizoglu, A. E., Lee, B. J., Zhang, Z. C., Fontoura, B. M., and Chook, Y. M. (2007). Structure-based design of a pathway-specific nuclear import inhibitor. *Nat. Struct. Mol. Biol.* 14, 452–454. doi:10.1038/nsmb1229
- Carlón-Andrés, I., Lagadec, F., Pied, N., Rayne, F., Lafon, M. E., Kehlenbach, R. H., et al. (2020). Nup358 and Transportin 1 cooperate in adenoviral genome import. *J. Virol.* 94, 33. doi:10.1128/JVI.00164-20
- Chook, Y. M., and Blobel, G. (1999). Structure of the nuclear transport complex karyopherin-beta2-Ran x GppNHp. *Nature* 399, 230–237. doi:10.1038/20375
- Chook, Y. M., Jung, A., Rosen, M. K., and Blobel, G. (2002). Uncoupling Kapbeta2 substrate dissociation and ran binding. *Biochemistry* 41, 6955–6966. doi:10.1021/bi012122p
- Chook, Y. M., and Süel, K. E. (2011). Nuclear import by karyopherin- β s: recognition and inhibition. *Biochim. Biophys. Acta* 1813, 1593–606. doi:10.1016/j.bbamcr.2010.10.014
- Cingolani, G., Petosa, C., Weis, K., and Müller, C. W. (1999). Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature* 399, 221–229. doi:10.1038/20367
- Cohen, S., Au, S., and Panté, N. (2011). How viruses access the nucleus. *Biochim. Biophys. Acta* 1813, 1634–45. doi:10.1016/j.bbamcr.2010.12.009
- Conti, E., and Izaurrealde, E. (2001). Nucleocytoplasmic transport enters the atomic age. *Curr. Opin. Cell Biol.* 13, 310–319. doi:10.1016/s0955-0674(00)00213-1
- Conti, E., Müller, C. W., and Stewart, M. (2006). Karyopherin flexibility in nucleocytoplasmic transport. *Curr Opin Struct Biol* 16, 237–244. doi:10.1016/j.sbi.2006.03.010
- Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* 76, 647–671. doi:10.1146/annurev.biochem.76.052705.161529
- Cook, A. G., and Conti, E. (2010). Nuclear export complexes in the frame. *Curr. Opin. Struct. Biol.* 20, 247–252. doi:10.1016/j.sbi.2010.01.012
- Darovic, S., Prpar Mihevc, S., Župunski, V., Gunčar, G., Stalekar, M., Lee, Y. B., et al. (2015). Phosphorylation of C-terminal tyrosine residue 526 in FUS impairs its nuclear import. *J. Cell Sci.* 128, 4151–4159. doi:10.1242/jcs.176602
- Dasso, M. (2002). The Ran GTPase: theme and variations. *Curr Biol* 12, R502–R508. doi:10.1016/s0960-9822(02)00970-3
- Delphin, C., Guan, T., Melchior, F., and Gerace, L. (1997). RanGTP targets p97 to RanBP2, a filamentous protein localized at the cytoplasmic periphery of the nuclear pore complex. *Mol. Biol. Cell* 8, 2379–2390. doi:10.1091/mbc.8.12.2379
- Devos, D., Dokudovskaya, S., Alber, F., Williams, R., Chait, B. T., Sali, A., Rout, M. P., et al. (2004). Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* 2, e380. doi:10.1371/journal.pbio.0020380
- Dishinger, J. F., Kee, H. L., Jenkins, P. M., Fan, S., Hurd, T. W., Hammond, J. W., et al. (2010). Ciliary entry of the kinesin-2 motor KIF17 is regulated by importin-beta2 and RanGTP. *Nat. Cell Biol.* 12, 703–710. doi:10.1038/ncb2073
- Dormann, D., Rodde, R., Edbauer, D., Bentmann, E., Fischer, I., Hruscha, A., et al. (2010). ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. *EMBO J.* 29, 2841–2857. doi:10.1038/emboj.2010.143
- Dormann, D., Madl, T., Valori, C. F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., et al. (2012). Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. *EMBO J.* 31, 4258–4275. doi:10.1038/emboj.2012.261
- Eckmann, C. R., Neunteufl, A., Pfaffstetter, L., and Jantsch, M. F. (2001). The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein. *Mol. Biol. Cell* 12, 1911–1924. doi:10.1091/mbc.12.7.1911
- Fan, X. C., and Steitz, J. A. (1998). HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15293. doi:10.1073/pnas.95.26.15293
- Fan, S., Fogg, V., Wang, Q., Chen, X. W., Liu, C. J., and Margolis, B. (2007). A novel Crumbs3 isoform regulates cell division and ciliogenesis via importin beta interactions. *J. Cell Biol.* 178, 387–398. doi:10.1083/jcb.200609096
- Fan, S., Whiteman, E. L., Hurd, T. W., McIntyre, J. C., Dishinger, J. F., Liu, C. J., et al. (2011). Induction of Ran GTP drives ciliogenesis. *Mol. Biol. Cell* 22, 4539–4548. doi:10.1091/mbc.E11-03-0267
- Fernandez, J., Machado, A. K., Lonnais, S., Chantontin, C., Gärtner, K., Léger, T., et al. (2019). Transportin-1 binds to the HIV-1 capsid via a nuclear localization signal and triggers uncoating. *Nat. Microbiol.* 4, 1840–1850. doi:10.1038/s41564-019-0575-6
- Floer, M., and Blobel, G. (1996). The nuclear transport factor karyopherin beta binds stoichiometrically to Ran-GTP and inhibits the Ran GTPase activating protein. *J. Biol. Chem.* 271, 5313–5316. doi:10.1074/jbc.271.10.5313
- Fridell, R. A., Truant, R., Thorne, L., Benson, R. E., and Cullen, B. R. (1997). Nuclear import of hnRNP A1 is mediated by a novel cellular cofactor related to karyopherin-beta. *J. Cell Sci.* 110 (Pt 11), 1325–1331.
- Fried, H., and Kutay, U. (2003). Nucleocytoplasmic transport: taking an inventory. *Cell Mol. Life Sci.* 60, 1659–1688. doi:10.1007/s00018-003-3070-3
- Fritz, J., Strehlbow, A., Taschner, A., Schopoff, S., Pasierbek, P., and Jantsch, M. F. (2009). RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1. *Mol. Cell Biol.* 29, 1487–1497. doi:10.1128/MCB.01519-08
- Fronz, K., Güttinger, S., Burkert, K., Kühn, U., Stöhr, N., Schierhorn, A., et al. (2011). Arginine methylation of the nuclear poly(a) binding protein weakens the interaction with its nuclear import receptor, transportin. *J. Biol. Chem.* 286, 32986–32994. doi:10.1074/jbc.M111.273912
- Fu, X., Liang, C., Li, F., Wang, L., Wu, X., Lu, A., et al. (2018). The rules and functions of nucleocytoplasmic shuttling proteins. *Int. J. Mol. Sci.* 19, 129. doi:10.3390/ijms19051445
- Fujii, R., and Takumi, T. (2005). TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. *J. Cell Sci.* 118, 5755–5765. doi:10.1242/jcs.02692
- Fukuhara, N., Fernandez, E., Ebert, J., Conti, E., and Svergun, D. (2004). Conformational variability of nucleo-cytoplasmic transport factors. *J. Biol. Chem.* 279, 2176–2181. doi:10.1074/jbc.M309112200
- Gamble, T. R., Vajdos, F. F., Yoo, S., Worthylake, D. K., Houseweart, M., Sundquist, W. I., et al. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* 87, 1285–1294. doi:10.1016/s0092-8674(00)81823-1

- Görllich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* 15, 607–660. doi:10.1146/annurev.cellbio.15.1.607
- Guo, L., Kim, H. J., Wang, H., Monaghan, J., Freyermuth, F., Sung, J. C., et al. (2018). Nuclear-import receptors reverse aberrant phase transitions of RNA-binding proteins with prion-like domains. *Cell* 173, 677–e20. doi:10.1016/j.cell.2018.03.002
- Güttinger, S., Mühlhäusser, P., Koller-Eichhorn, R., Brennecke, J., and Kutay, U. (2004). Transportin2 functions as importin and mediates nuclear import of HuR. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2918–2923. doi:10.1073/pnas.0400342101
- Hampoelz, B., Andres-Pons, A., Kastritsis, P., and Beck, M. (2019). Structure and assembly of the nuclear pore complex. *Annu. Rev. Biophys.* 48, 515–536. doi:10.1146/annurev-biophys-052118-115308
- Han, Y., Xiong, Y., Shi, X., Wu, J., Zhao, Y., and Jiang, J. (2017). Regulation of Gli ciliary localization and Hedgehog signaling by the PY-NLS/karyopherin- β 2 nuclear import system. *PLoS Biol.* 15, e2002063. doi:10.1371/journal.pbio.2002063
- Harrison, A. F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438. doi:10.1042/BCJ20160499
- Helenius, A. (2018). Virus entry: looking back and moving forward. *J. Mol. Biol.* 430, 1853–1862. doi:10.1016/j.jmb.2018.03.034
- Hock, E. M., Maniecka, Z., Hruska-Plochan, M., Reber, S., Laferrière, F., Sahadevan, M. K. S., et al. (2018). Hypertonic stress causes cytoplasmic translocation of neuronal, but not astrocytic, FUS due to impaired transportin function. *Cell Rep.* 24, 987–e7. doi:10.1016/j.celrep.2018.06.094
- Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., et al. (2018). Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation. *Cell* 173, 706–e13. doi:10.1016/j.cell.2018.03.004
- Huber, F. M., and Hoelz, A. (2017). Molecular basis for protection of ribosomal protein L4 from cellular degradation. *Nat. Commun.* 8, 14354. doi:10.1038/ncomms14354
- Hülsmann, B. B., Labokha, A. A., and Görllich, D. (2012). The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell* 150, 738–751. doi:10.1016/j.cell.2012.07.019
- Hurd, T. W., Fan, S., and Margolis, B. L. (2011). Localization of retinitis pigmentosa 2 to cilia is regulated by Importin beta2. *J. Cell Sci.* 124, 718–726. doi:10.1242/jcs.070839
- Iko, Y., Kodama, T. S., Kasai, N., Oyama, T., Morita, E. H., Muto, T., et al. (2004). Domain architectures and characterization of an RNA-binding protein, TLS. *J. Biol. Chem.* 279, 44834–44840. doi:10.1074/jbc.M408552200
- Imasaki, T., Shimizu, T., Hashimoto, H., Hidaka, Y., Kose, S., Imamoto, N., et al. (2007). Structural basis for substrate recognition and dissociation by human transportin 1. *Mol. Cell* 28, 57–67. doi:10.1016/j.molcel.2007.08.006
- Ishikawa, H., Thompson, J., Yates, J. R., 3rd, and Marshall, W. F. (2012). Proteomic analysis of mammalian primary cilia. *Curr. Biol.* 22, 414–419. doi:10.1016/j.cub.2012.01.031
- Izaurrealde, E., Kutay, U., von Kobbe, C., Mattaj, I. W., and Görllich, D. (1997). The asymmetric distribution of the constituents of the Ran system is essential for transport into and out of the nucleus. *EMBO J.* 16, 6535–6547. doi:10.1093/emboj/16.21.6535
- Jäkel, S., and Görllich, D. (1998). Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.* 17, 4491–4502. doi:10.1093/emboj/17.15.4491
- Kim, J., Kato, M., and Beachy, P. A. (2009). Gli2 trafficking links Hedgehog-dependent activation of Smoothened in the primary cilium to transcriptional activation in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21666–21667. doi:10.1073/pnas.0912180106
- Kimura, M., Kose, S., Okumura, N., Imai, K., Furuta, M., Sakiyama, N., et al. (2013). Identification of cargo proteins specific for the nucleocytoplasmic transport carrier transportin by combination of an in vitro transport system and stable isotope labeling by amino acids in cell culture (SILAC)-based quantitative proteomics. *Mol. Cell Proteomics* 12, 145–157. doi:10.1074/mcp.M112.019414
- Kimura, M., Morinaka, Y., Imai, K., Kose, S., Horton, P., and Imamoto, N. (2017). Extensive cargo identification reveals distinct biological roles of the 12 importin pathways. *Life* 6, 121. doi:10.7554/eLife.21184
- Klebe, C., Bischoff, F. R., Ponstingl, H., and Wittinghofer, A. (1995). Interaction of the nuclear GTP-binding protein Ran with its regulatory proteins RCC1 and RanGAP1. *Biochemistry* 34, 639–647. doi:10.1021/bi00002a031
- Klucsevsek, K., Daley, J., Darshan, M. S., Bordeaux, J., and Moroianu, J. (2006). Nuclear import strategies of high-risk HPV18 L2 minor capsid protein. *Virology* 352, 200–208. doi:10.1016/j.virol.2006.04.007
- Korge, S., Maier, B., Brünig, F., Ehrhardt, L., Korte, T., Mann, M., et al. (2018). The non-classical nuclear import carrier Transportin 1 modulates circadian rhythms through its effect on PER1 nuclear localization. *PLoS Genet.* 14, e1007189. doi:10.1371/journal.pgen.1007189
- Kuiken, C., Korber, B., and Shafer, R. W. (2003). HIV sequence databases. *AIDS Rev.* 5, 52–61. doi:10.2172/1566101
- Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D. W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* 19, R46–R64. doi:10.1093/hmg/ddq137
- Lahaye, X., Satoh, T., Gentili, M., Cerboni, S., Conrad, C., Hurbain, I., et al. (2013). The capsids of HIV-1 and HIV-2 determine immune detection of the viral cDNA by the innate sensor cGAS in dendritic cells. *Immunity* 39, 1132–1142. doi:10.1016/j.immuni.2013.11.002
- Le Roux, L. G., and Moroianu, J. (2003). Nuclear entry of high-risk human papillomavirus type 16 E6 oncoprotein occurs via several pathways. *J. Virol.* 77, 2330–2337. doi:10.1128/jvi.77.4.2330-2337.2003
- Lee, S. J., Matsuura, Y., Liu, S. M., and Stewart, M. (2005). Structural basis for nuclear import complex dissociation by RanGTP. *Nature* 435, 693–696. doi:10.1038/nature03578
- Lee, B. J., Cansizoglu, A. E., Süel, K. E., Louis, T. H., Zhang, Z., and Chook, Y. M. (2006). Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* 126, 543–558. doi:10.1016/j.cell.2006.05.049
- Leemann-Zakaryan, R. P., Pahlisch, S., Grossenbacher, D., and Gehring, H. (2011). Tyrosine Phosphorylation in the C-terminal nuclear localization and retention signal (C-NLS) of the EWS protein. *Sarcoma* 2011, 218483. doi:10.1155/2011/218483
- Liu, Q., Tan, G., Levenkova, N., Li, T., Pugh, E. N., Jr, Rux, J. J., et al. (2007). The proteome of the mouse photoreceptor sensory cilium complex. *Mol. Cell Proteomics* 6, 1299–1317. doi:10.1074/mcp.M700054-MCP200
- López-Erauskin, J., Tadokoro, T., Baughn, M. W., Myers, B., McAlonis-Downes, M., Chillon-Marinas, C., et al. (2020). ALS/FTD-linked mutation in FUS suppresses intra-axonal protein synthesis and drives disease without nuclear loss-of-function of FUS. *Neuron* 106, 354. doi:10.1016/j.neuron.2020.04.006
- Lukong, K. E., Larocque, D., Tyner, A. L., and Richard, S. (2005). Tyrosine phosphorylation of sam68 by breast tumor kinase regulates intranuclear localization and cell cycle progression. *J. Biol. Chem.* 280, 38639–38647. doi:10.1074/jbc.M505802200
- Mackmull, M. T., Klaus, B., Heinze, I., Chokkalingam, M., Beyer, A., Russell, R. B., et al. (2017). Landscape of nuclear transport receptor cargo specificity. *Mol. Syst. Biol.* 13, 962. doi:10.15252/msb.20177608
- Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., et al. (2018). RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* 360, 918–921. doi:10.1126/science.aar7366
- Matreyek, K. A., and Engelman, A. (2013). Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. *Viruses* 5, 2483–2511. doi:10.3390/v5102483
- Michael, W. M., Choi, M., and Dreyfuss, G. (1995). A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* 83, 415–422. doi:10.1016/0092-8674(95)90119-1
- Miyake, Y., Keusch, J. J., Decamps, L., Ho-Xuan, H., Iketani, S., Gut, H., et al. (2019). Influenza virus uses transportin 1 for vRNP debundling during cell entry. *Nat. Microbiol.* 4, 578–586. doi:10.1038/s41564-018-0332-2
- Mordovkina, D. A., Kim, E. R., Buldakov, I. A., Sorokin, A. V., Eliseeva, I. A., Lyabin, D. N., et al. (2016). Transportin-1-dependent YB-1 nuclear import. *Biochem. Biophys. Res. Commun.* 480, 629–634. doi:10.1016/j.bbrc.2016.10.107

- Mosammaparast, N., Guo, Y., Shabanowitz, J., Hunt, D. F., and Pemberton, L. F. (2002). Pathways mediating the nuclear import of histones H3 and H4 in yeast. *J. Biol. Chem.* 277, 862–888. doi:10.1074/jbc.M106845200
- Mühlhäusser, P., Müller, E. C., Otto, A., and Kutay, U. (2001). Multiple pathways contribute to nuclear import of core histones. *EMBO Rep.* 2, 690–696. doi:10.1093/embo-reports/kve168
- Narita, K., Kozuka-Hata, H., Nonami, Y., Ao-Kondo, H., Suzuki, T., Nakamura, H., et al. (2012). Proteomic analysis of multiple primary cilia reveals a novel mode of ciliary development in mammals. *Biol. Open* 1, 815–825. doi:10.1242/bio.20121081
- Neumann, M., Bentmann, E., Dormann, D., Jawaide, A., DeJesus-Hernandez, M., Ansoorge, O., et al. (2011). FET proteins TAF15 and EWS are selective markers that distinguish FTLD with FUS pathology from amyotrophic lateral sclerosis with FUS mutations. *Brain* 134, 2595–609. doi:10.1093/brain/awr201
- Niaki, A. G., Sarkar, J., Cai, X., Rhine, K., Vidaurre, V., Guy, B., et al. (2020). Loss of dynamic RNA interaction and aberrant phase separation induced by two distinct types of ALS/FTD-linked FUS mutations. *Mol. Cell* 77, 82–e4. doi:10.1016/j.molcel.2019.09.022
- Niewiadomski, P., Kong, J., H., Ahrends, R., Ma, Y., Humke, E. W., Khan, S., et al. (2014). Gli protein activity is controlled by multisite phosphorylation in vertebrate Hedgehog signaling. *Cell Rep.* 6, 168–181. doi:10.1016/j.celrep.2013.12.003
- Niu, C., Zhang, J., Gao, F., Yang, L., Jia, M., Zhu, H., et al. (2012). FUS-NLS/Transportin 1 complex structure provides insights into the nuclear targeting mechanism of FUS and the implications in ALS. *PLoS One* 7, e47056. doi:10.1371/journal.pone.0047056
- Patil, S. B., Hurd, T. W., Ghosh, A. K., Murga-Zamalloa, C. A., and Khanna, H. (2011). Functional analysis of retinitis pigmentosa 2 (RP2) protein reveals variable pathogenic potential of disease-associated missense variants. *PLoS One* 6, e21379. doi:10.1371/journal.pone.0021379
- Pollard, V. W., Michael, W. M., Nakielnny, S., Siomi, M. C., Wang, F., and Dreyfuss, G. (1996). A novel receptor-mediated nuclear protein import pathway. *Cell* 86, 985–994. doi:10.1016/s0092-8674(00)80173-7
- Putker, M., Madl, T., Vos, H. R., de Ruiter, H., Visscher, M., van den Berg, M. C., et al. (2013). Redox-dependent control of FOXO/DAF-16 by transportin-1. *Mol. Cell* 49, 730–742. doi:10.1016/j.molcel.2012.12.014
- Qamar, S., Wang, G., Randle, S. J., Ruggeri, F. S., Varela, J. A., Lin, J. Q., et al. (2018). FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation- π interactions. *Cell* 173, 720–e15. doi:10.1016/j.cell.2018.03.056
- Rasaiyaah, J., Tan, C. P., Fletcher, A. J., Price, A. J., Blondeau, C., Hilditch, L., et al. (2013). HIV-1 evades innate immune recognition through specific cofactor recruitment. *Nature* 503, 402–405. doi:10.1038/nature12769
- Rebane, A., Aab, A., and Steitz, J. A. (2004). Transportins 1 and 2 are redundant nuclear import factors for hnRNP A1 and HuR. *RNA* 10, 590–599. doi:10.1261/rna.5224304
- Reiter, J. F., and Leroux, M. R. (2017). Genes and molecular pathways underpinning ciliopathies. *Nat. Rev. Mol. Cell Biol.* 18, 533–547. doi:10.1038/nrm.2017.60
- Rout, M. P., Aitchison, J. D., Magnasco, M. O., and Chait, B. T. (2003). Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* 13, 622–688. doi:10.1016/j.tcb.2003.10.007
- Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W., and Wittinghofer, A. (1995). Crystal structure of the nuclear Ras-related protein Ran in its GDP-bound form. *Nature* 374, 378–381. doi:10.1038/374378a0
- Schmidt, H. B., and Görlich, D. (2016). Transport selectivity of nuclear pores, phase separation, and membraneless organelles. *Trends Biochem. Sci.* 41, 46–61. doi:10.1016/j.tibs.2015.11.001
- Seewald, M. J., Körner, C., Wittinghofer, A., and Vetter, I. R. (2002). RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* 415, 662–666. doi:10.1038/415662a
- Shamsher, M. K., Ploski, J., and Radu, A. (2002). Karyopherin beta 2B participates in mRNA export from the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 99, 14195–14196. doi:10.1073/pnas.212518199
- Siomi, H., and Dreyfuss, G. (1995). A nuclear localization domain in the hnRNP A1 protein. *J. Cell Biol.* 129, 551–560. doi:10.1083/jcb.129.3.551
- Siomi, M. C., Eder, P. S., Kataoka, N., Wan, L., Liu, Q., and Dreyfuss, G. (1997). Transportin-mediated nuclear import of heterogeneous nuclear RNP proteins. *J. Cell Biol.* 138, 1181–1192. doi:10.1083/jcb.138.6.1181
- Soniati, M., Sampathkumar, P., Collett, G., Gizzi, A. S., Banu, R. N., Bhosle, R. C., et al. (2013). Crystal structure of human Karyopherin β 2 bound to the PY-NLS of *Saccharomyces cerevisiae* Nab2. *J. Struct. Funct. Genom.* 14, 31–35. doi:10.1007/s10969-013-9150-1
- Soniati, M., and Chook, Y. M. (2016). Karyopherin- β 2 recognition of a PY-NLS variant that lacks the proline-tyrosine motif. *Structure* 24, 1802–1809. doi:10.1016/j.str.2016.07.018
- Soniati, M., Gağatay, T., and Chook, Y. M. (2016). Recognition Elements in the Histone H3 and H4 Tails for Seven Different Importins. *J. Biol. Chem.* 291, 21171–21183. doi:10.1074/jbc.M116.730218
- Springhower, C. E., Rosen, M. K., and Chook, Y. M. (2020). Karyopherins and condensates. *Curr. Opin. Cell Biol.* 64, 112–123. doi:10.1016/j.ccb.2020.04.003
- Stewart, M., Kent, H. M., and McCoy, A. J. (1998). The structure of the Q69L mutant of GDP-Ran shows a major conformational change in the switch II loop that accounts for its failure to bind nuclear transport factor 2 (NTF2). *J. Mol. Biol.* 284, 1517–1527. doi:10.1006/jmbi.1998.2204
- Strehlow, A., Hallegger, M., and Jantsch, M. F. (2002). Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol. Biol. Cell* 13, 3822–3835. doi:10.1091/mbc.E02-03-0161
- Suárez-Calvet, M., Neumann, M., Arzberger, T., Abou-Ajram, C., Funk, E., Hartmann, H., et al. (2016). Monomethylated and unmethylated FUS exhibit increased binding to Transportin and distinguish FTLD-FUS from ALS-FUS. *Acta Neuropathol.* 131, 587–604. doi:10.1007/s00401-016-1544-2
- Süel, K. E., Gu, H., and Chook, Y. M. (2008). Modular organization and combinatorial energetics of proline-tyrosine nuclear localization signals. *PLoS Biol.* 6, e137. doi:10.1371/journal.pbio.0060137
- Sun, Y., Zhao, K., Xia, W., Feng, G., Gu, J., Ma, Y., et al. (2020). The nuclear localization sequence mediates hnRNP A1 amyloid fibril formation revealed by cryoEM structure. *Nat. Commun.* 11, 6349. doi:10.1038/s41467-020-20227-8
- Suzuki, M., Iijima, M., Nishimura, A., Tomozoe, Y., Kamei, D., and Yamada, M. (2005). Two separate regions essential for nuclear import of the hnRNP D nucleocytoplasmic shuttling sequence. *FEBS J.* 272, 3975–3987. doi:10.1111/j.1742-4658.2005.04820.x
- Tai, L. R., Chou, C. W., Lee, I. F., Kirby, R., and Lin, A. (2013). The quantitative assessment of the role played by basic amino acid clusters in the nuclear uptake of human ribosomal protein L7. *Exp. Cell Res.* 319, 367–375. doi:10.1016/j.yexcr.2012.12.007
- Tessier, T. M., Dodge, M. J., Prusinkiewicz, M. A., and Mymryk, J. S. (2019). Viral appropriation: laying claim to host nuclear transport machinery. *Cell* 8, 13. doi:10.3390/cells8060559
- Tran, E. J., and Wente, S. R. (2006). Dynamic nuclear pore complexes: life on the edge. *Cell* 125, 1041–1053. doi:10.1016/j.cell.2006.05.027
- Truant, R., Kang, Y., and Cullen, B. R. (1999). The human tap nuclear RNA export factor contains a novel transportin-dependent nuclear localization signal that lacks nuclear export signal function. *J. Biol. Chem.* 274, 32167–32171. doi:10.1074/jbc.274.45.32167
- Twyffels, L., Wauquier, C., Soin, R., Decaestecker, C., Gueydan, C., and Kruys, V. (2013). A masked PY-NLS in *Drosophila* TIS11 and its mammalian homolog tristetraprolin. *PLoS One* 8, e71686. doi:10.1371/journal.pone.0071686
- Twyffels, L., Gueydan, C., and Kruys, V. (2014). Transportin-1 and Transportin-2: protein nuclear import and beyond. *FEBS Lett.* 588, 1857–1868. doi:10.1016/j.febslet.2014.04.023
- Verhey, K. J., Dishinger, J., and Kee, H. L. (2011). Kinesin motors and primary cilia. *Biochem. Soc. Trans.* 39, 1120–1125. doi:10.1042/BST0391120
- Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J., and Wittinghofer, A. (1999). Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature* 398, 39–46. doi:10.1038/17969

- Wang, X., Xu, X., Zhu, S., Xiao, Z., Ma, Z., Li, Y., et al. (2012). Molecular dynamics simulation of conformational heterogeneity in transportin 1. *Proteins* 80, 382–397. doi:10.1002/prot.23193
- Weighardt, F., Biamonti, G., and Riva, S. (1995). Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J. Cell Sci.* 108 (Pt 2), 545–555.
- Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* 112, 441–451. doi:10.1016/s0092-8674(03)00082-5
- Weis, K. (2007). The nuclear pore complex: oily spaghetti or gummy bear?. *Cell* 130, 405–407. doi:10.1016/j.cell.2007.07.029
- Wente, S. R., and Rout, M. P. (2010). The nuclear pore complex and nuclear transport. *Cold Spring Harb. Perspect. Biol.* 2, a000562. doi:10.1101/cshperspect.a000562
- Xu, D., Farmer, A., and Chook, Y. M. (2010). Recognition of nuclear targeting signals by Karyopherin- β proteins. *Curr. Opin. Struct. Biol.* 20, 782–790. doi:10.1016/j.sbi.2010.09.008
- Yamauchi, Y., and Greber, U. F. (2016). Principles of virus uncoating: cues and the snooker ball. *Traffic* 17, 569–592. doi:10.1111/tra.12387
- Yamauchi, Y. (2020). Influenza A virus uncoating. *Adv. Virus Res.* 106, 1–38. doi:10.1016/bs.aivir.2020.01.001
- Yoo, S., Myszka, D. G., Yeh, C., McMurray, M., Hill, C. P., and Sundquist, W. I. (1997). Molecular recognition in the HIV-1 capsid/cyclophilin A complex. *J. Mol. Biol.* 269, 780–95. doi:10.1006/jmbi.1997.1051
- Yoshizawa, T., Ali, R., Jiou, J., Fung, HYJ, Burke, K. A., Kim, S. J., et al. (2018). Nuclear import receptor inhibits phase separation of FUS through binding to multiple sites. *Cell* 173, 693–e22. doi:10.1016/j.cell.2018.03.003
- Zhang, Z. C., and Chook, Y. M. (2012). Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). *Proc. Natl. Acad. Sci. U.S.A.* 109, 12017–12021. doi:10.1073/pnas.1207247109
- Zhang, S. B., Lin, S. Y., Liu, M., Liu, C. C., Ding, H. H., Sun, Y., et al. (2019). CircAnks1a in the spinal cord regulates hypersensitivity in a rodent model of neuropathic pain. *Nat. Commun.* 10, 4119. doi:10.1038/s41467-019-12049-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Mboukou, Rajendra, Kleinova, Tisné, Jantsch and Barraud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Résumé

Une caractéristique des cellules eucaryotes est la séparation physique du matériel génétique de la machinerie de traduction. Le transport de macromolécules entre le noyau et le cytoplasme est un processus efficace orchestré par des récepteurs de transport. La Transportine-1 (Trn1) est un récepteur d'import nucléaire impliqué dans la reconnaissance d'une classe particulière de NLS, les PY-NLS. Cependant, plusieurs protéines cargos transportées par Trn1 ne possèdent pas de PY-NLS, et leur reconnaissance est mal caractérisée. C'est le cas de l'enzyme d'édition d'ARN, ADAR1 dont le NLS dépendant de Trn1 est atypique et formé de deux modules qui adoptent une conformation fonctionnelle grâce au repliement d'un des domaines de liaison à l'ARN double-brin (dsRBD) d'ADAR1. Jusqu'à présent, la compréhension du fonctionnement de ce NLS atypique reposait sur un modèle basé sur des données de modélisations qui suggéraient que ce dernier ne peut être reconnu par Trn1 que lorsqu'il n'est pas lié à des ARNs substrats.

L'objectif de mes travaux de thèse est la validation expérimentale de ce modèle par des données structurales. D'une part, la résolution de la structure cristallographique du dsRBD portant le NLS d'ADAR1 en complexe avec un ARN double-brin a montré que ce domaine dsRBD se lie à l'ARN de façon canonique sans réarrangement majeur de structure, validant ainsi une partie du modèle. D'autre part, les travaux réalisés sur le complexe entre le NLS d'ADAR1 et son récepteur d'import Trn1 ont permis de mieux comprendre le mode de reconnaissance de ce NLS. La combinaison de différentes techniques biophysiques a permis de montrer que le NLS d'ADAR1 ne présente pas exactement le même mode de reconnaissance que les PY-NLS. En effet, la boucle H8 de Trn1 qui est connue pour être impliquée dans le relargage des PY-NLS, est dans le cas d'ADAR1 impliquée dans la reconnaissance de son NLS atypique. Dans leur ensemble, mes travaux apportent un éclairage nouveau sur le mode de fonctionnement du NLS atypique d'ADAR1.

Mots clefs : transport nucléo-cytoplasmique, Transportine-1, ADAR1, signal de localisation nucléaire, import nucléaire, dsRBD, interaction ARN-protéine, cristallographie.

Abstract

A hallmark of eukaryotic cells is the physical separation of the genetic material from the translation machinery. The transport of macromolecules between the nucleus and cytoplasm is an effective process orchestrated by transport receptors. Transportin-1 (Trn1) is a nuclear import receptor involved in the recognition of a particular class of NLS, the PY-NLS. However, besides PY-NLSs, many Trn1 cargoes harbor NLSs that do not resemble the archetypical PY-NLS. This is the case of the RNA-editing enzyme ADAR1, the NLS of which is Trn1-dependent. This atypical NLS consists of two modules that adopt functional conformation through the folding of one of the ADAR1 double-stranded RNA-binding domains (dsRBD). Until now, the understanding of the functioning of this atypical NLS was based on a model built on modelling data, which suggested that the atypical NLS of ADAR1 can only be recognized by Trn1 when it is not associated with substrate RNAs. This would prevent the transport of ADAR1 bound to cytoplasmic RNAs into the nucleus.

The aim of my thesis was to validate experimentally this model using structural data. On the one hand, the resolution of the crystallographic structure of ADAR1 dsRBD carrying the NLS of ADAR1 in complex with a double-stranded RNA showed that this dsRBD binds to RNA in a canonical manner without structural rearrangement, thus validating one part of the model. On the other hand, my work on the complex between the ADAR1 NLS and its nuclear import receptor Trn1, has led to a better understanding of the recognition mode of this atypical NLS. The combination of different biophysical techniques showed that the NLS of ADAR1 does not have the same recognition mode as PY-NLSs. Indeed, the Trn1 H8 loop, known to be involved in the release of PY-NLSs, is in the case of ADAR1 involved in the recognition of its atypical NLS. As a whole, my work sheds new light on the functioning of the atypical NLS of ADAR1.

Keywords: transport in eukaryotic cell, Transportin-1, ADAR1, nuclear localization signal, nuclear import, dsRBD, RNA-protein interaction, crystallography.