



HAL
open science

Dynamique structurale et fonctionnelle du domaine C-terminal de la protéine PB2 du virus de la grippe A

Elise Delaforge

► **To cite this version:**

Elise Delaforge. Dynamique structurale et fonctionnelle du domaine C-terminal de la protéine PB2 du virus de la grippe A. Biologie moléculaire. Université Grenoble Alpes, 2015. Français. NNT : 2015GREAV037 . tel-01485269

HAL Id: tel-01485269

<https://theses.hal.science/tel-01485269>

Submitted on 8 Mar 2017

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Elise Delaforge

Thèse dirigée par **Martin Blackledge**
et codirigée par **Darren Hart**

préparée au sein de l'**Institut de Biologie Structurale de Grenoble**
et de l'**Ecole Doctorale Chimie et Sciences du Vivant**

Dynamique structurale et fonctionnelle du domaine C-terminal de la protéine PB2 du virus de la grippe A

Thèse soutenue publiquement le **4 décembre 2015**,
devant le jury composé de :

Dr Bruno Kieffer

Institut de Génétique et de Biologie Moléculaire et Cellulaire, Strasbourg,
Président

Dr Véronique Receveur-Brechot

Centre de Recherche en Cancérologie de Marseille, Marseille, Rapporteur

Dr Cameron Mackereth

Institut Européen de Chimie et de Biologie, Bordeaux, Rapporteur

Dr Pau Bernado

Centre de Biologie Structurale, Montpellier, Examineur

Dr Darren Hart

Unit of Virus Host Cell Interaction, Grenoble, Co-Directeur de thèse

Dr Martin Blackledge

Institut de Biologie Structurale, Grenoble, Directeur de thèse



J'aimerais remercier les Dr. Véronique Receveur-Brechot et Dr. Cameron Mac-kereth d'avoir accepté d'être juges de mon travail, prenant en charge la lourde tâche de rapporteur. J'aimerais également remercier les Dr. Pau Bernado et Dr. Bruno Kieffer, respectivement examinateur et président de mon jury.

Je tiens également à adresser un immense merci à mon directeur de thèse, Dr. Martin Blackledge, premier contact à l'IBS, qui après m'avoir offert un stage en Master 1, m'a permis de faire aboutir le projet qui a fait l'objet de ma thèse au sein de son laboratoire. Sa porte m'a toujours été ouverte, pour des discussions de qualité. C'est grâce à lui que j'ai découvert la richesse de la RMN et l'importance de la compréhension des phénomènes dynamiques, tout particulièrement par rapport aux échelles de temps observées. Cette thèse n'aurait également pas pu avoir lieu sans le concours de mon co-directeur de thèse, Dr. Darren Hart, qui m'a enseigné la nécessité de la rigueur bibliographique et du fait de connaître les expériences réalisées par les personnes ayant travaillé sur le projet par le passé. Je me suis toujours sentie à ma place dans son laboratoire, et pour ça Darren je te remercie.

Merci également aux membres de mon comité de suivi de thèse, Dr. Pau Bernado et Pr. Rob Ruigrok. Ils ont su poser les bonnes questions (pas toujours les plus simples) et m'orienter au moment où il le fallait. Merci aussi aux collaborateurs, notamment Anne Martel, Adam Round et Martha Brennich pour leur aide.

Un personnage important de cette histoire : Dr. Denis Bouvier. Post-doctorant en charge du projet à mon arrivée, merci de m'avoir supportée pendant les 4 mois de passage de flambeau, et de m'avoir tout appris sur le sujet. Ton aide m'a été extrêmement précieuse.

Je souhaite remercier les membres des différents laboratoires dont j'ai fait partie : deux labos, deux groupes bien différents mais deux endroits où je me suis sentie chez moi. La Hart team, avec Philippe, Rasa, Eva, Saijo, Sri, Paul : merci pour votre soutien permanent, les fous rires et tout le reste. Le FDP : Malene, Valery, Guillaume², Luca, Robert, Jie-Rong, Loïc, Jaka, Damien, Stefaniia, Koji, Nicola, Anton, Sigrid, Paul, Malmed, Pavel, Amir, Benjamin, Emma, Kathleen... Grâce à vous, j'ai bénéficié pendant 3 ans d'un environnement de travail des plus agréables. Vous me manquerez.

Un merci spécial à Malene, qui m'a initiée à l'utilisation des spectros, m'a soutenue face à un MARS peu coopératif, et a toujours été présente lorsque j'en avais besoin. Un merci spécial également à Luca, pour les heures passées à se battre avec moi contre les spectres Bruker mal processés, décalés (ah les 2.7 ppm...), lors de mes débuts en RMN.

Jaka, camarade de galère, compagnon de peak-picking tardif. Sans toi les heures d'analyses de données auraient été beaucoup plus longues (et le mur NMR-art n'aurait jamais existé!).

Damien et Philippe, toujours un protocole parfait dans la poche, toujours présents pour résoudre les galères de paillasse, et compagnons de pause indéfectibles : merci.

Merci au Dr. Nicolas Tarbouriech, qui m'a initiée à la paillasse il y a de ça 8 ans (ouch)... Merci de m'avoir fait découvrir le monde des protéines et de la biologie structurale, de m'avoir donné le goût de poser les bonnes questions. Tu as été une étape essentielle de mon parcours dans les laboratoires. Martial, toi aussi à ta façon, tu as fais partie de ceux qui m'ont donné le goût de la recherche, et pour cela je te dis merci.

Je veux également citer tous les autres, ceux qui ne faisaient pas partie de mes équipes d'accueil, mais qui ont été des soutiens intra-institut précieux : Sriharsha, Didier, Maxime, Marjolaine, Flore, Isabel, Isabelle, Yoan, Yann HDK, Vanessa, Ziad, Vincent, Irina, Guillaume, Samira, Caroline... Merci à eux.

Un grand merci aussi à toute l'équipe de la DIR, sans qui on ne pourrait pas faire grand chose : Sylvie, Morgane, Domi, Carole, Fabrice, Jessy, Didier, Fred, Alain², Laure, Anaïs², Blandine, Chantal, Isabelle, Nathalie, Odile, Eva, Winfried, Jacques... Et aussi l'équipe de gestion de l'EMBL/CIBB : Régis, Rokhaya, Dominique, Tanja, Annie, Audrey, Kévin... Merci à tous pour votre disponibilité, votre réactivité, et votre compréhension.

Un gigantesque merci à tout mon entourage extérieur au monde de la recherche : Defne, Sriharsha, Théo, Rix, Yannick, Thomas, Amandine, Félix, Souad, Hitoshi, Jacques, Stéphanie, Jules. Votre soutien m'a été essentiel. Merci aussi au Shamrock's, et à tous les dartistes. Merci pour votre patience lors des tournois partagés entre jeu et boulot.

Merci évidemment à ma famille, soutien de toujours et à tous les niveaux dans mes choix, même les plus surprenants. Merci de votre patience lors de mes tentatives de réponses au sempiternel "Mais c'est quoi exactement, ce que tu fais ?", du support moral que vous avez été, et pour tout le reste.

Enfin, ces remerciements ne seraient pas complets sans remercier celui qui partage ma vie. Tu y as débarqué au milieu de ma thèse, et malgré mes nombreuses mises en garde, tu n'as pas eu peur. Ton soutien inconditionnel a été essentiel. Merci d'avoir cru en moi, su être là, trouver les mots. Je t'aime.

Table des matières

I	Le virus de la grippe	13
1	Structure et génome du virus	14
1.1	Les protéines associées à la membrane	14
1.2	Les ribonucléoprotéines	16
1.3	Autres protéines	18
2	Cycle de réplication du virus	20
2.1	Entrée dans la cellule	20
2.2	Import des vRNPs au noyau	20
2.3	Réplication et transcription	21
2.4	Import au noyau de l'ARN polymérase néo-synthétisée	22
2.5	Export du noyau et bourgeonnement	23
3	Mécanismes d'évolution	23
3.1	Grippe A	23
3.2	Grippe B	24
4	Adaptation inter-espèces de la grippe A	24
II	Le domaine 627-NLS	27
1	Structure	27
2	Interaction avec l'ARN	29
3	Interaction avec l'importine α	30
3.1	Importine α et import au noyau canonique	30
3.2	Interaction du domaine NLS et de l'importine α	32
4	Mutations adaptatives	35
4.1	Résidu 627	35
4.2	Résidu 701	37
4.3	Résidu 588	37
4.4	Résidus 590/591	38
4.5	Résidu 613	38
4.6	Résidu 676	39

4.7	Résidus 661 et 684	39
4.8	Résidu 702	39
4.9	Résidu 714	39
5	Objectifs	40
III	Résonance Magnétique Nucléaire	43
1	Généralités	43
2	Résonance des noyaux à spin demi-entier	44
3	La spectroscopie par RMN	45
4	Les spectres multidimensionnels	47
5	Les paramètres observés	48
5.1	Les échelles de temps en RMN	48
5.2	Les couplages	49
5.3	Le déplacement chimique	50
5.4	La relaxation	55
5.5	L'échange chimique	58
5.6	L'augmentation paramagnétique de la relaxation	61
6	Limitations de taille	63
6.1	La deutération	63
6.2	Les séquences TROSY	64
IV	Autres techniques	67
1	Diffusion des rayons X aux petits angles	68
1.1	Principe	68
1.2	Expériences	69
1.3	Analyse des données	70
1.4	Modélisation	72
1.5	Conclusion	72
2	Diffusion des neutrons aux petits angles	73
2.1	Principe	73
2.2	Variation de contraste	73
2.3	Conclusion	74
3	Transfert d'énergie par résonance de Förster sur molécules uniques (sm-FRET)	74
3.1	Principe	74
3.2	Paramètres	76
3.3	Mesures sur molécules uniques	77
3.4	Analyse des données	78
3.5	Conclusion	80
4	Résonance des plasmons de surface	81
4.1	Principe	81
4.2	SPR en biologie	81

4.3 Conclusion	82
5 Thermophorèse à échelle microscopique	83
5.1 Principe	83
5.2 Considérations expérimentales	83
5.3 Conclusion	84
V Matériel et Méthodes	85
1 Préparation des échantillons	85
1.1 Expression et purification : 627-NLS et importine α	85
1.2 Marquage par des sondes	86
2 Biologie moléculaire	87
2.1 Clonage du 627-NLS de la grippe B	87
2.2 Mutagénèse dirigée	88
3 Mesures par résonance magnétique nucléaire	88
3.1 Expériences d'attribution	88
3.2 CEST	89
3.3 PREs	89
3.4 Relaxation	89
4 Diffusion aux petits angles	90
4.1 Des rayons X	90
4.2 Des neutrons	90
5 Transfert d'énergie par résonance de Förster	90
6 Création et sélection d'ensembles	91
7 Détermination de constantes d'affinité	92
7.1 Thermophorèse à échelle microscopique	92
7.2 Résonance des plasmons de surface	92
VI Structure et dynamique du domaine 627-NLS aviaire et de son interaction avec l'importine α	93
1 Résumé de la publication	93
1.1 Equilibre entre conformations ouvertes et fermées	93
1.2 La forme fermée est stabilisée par un pont salin	94
1.3 L'équilibre dépend de la température	95
1.4 Confirmation par SAXS	95
1.5 La conformation ouverte interagit avec l'importine α	96
1.6 Propriétés dynamiques et mode d'interaction du 627-NLS avec l'importine α par smFRET	96
1.7 Conclusion	97
2 Publication	97
3 Résultats complémentaires	140
3.1 Relaxation et temps de corrélation	140
3.2 Equilibre conformationnel du 627-NLS humain	140

3.3	La forme ouverte	142
3.4	Interaction avec l'importine α par RMN	147
3.5	Interaction avec l'importine α par smFRET	151
3.6	Etudes d'affinité	153
VII	Discussion, conclusion et perspectives	159
1	Discussion	159
1.1	Dynamique conformationnelle du 627-NLS et de son interaction avec l'importine α	159
1.2	Affinité du 627-NLS pour l'importine α	162
2	Conclusion et perspectives	165
VIII	Structure du domaine 627 de la grippe B	169
1	Introduction	169
2	Cristallographie aux rayons X	170
2.1	Cristallogénèse des protéines	170
2.2	Diffraction aux rayons X	172
2.3	Conclusion	174
3	Définition du domaine, clonage et production	174
4	Cristallisation et diffraction des rayons X	175
5	Discussion	178
6	Conclusion	180
	Annexes	181
A	Séquences des constructions utilisées	182
1	627-NLS	182
1.1	Aviaire	182
1.2	Humain	183
2	Importines α	183
B	Attributions	185
C	Alignements de séquence	193

Abréviations

ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
ARNc	Acide ribonucléique complémentaire
ARNv	Acide ribonucléique viral
BMRB	<i>Biological magnetic resonance data bank</i>
BSL	<i>Biosafety level</i>
CDC	<i>Centre for disease control</i>
CEST	<i>Chemical exchange saturation transfer</i>
CRM1	<i>Chromosome region maintenance protein 1</i>
cRNP	Ribonucléoprotéine complémentaire
CSA	<i>Chemical shift anisotropy</i>
CSP	<i>chemical shift perturbation</i>
DSS	Acide 4,4-diméthyl-4-silapentane-1-sulfonique
DTT	Dithiothréitol
EDC	Hydrochlorure de 1-éthyl-3-(3-diméthylamino-propyl)carbodiimide
ESRF	<i>European synchrotron radiation facility</i>
FID	<i>Free induction decay</i>
FRET	<i>Förster resonance energy transfer</i>
HA	Hémagglutinine
Hsp90	<i>Heat shock protein 90</i>

HSQC	<i>Heteronuclear single quantum coherence</i>
IBB	<i>Importin β binding</i>
ICTV	<i>International committee on taxonomy of viruses</i>
ILL	Institut Laue-Langevin
IMAC	<i>Ion metal affinity chromatography</i>
IPTG	Isopropyl β -D-thiogalactopyranoside
ISIS	
LB	Luria Broth
MST	<i>Microscale thermophoresis</i>
MTSL	S-(2,2,5,5-tetraméthyle-2,5-dihydro-1H-pyrrol-3-yl) méthyle méthanesulfonothioate
NA	Neuraminidase
NES	<i>Nuclear export signal</i>
NHS	N-hydroxysuccinimide
NLS	<i>Nuclear localisation signal</i>
nOe	<i>Nuclear Overhauser effect</i>
NP	Nucléoprotéine
Nup	Nucléoporine
PA	Protéine acide
PB1	Protéine basique 1
PB2	Protéine basique 2
PCR	<i>Polymerase chain reaction</i>
PDB	<i>Protein data bank</i>
PRE	<i>Paramagnetic relaxation enhancement</i>
RMN	Résonance magnétique nucléaire
RNP	Ribonucléoprotéine
RU	Unité de réponse
SANS	Diffusion des neutrons aux petits angles
SAXS	Diffusion des rayons X aux petits angles

SCK	<i>Single cycle kinetics</i>
SLD	<i>Scattering length density</i>
SPR	Résonance des plasmons de surface
TEMPO	2,2,6,6-Tetramethylpiperidine 1-oxyl
TEV	<i>Tobacco etch virus</i>
TROSY	<i>Transverse relaxation optimized spectroscopy</i>
vRNP	Ribonucléoprotéine virale

Le virus de la grippe

La grippe est un problème de santé publique. En effet, chaque hiver, elle est la cause d'importantes épidémies saisonnières qui tuent entre 250 000 et 500 000 personnes dans le monde (0,1 % de mortalité), essentiellement chez les personnes âgées [WHO, 2014]. Cette maladie est causée par le virus de la grippe, ou influenza, qui appartient à la famille des *Orthomyxoviridae*, au genre *influenza A, B* ou *C* (Figure 1.1) [ICTV, 2014]. Il infecte principalement les oiseaux et certains mammifères, dont l'homme et le porc [Noah and Krug, 2005]. La grippe A a été isolée pour la première fois en laboratoire en 1932. Cependant, la maladie est décrite dans de nombreux textes, les premiers rapports confirmés datant de 1694, et les premiers rapports non confirmés remontant jusqu'à 412 av. J.-C., par les Grecs [Potter, 2001]. Il existe des vaccins contre le virus de la grippe, cependant ce dernier est en constante évolution, obligeant à renouveler les vaccins à chaque saison [Drake, 1993]. De plus, un virus particulièrement mortel (jusqu'à 2 % de mortalité) émerge sporadiquement, généralement suite à un réassortiment avec un virus aviaire ou porcine. On observe alors des pandémies pouvant causer des dizaines de millions de morts, comme la grippe espagnole qui causa en 1918 de 20 à 50 millions de morts, surtout chez les jeunes adultes [WHO, 2012b].

Plus récemment, le virus pandémique H1N1, ou les virus aviaires H7N9 ou H5N1 (mortalité de 60 %), capables de se transmettre de l'oiseau à l'homme mais pas entre humains, ont fortement inquiété les populations et la communauté scientifique [WHO, 2012a]. Deux études publiées en 2012 dans les journaux Nature et Science identifiant les mutations nécessaires au virus de la grippe aviaire H5N1 pour se transmettre efficacement par voies aérosols entre furets (animal modèle pour l'humain) ont suscité beaucoup de réactions, notamment par peur du bioterrorisme [Imai et al., 2012, Russell et al., 2012, Osterholm and Kelley, 2012, Palese and Wang, 2012]. A tel point qu'un moratoire volontaire de 60 jours a été appliqué, suspendant toutes les recherches utilisant des virus de la grippe actifs appartenant au groupe "H5". Cette

“pause” a permis aux acteurs internationaux de la recherche sur la transmission de la grippe aviaire de se rencontrer pour préciser les règles selon lesquelles ces virus devraient être manipulés, augmentant le niveau de sécurité à BSL-3+, voire BSL-4 dans certains pays [Sleator, 2012, Fouchier et al., 2013].

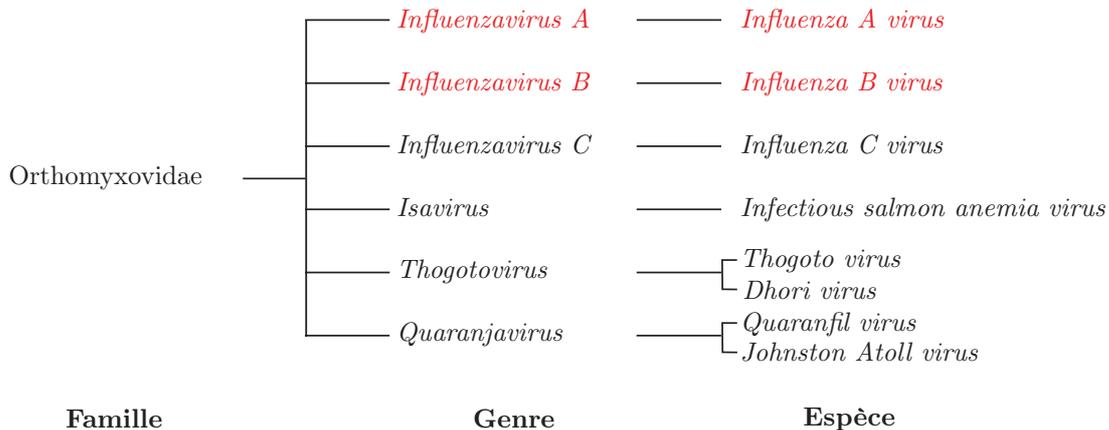


FIGURE 1.1 – Famille des Orthomyxovirus.

En rouge, les genres et espèces qui seront discutés dans ce manuscrit.

1 Structure et génome du virus

Les particules virales sont généralement sphériques, avec un diamètre de 100 nm en moyenne (Figure 1.2A) et possèdent une membrane lipidique riche en rafts [Fujiyoshi et al., 1994, Scheiffele et al., 1999]. Le génome du virus est constitué d’ARN négatif, segmenté, simple brin. Ce génome code pour 10 protéines majeures et 6 polypeptides mineurs, répartis sur 8 brins d’ARN viral (ARNv) [Fodor, 2013]. Une copie de chaque ARNv est présente dans chaque particule virale, sous forme de ribonucléoprotéines (Figure 1.2B,C).

1.1 Les protéines associées à la membrane

L’hémagglutinine (HA) est une glycoprotéine transmembranaire trimérique majeure de l’enveloppe du virus (80 %). Son domaine soluble comporte une tige et une tête, qui contient un site de reconnaissance pour les sucres situés en surface des cellules cibles du virus. Il en existe 16 types différents pour les virus de la grippe A, caractérisés par la variabilité de la tête. Cette dernière est une des cibles privilégiées du système immunitaire, puisque c’est une des parties les plus exposées du virus [Webster et al., 2014, CDC, 2015a]. De récentes avancées prometteuses de développement de vaccins protégeant contre une large gamme de virus de la grippe A visent la partie conservée (tige) de l’hémagglutinine [Yassine et al., 2015, Impagliazzo et al., 2015].

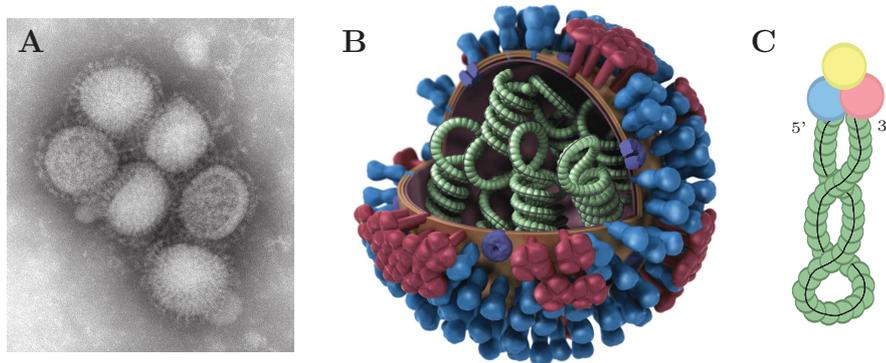


FIGURE 1.2 – Structure du virus de la grippe.

A. Cliché de microscopie électronique du virus de la grippe H1N1 (souche pandémique de 2009) [CDC, 2010]. **B.** Représentation tridimensionnelle du virus de la grippe. Les protéines de surfaces HA, NA et M2 sont en bleu, rouge et violet respectivement. La membrane lipidique est en brun et la protéine de matrice M1 en mauve. A l'intérieur du virus, les ribonucléoprotéines : chaque ARNv (noir) est protégé par des nucléoprotéines (vert). [CDC, 2010] **C.** Représentation d'une ribonucléoprotéine virale, avec les nucléoprotéines en vert, l'ARN en noir, et aux extrémités 5' et 3' de l'ARN, un trimère de polymérase (PA, PB1 et PB2).

La neuraminidase (NA) est une enzyme transmembranaire tétramérique moins abondante que HA (17 %) dont le rôle est de cliver l'acide sialique retenant les virions nouvellement synthétisés à la membrane de la cellule de l'hôte [Webster et al., 2014]. Il en existe 11 types différents pour le virus de la grippe A [CDC, 2015a]. Plusieurs antiviraux utilisés contre la grippe inhibent cette enzyme : l'oseltamivir (Tamiflu), le peramivir (Rapivab) et le zanamivir (Relenza) [FDA, 2015].

HA et NA sont ancrées dans les rafts lipidiques de l'enveloppe du virus [Skibbens et al., 1989, Kundu et al., 1996].

M1 est une protéine structurale recouvrant l'intérieur de la membrane lipidique, formant une matrice à laquelle s'attachent les ribonucléoprotéines virales. Associée à M2, elle contrôle la forme du virion. Elle interagit également avec la nucléoprotéine et la protéine d'export nucléaire du virus pour permettre l'export des ribonucléoprotéines nouvellement synthétisées [Samji, 2009, Schroeder et al., 2004].

M2 (BM2 pour la grippe B) est le produit d'épissage du gène *m1*. C'est un tétramère traversant à la fois la membrane et la couche de protéine matrice M1. C'est un canal ionique sélectif des protons qui permet l'acidification de la particule virale après son entrée dans la cellule cible. Bien que peu abondant (16 à 50 molécules par virion), son rôle est critique pour le bon déroulement du cycle viral. Il peut être bloqué par l'amantadine (Symmetrel) et la rimantadine (Flumadine) [Webs-

ter et al., 2014, FDA, 2015]. La structure de la partie transmembranaire de M2 a été résolue par résonance magnétique nucléaire (RMN) et par cristallographie aux rayons X, dans la forme fermée et la forme ouverte du canal [Pielak and Chou, 2011]. Différents sites d'interactions de la rimantadine et de l'amantadine ont été proposés à partir de mutants inhibant l'effet de ces molécules, tous situés à l'intérieur du canal formé par les segments transmembranaires de M2. Cependant, une étude par RMN, supportée par différentes études par mutagenèse et par RMN, a déterminé que l'amantadine et la rimantadine interagissent avec M2 dans une poche à l'interface lipide/protéine en C-terminal de M2 [Schnell and Chou, 2008]. Enfin, d'autres études utilisant une chimère entre le N-terminal de M2 et le C-terminal de BM2, cette dernière étant naturellement résistante à l'amantadine, ont montré que la protéine obtenue est sensible à l'amantadine. Ces résultats sont contradictoires avec l'interaction proposée à l'interface lipide/protéine. Le mécanisme d'action de l'amantadine reste donc controversé, et de futures études seront nécessaires pour établir son mécanisme d'action [Pielak and Chou, 2010]. Le domaine soluble de M2 est, quant à lui, structuré et permet une interaction spécifique avec M1 [Pielak and Chou, 2011].

M42 est un variant de M2 encodé par un autre fragment d'ARN. M42 n'est pas essentiel pour le cycle de réplication du virus, mais il permet une réplication efficace du virus en absence de M2 [Wise et al., 2012].

NB est une protéine membranaire trouvée uniquement dans le virus de la grippe B. C'est un canal ionique putatif dont la fonction reste obscure. Il est palmitoylé pour l'import à la membrane lors du bourgeonnement de virions [Demers et al., 2014].

1.2 Les ribonucléoprotéines

La nucléoprotéine (NP) est en échange lent entre un état monomérique et un état trimérique (grippe A) ou tétramérique (grippe B) lorsqu'elle n'est pas associée à l'ARN [Chenavas et al., 2013]. Seule la forme monomérique est capable de s'associer à l'ARN, interagissant avec 24 à 27 bases d'ARN via un sillon basique, protégeant l'ARN viral de la dégradation dans la cellule hôte, et formant ainsi une ribonucléoprotéine virale (vRNP) [Chenavas et al., 2013, Ye et al., 2006, Ortega et al., 2000]. Dans la RNP, NP multimérise grâce à l'insertion d'une boucle dans la molécule voisine, et est capable de former des multimères rigides ou non [Ye et al., 2006]. Les vRNPs sont ciblées par RIG-I qui les déstabilise en interagissant avec la structure secondaire "panhandle" formée par l'ARN viral 5' triphosphate, activant ainsi la réponse immunitaire de l'hôte [Weber et al., 2013].

1.2.1 L'ARN polymérase

L'ARN polymérase ARN-dépendante est un hétérotrimère, composé d'une protéine acide (PA) et de deux protéines basiques (PB1 et PB2). PA et PB1 possèdent un signal de localisation nucléaire (NLS) tandis que PB2 en possède deux, bien que l'un d'entre eux soit situé au centre de la protéine et donc inutilisable pour l'import au vu de la structure de la protéine [Boulo et al., 2007, Mukaigawa and Nayak, 1991]. Les protéines de la polymérase entrent dans le noyau via interaction avec l'importine α ou par la voie RAN-BP5 (*Ran Binding Protein 5*) [Deng et al., 2006]. La polymérase de la grippe cumule un certain nombre de fonctions essentielles à la réplication : PA comporte une activité endonucléase, PB1 porte l'activité polymérase, et PB2 contient un domaine d'interaction avec la coiffe des ARN messagers cellulaires, ainsi qu'un domaine en C-terminal, le 627-NLS, appelé domaine d'adaptation à l'hôte du fait des nombreuses mutations qui apparaissent lors de l'adaptation du virus d'un hôte aviaire à un hôte humain. Dans la particule virale, on trouve un trimère de polymérase associé aux extrémités 5' et 3' conservées de chaque ARNv [Hutchinson and Fodor, 2012, Boivin et al., 2010]. Les structures des domaines d'interaction entre les différentes sous-unités, du domaine endonucléase de PA et du domaine d'interaction avec la coiffe de PB2, et plus récemment celle de la polymérase entière en complexe avec un ARN viral, ont été résolues par cristallographie aux rayons X (Figure 1.3) [He et al., 2008, Obayashi et al., 2008, Sugiyama et al., 2009, Dias et al., 2009, Yuan et al., 2009, Guilligay et al., 2008, Reich et al., 2014, Pflug et al., 2014].

Le domaine N-terminal de PA est une endonucléase de la famille PD-(D/E)XK, coordonnant deux ions magnésium (Figure 1.3C) [Doan et al., 1999, Dias et al., 2009]. Il est capable d'hydrolyser de l'ARN simple brin et de l'ADN double brin [Dias et al., 2009]. L'inactivation de ce domaine par mutations du site actif produit une polymérase capable d'effectuer la réplication ARN-indépendante, mais incapable d'effectuer la transcription qui nécessite le clivage par l'endonucléase d'une coiffe m⁷GTP cellulaire [Yuan et al., 2009]. Le domaine C-terminal de PA permet l'interaction avec PB1 dans le contexte de la polymérase entière (Figure 1.3A) [Reich et al., 2014, Obayashi et al., 2008, He et al., 2008].

Les N-terminal et C-terminal de PB1 interagissent avec le C-terminal de PA et le N-terminal de PB2 respectivement (Figure 1.3A,B) [He et al., 2008, Obayashi et al., 2008, Sugiyama et al., 2009]. Le domaine central de PB1 porte l'activité ARN polymérase à proprement parler (motif classique S-D-D), et sa structure n'a été que très récemment caractérisée à haute résolution (Figure 1.3E) [Fodor, 2013, Reich et al., 2014]. Elle a permis de mettre en évidence les nombreux contacts entre PA et PB1 d'une part, et PB2 et PB1 d'autre part. En effet, PA et PB2 sont enroulées autour de PB1, l'ARN viral étant fixé dans un canal traversant la polymérase de

part en part et permettant l'accès au site actif de polymérisation [Reich et al., 2014]. PB1 interagit avec le 5' et le 3' de l'ARN viral, interaction nécessaire à la formation de RNPs [González and Ortín, 1999, Reich et al., 2014].

Le domaine N-terminal de PB2 permet son interaction avec le peptide C-terminal de PB1 (Figure 1.3B) [Sugiyama et al., 2009]. Le domaine central de PB2 (Figure 1.3C), dit "d'interaction avec la coiffe" [Guilligay et al., 2008], permet à la polymérase de fixer la coiffe m⁷GTP des ARN messagers cellulaires, la coupure nécessaire au vol étant effectuée par le domaine endonucléase de PA [Plotch et al., 1981, Li et al., 2001]. En effet, dans la structure de la polymérase entière, ces deux domaines sont proches l'un de l'autre. D'autre part, PA et PB2 interagissent toutes les deux avec l'ARN promoteur nécessaire pour l'initiation de la transcription, ce qui est cohérent avec leurs positions respectives dans le contexte de la polymérase entière [Fodor et al., 1993, Reich et al., 2014]. Enfin, le domaine 627-NLS, situé en C-terminal de PB2, est un double domaine localisé près du canal de sortie putatif de l'ARN synthétisé [Reich et al., 2014]. Il est également impliqué dans l'adaptation inter-espèces [Tarendeau et al., 2008, Boivin et al., 2010]. Les détails de sa structure et de sa fonction seront introduits dans le chapitre II.

Le gène *pa* code pour trois protéines mineures : PA-X est composée du domaine endonucléase de PA et d'un domaine correspondant à la traduction d'un second cadre de lecture en C-terminal. Elle réprime l'expression de gènes cellulaires [Jagger et al., 2012]. PA-N155 et PA-N182 sont des formes tronquées en N-terminal (à l'acide aminé 155 ou 182) de PA. Elles ne sont pas suffisantes pour la répllication mais jouent probablement un rôle significatif dans ce processus, les virus qui ne les expriment pas se répliquant plus lentement [Muramoto et al., 2013]

Le gène *pb1* code également pour deux protéines mineures : N40, une forme tronquée en N-terminal grâce à un codon start alternatif, qui ne possède pas le domaine d'interaction avec PA, et dont la fonction n'est pas élucidée [Wise et al., 2009]. PB1-F2 correspond à un décalage (+1) du cadre de lecture du gène, donnant une protéine complète de 87 acides aminés (facteur de virulence à l'effet pro-apoptotique) chez un hôte aviaire, mais devenant un court peptide (58 acides aminés) du fait de mutations lors de l'adaptation à un hôte humain [Chen et al., 2001, McAuley et al., 2010].

1.3 Autres protéines

NS1 est un facteur de virulence impliqué dans l'évasion du système immunitaire inné, dont les rôles sont multiples. C'est un dimère comportant un domaine de dimérisation et un domaine effecteur, capable d'interagir avec divers partenaires,

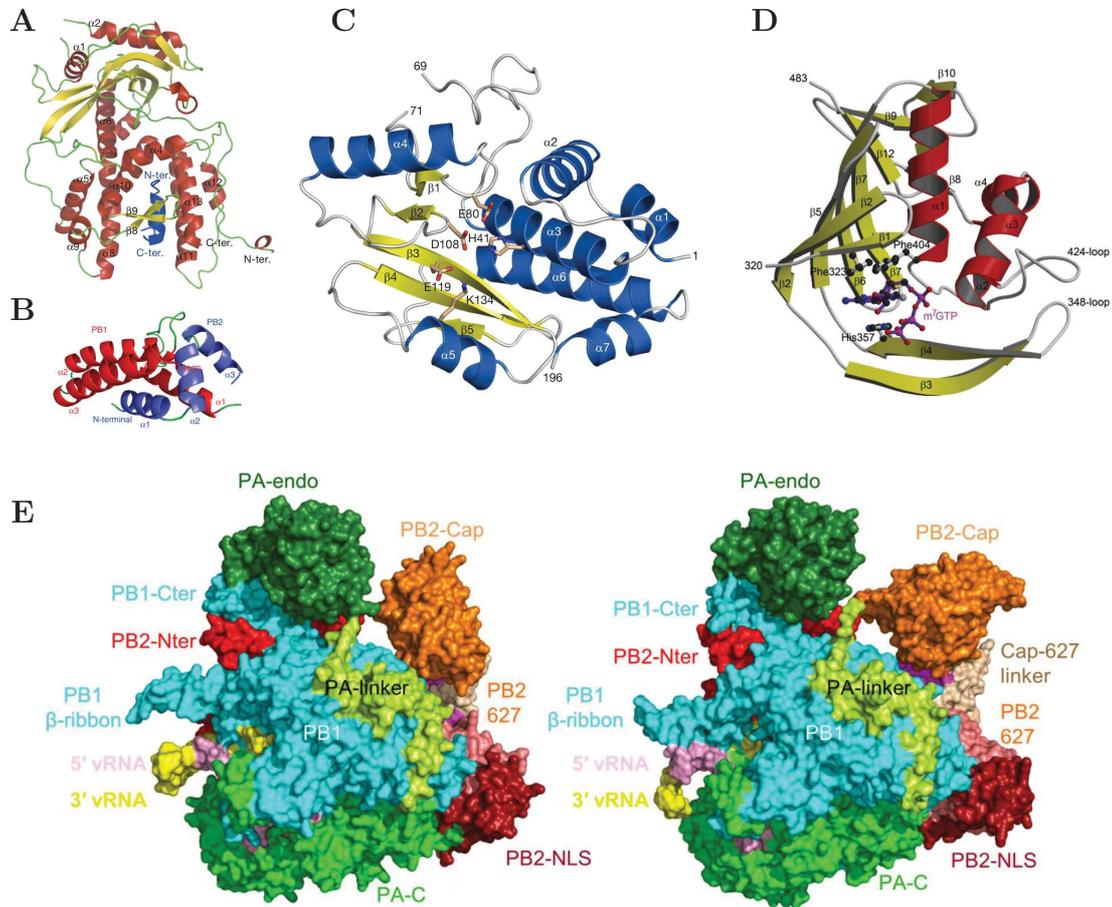


FIGURE 1.3 – Structures cristallographiques de l’ARN polymérase ARN dépendante de la grippe. **A.** Domaine d’interaction en “tête de dragon” entre le C-terminal de PA (hélices α en rouge, feuillets β en jaune, boucles en vert) et le N-terminal de PB1 (bleu). Adapté de [Obayashi et al., 2008]. **B.** Domaine d’interaction entre le C-terminal de PB1 (rouge) et le N-terminal de PB2 (bleu). Adapté de [Sugiyama et al., 2009]. **C.** Domaine N-terminal de PA, portant l’activité endonculéase (hélices α en bleu, feuillets β en jaune, boucles en gris). En bâtonnets, les acides aminés du site actif. Adapté de [Dias et al., 2009]. **D.** Domain central de PB2, portant le site d’interaction avec la coiffe, en complexe avec un m⁷GTP (hélices α en rouge, feuillets β en jaune, boucles en gris). En bâtonnets, les acides aminés réalisant l’interaction ainsi que le ligand. Adapté de [Guilligay et al., 2008]. **E.** Représentation en surface du trimère PA/PB1/PB2 d’une souche de la grippe B (à gauche) ou A (à droite) en complexe avec un ARN viral. Chaque domaine est coloré comme son nom l’indique. La différence majeure entre ces deux structures est l’orientation du domaine d’interaction avec la coiffe, en orange. Adapté de [Reich et al., 2014].

tels que de l'ARN ou la phosphoinositide 3-kinase [Hale et al., 2010, Cheng et al., 2009]. NS1 augmente le niveau d'expression des gènes codant pour des cytokines. Cependant, NS1 inhibe le traitement de l'extrémité 3' des ARN pré-messagers de la cellule en interagissant avec des facteurs cruciaux de ce processus, empêchant la cellule de traduire ces ARN en protéines. NS1 interfère également avec la machinerie d'épissage de la cellule, forçant la relocalisation des ARN messagers au noyau [Noah and Krug, 2005, Lu et al., 1994].

NS2, aussi appelée NEP (*Nuclear Export Protein*) est le produit d'épissage du gène *ns1*. NEP joue un rôle essentiel dans l'export des vRNPs hors du noyau, grâce à un signal d'export nucléaire (NES), permettant l'export du noyau médié par l'exportine CRM1 (*Chromosome Region Maintenance protein 1*). De ce fait, une petite quantité de protéines NEP se trouve également associée aux vRNPs dans les virions produits [Iwatsuki Horimoto et al., 2004]. NEP interagit également avec M1 via un domaine en épingle à cheveux hélicale amphipatique, qui dimérise pour former un ensemble de quatre hélices [Akarsu et al., 2003].

Des protéines de l'hôte sont également incorporées dans les virions. On peut notamment trouver des protéines impliquées dans la structure de la cellule et le trafic intra-cellulaire (actine β , tubuline α , famille des tropomyosines et cofilines), mais également des enzymes impliquées dans la glycolyse (aldolase, pyruvate kinase, glyceraldéhyde-3-phosphate déshydrogénase, ...). Bien qu'une partie le soit de façon non spécifique, des similarités entre les protéines incluses par différents virus suggèrent l'implication de ces protéines dans certaines étapes du cycle du virus [Shaw et al., 2008]. D'autre part, certaines protéines de l'hôte ont un rôle dans le cycle de réplication du virus, comme PABP1 et eIF4GI qui s'associent avec NS1 pour l'initiation de la traduction des ARN messagers viraux, ou encore PA qui interagit avec des protéines de la famille des facteurs de transcription, augmentant l'activité de la polymérase [Burgui et al., 2003, Huarte et al., 2001].

2 Cycle de réplication du virus

2.1 Entrée dans la cellule

Lors de l'infection, le virus de la grippe entre dans les cellules cibles (cellules épithéliales) par endocytose grâce à la reconnaissance par HA des acides sialiques terminaux des glycoprotéines ou glycolipides présents en surface de ces cellules (Figure 1.4-1). Suite à la baisse de pH dans la vésicule d'endocytose (pH 5-6), le processus de fusion de la membrane virale avec celle de l'endosome est déclenché (Figure 1.4-2). HA subit alors un changement conformationnel important, permettant d'amener son peptide de fusion à proximité de son attachement dans la membrane virale. Le

peptide de fusion s'insère dans la membrane endosomale et la fusion a lieu par un mécanisme proche de celui des protéines SNARE : rapprochement des deux membranes, fusions des feuillettes externes, puis internes [Skehel and Wiley, 2000, Stegmann, 2000]. Dans le même temps, l'acidification est répercutée à l'intérieur du virus par l'entrée de protons à travers le canal M2, induisant un changement conformationnel de la protéine M1 et permettant ainsi la libération dans le cytoplasme des vRNPs (Figure 1.4-3) [Pinto and Lamb, 2006, Okada et al., 2003].

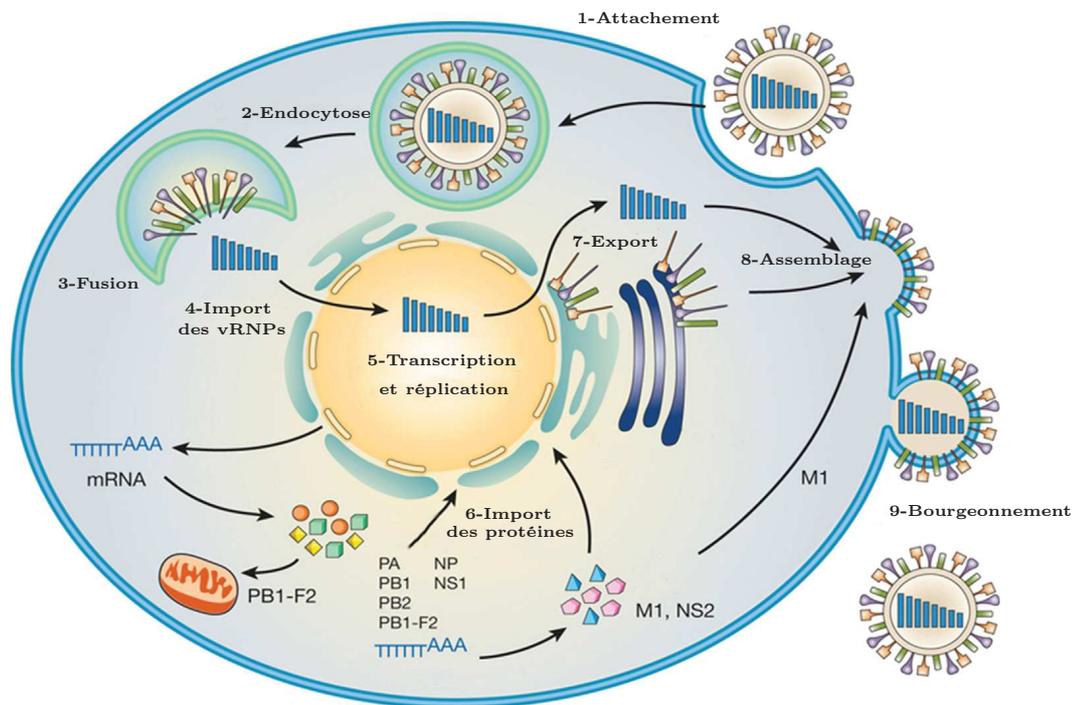


FIGURE 1.4 – Cycle de réplication du virus de la grippe.

Après reconnaissance de récepteurs en surface de la cellule cible (étape 1), le virus est endocyté (étape 2) puis relâche son contenu dans la cellule infectée (étape 3). Les vRNPs sont importées au noyau (étape 4) puis la transcription et la réplication commencent (étape 5). Les protéines néo-synthétisées sont ré-importées au noyau pour amplifier la transcription et la réplication (étape 6). Les vRNPs nouvellement synthétisés sont exportés dans le cytoplasme (étape 7) et tous les composants du virus sont assemblés à la membrane (étape 8). Enfin, le virus bourgeoine et est libéré hors de la cellule (étape 9). Adapté de [Sediri et al., 2015].

2.2 Import des vRNPs au noyau

Les vRNPs sont ensuite importées dans le noyau (Figure 1.4-4). Les différentes protéines composant les vRNPs contiennent chacune au moins un NLS, et une étude récente propose que des importine α interagissent avec les NLS des NP pour l'import

au noyau. Une mutation permettant l'exposition du NLS de PB2 (D701N) augmenterait l'efficacité de cet import en favorisant une orientation qui facilite le passage des vRNPs à travers le pore nucléaire [Tarendeau et al., 2007, Sediri et al., 2015].

2.3 Réplication et transcription

Une fois dans le noyau, la polymérase va effectuer deux tâches principales : la réplication et la transcription (Figure 1.4-5). La réplication requiert la synthèse d'ARN positif complémentaire (ARNc), à partir duquel les ARNv sont de nouveau transcrits. Ces deux actions sont effectuées par la polymérase virale, et ne requièrent pas d'amorce, une boucle - commune à toutes les polymérases des virus à ARN - contenant un résidu aromatique servant à stabiliser l'interaction des premiers nucléotides [Reich et al., 2014]. L'ARNc et l'ARNv produits forment avec des NP des cRNPs et vRNPs, potentiellement avec l'aide du domaine 627 situé à la sortie du canal de synthèse de l'ARN [Reich et al., 2014]. Il a été proposé que ces étapes soient effectuées en "trans", c'est à dire que la polymérase synthétisant l'ARNc ou ARNv ne soit pas celle qui est attachée à la vRNP ou cRNP du brin d'ARN modèle [Fodor, 2013].

La transcription est, quant à elle, effectuée en "cis", par la polymérase attachée à la vRNP. Une première étape consiste à voler une coiffe m⁷GTP des ARN pré-messagers de la cellule hôte. PB2 reconnaît la coiffe, et l'activité endonucléase de PA permet son vol [Plotch et al., 1981, Li et al., 2001]. Le brin d'ARN coiffé volé (10-15 nucléotides) est ensuite amené au site actif de la polymérase et sert d'amorce pour l'initiation de la transcription de l'ARNv en ARN messenger viral [Sikora et al., 2014]. La polyadénylation de cet ARNm a lieu grâce à une séquence poly-U située en 5' de l'ARNv. L'ARNm est ensuite pris en charge par la machinerie cellulaire, et les protéines virales sont produites dans le cytosol [Fodor, 2013].

D'autres rôles de l'importine α ont été proposés au-delà de l'import nucléaire. Par exemple, plusieurs études suggèrent que l'importine α , interagissant avec les sous-unités de la polymérase, recrute d'autres cofacteurs permettant la régulation de l'activité de la polymérase. L'importine α pourrait également agir seule comme cofacteur de la réplication et de la transcription à l'intérieur du noyau [Resa Infante et al., 2008, Hudjetz and Gabriel, 2012, Yasuda et al., 2012].

2.4 Import au noyau de l'ARN polymérase néo-synthétisée

Les protéines néo-synthétisées sont ensuite ré-importées au noyau. PA et PB1 sont importées ensemble sous la forme d'un hétérodimère via la voie non canonique Ran-BP5, alors que PB2 et NP sont importées séparément, via l'interaction de leur signal

NLS avec l'importine α (Figure 1.4-6) [Deng et al., 2006, Huet et al., 2010, Hutchinson and Fodor, 2012]. Cependant, il a été proposé que des dimères de PB1/PB2 ou PB1/PA soient importés au noyau par Hsp90, suggérant également que cette chaperonne soit impliquée dans la formation d'un trimère actif dans le noyau [Naito et al., 2007]. Les mécanismes d'assemblage de la polymérase à l'intérieur du noyau restent mal compris, cependant il a été montré que PB2 seul peut s'accumuler efficacement dans le noyau, contrairement à l'hétérodimère PA/PB1 [Huet et al., 2010]. En effet, ce dernier requiert non seulement la présence de PB2 pour s'accumuler efficacement au noyau, mais également que PB2 soit capable d'être importé au noyau, ce qui a été démontré à l'aide de mutants dans le NLS de PB2 [Huet et al., 2010]. Cependant, des mutations du signal NLS de PB2 n'empêchent pas le bon assemblage de la polymérase, mais donnent naissance à des vRNPs peu actives, suggérant que la fonction de l'interaction de PB2 avec l'importine α n'est pas seulement l'import au noyau, mais également de chaperonner la formation de vRNPs actives [Huet et al., 2010, Resa Infante et al., 2008]. Cette proposition est cohérente avec l'hypothèse selon laquelle l'interaction de PB2 avec l'importine α induit une réorganisation du complexe de la polymérase et participe activement à la formation de vRNPs actives [Boivin and Hart, 2011]. Il a été proposé qu'un intermédiaire de formation du trimère soit l'interaction de PA/PB1 toujours couplé à Ran-BP5 avec l'ARN viral, l'interaction de PB2 étant alors promue par le départ de Ran-PB5 lors de son association à Ran-GTP [Deng et al., 2006]. NP comporte deux signaux de localisation nucléaire, cependant un seul est exposé dans sa structure tridimensionnelle, suggérant que c'est celui-ci qui permet l'import au noyau des NPs [Ye et al., 2006]. D'autre part, il a été proposé que l'importine $\alpha 5$ serve de chaperonne à NP, lui permettant de rester monomérique jusqu'à la formation de RNPs [Boulo et al., 2011]. Les vRNPs nouvellement assemblées effectuent ensuite les cycles de transcription et réplication suivants [Hutchinson and Fodor, 2012].

2.5 Export du noyau et bourgeonnement

Une fois les vRNPs synthétisées, elles doivent être exportées hors du noyau pour former de nouveaux virions. M1 interagit avec les RNPs et avec NEP, cette dernière interagissant à son tour avec CRM1, permettant ainsi l'export des vRNPs par la voie CRM1/Ran (Figure 1.4-7) [Boulo et al., 2007].

Enfin, la dernière étape du cycle de réplication du virus est de former de nouvelles particules virales (Figure 1.4-8). Pour ce faire, HA, NA et M2 sont exportées à la membrane apicale de la cellule grâce à un signal contenu dans leur segment transmembranaire. HA et NA s'insèrent dans des rafts lipidiques et M2 interagit avec du cholestérol [Schroeder et al., 2004]. M1 est également exportée, probablement grâce à son interaction avec la queue cytoplasmique de HA et NA, et multimérise au contact

des membranes lipidiques [Hilsch et al., 2014]. Il a été montré que les vRNPs interagissent avec les filaments d’actine et s’associent avec les rafts lipidiques, il est donc probable que ce soit de cette façon que les vRNPs et M1 associées soient dirigées vers le site de bourgeonnement du virus [Nayak et al., 2009]. Le bourgeonnement est initié par la courbure du feuillet externe de la membrane, causée par la composition particulière en lipides du site de bourgeonnement et par l’interaction, dans ou sous la membrane, des protéines HA, NA, M1 et M2. La croissance du bourgeon est ensuite due à la force exercée par les vRNPs associées à M1. Puis, le bourgeon est refermé grâce à la protéine M2 qui amène à proximité des régions lipidiques “non-raft” au niveau du site de fermeture [Schroeder et al., 2004]. Enfin, le virion, encore attaché à la membrane cellulaire par l’interaction avec un acide sialique, est clivé par NA et relâché dans l’environnement extracellulaire (Figure 1.4-9) [Nayak et al., 2009].

3 Mécanismes d’évolution

3.1 Grippe A

Deux mécanismes principaux d’évolution sont utilisés par la grippe A : la dérive antigénique et le réassortiment.

La dérive antigénique correspond à l’accumulation de mutations ponctuelles. En effet, la polymérase de la grippe, comme toutes les ARN polymérases, n’effectue ni relecture ni correction du brin synthétisé. Ainsi, un taux de mutation supérieur à $7,3 \cdot 10^{-5}$ par paire de base et de plus de 0,99 paire de base par génome est observé [Drake, 1993]. Ces mutations sont généralement observées sur HA et NA, et abolissent l’immunité acquise lors des précédentes infections, les cibles majeures des anticorps acquis étant ces deux protéines [Treanor, 2004].

Le réassortiment est un mécanisme beaucoup plus drastique par lequel, lors d’une co-infection par plusieurs virus, des virions contenant des segments d’ARNv de différentes souches sont produits. Le réassortiment a généralement lieu chez un hôte pouvant être infecté par différents virus, comme le porc. En effet, la surface de ses cellules épithéliales possède des acides sialiques liés aux glycoprotéines en $\alpha(2,3)$ et $\alpha(2,6)$. Les virus humains reconnaissent les liaisons $\alpha(2,6)$ (bronches supérieures humaines) alors que les virus aviaires reconnaissent les liaisons $\alpha(2,3)$ (intestins aviaires, bronches inférieures humaines). Les porcs possèdent les deux types de récepteurs, faisant d’eux l’hôte idéal pour les réassortiments entre virus humains, aviaires et porcins, comme ce fut le cas pour la souche pandémique H1N1 de 2009 [Skehel and Wiley, 2000, Vijaykrishna et al., 2010]. C’est de cette façon qu’émergent les pandémies observées à travers l’histoire : le virus produit étant très différent des virus saisonniers, la réaction du système immunitaire est beaucoup plus forte.

L'origine historique de ces pandémies est souvent située en Asie car le réassortiment y est facilité par la proximité des animaux et des humains, couplée à la densité de population élevée [Potter, 2001].

3.2 Grippe B

La grippe B a été isolée pour la première fois lors d'une épidémie hivernale à New-York en 1940 [Nerome et al., 2014]. Deux lignées principales ont depuis été identifiées : B/Victoria, prédominante dans les années 80, et B/Yamagata, prédominante dans les années 90. Depuis le début des années 2000, les deux lignées co-dominent [Ni et al., 2013]. La particularité de la grippe B est qu'elle n'infecte principalement que l'homme, bien que des infections aient été recensées chez le phoque [Bodewes et al., 2013]. Elle ne possède qu'un seul type d'hémagglutinine (HA), et évolue par dérive antigénique et par délétions/insertions systématiques de nucléotides, sur un cycle de 50 ans [Nerome et al., 2014]. La grippe B fait partie des souches causant des épidémies hivernales, et une, voire deux, de ses souches font partie de celles incluses dans les vaccins annuels [CDC, 2015b].

4 Adaptation inter-espèces de la grippe A

Lors de son adaptation d'un hôte aviaire à un hôte humain, le virus de la grippe doit s'adapter d'une part aux protéines de l'hôte, différentes, et d'autre part à la température, puisque chez l'oiseau la grippe se réplique dans l'environnement chaud de l'intestin (40°C) alors que chez l'homme elle se réplique dans les voies respiratoires, plus fraîches (environ 33°C). La transmission inter-espèces du virus de la grippe A est également un problème de santé publique, puisque de nouveaux virus à mortalité potentiellement élevée pour l'homme pourraient ainsi émerger, causant des pandémies. En effet, aucune immunité naturelle (innée ou acquise) n'existant contre ces virus, toute la population humaine y est alors sensible. De façon intéressante, seuls les virus possédant les types H1, H2, H3 et N1, N2, ont causé des pandémies par le passé et infectent de façon régulière les humains lors des épidémies hivernales [Kilbourne, 2006]. Cependant, des infections sporadiques par H5N1, H7N3, H7N7 et H9N2 sont recensées de façon régulière, nourrissant l'inquiétude d'une possible pandémie [Lin et al., 2000].

Le rôle de HA et NA dans l'adaptation inter-espèces est bien compris, cependant il est récemment devenu clair que les protéines composant la polymérase sont également impliquées [Naffakh et al., 2000, Chen et al., 2006, Finkelstein et al., 2007, Naffakh et al., 2008]. La majorité des mutations observées lors de ce phénomène se trouvent sur le domaine C-terminal de PB2, le double domaine 627-NLS (Figure 1.5). Le 627-NLS est aussi appelé "domaine d'adaptation à l'hôte". Il contient

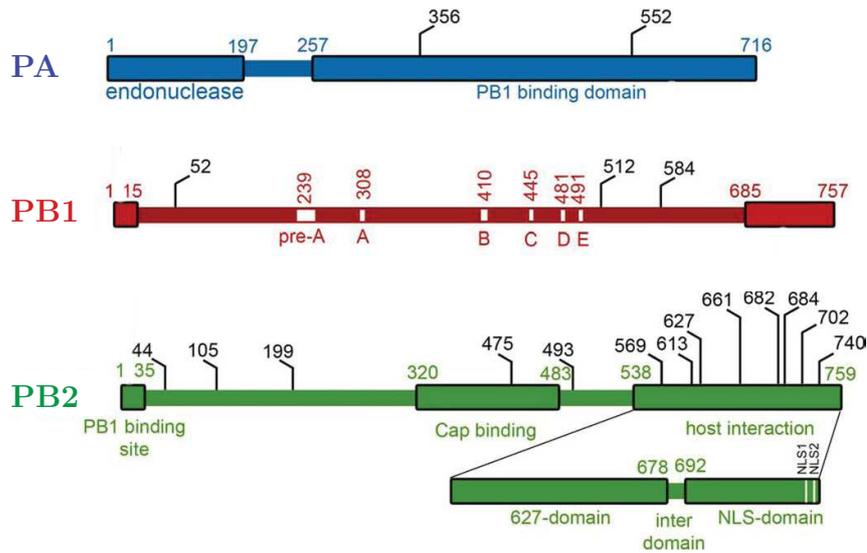


FIGURE 1.5 – Mutations de la polymérase lors de l’adaptation inter-espèces.
 Représentation schématique des protéines de la polymérase virale. Les mutations impliquées dans l’adaptation inter-espèce sont indiquées en noir [Tamuri et al., 2009]. Adapté de [Boivin et al., 2010].

le résidu 627 qui a été nommé “résidu déterminant l’hôte” du fait que la seule mutation E627K permet au virus aviaire de s’adapter à la température de répliation d’un hôte humain [Massin et al., 2001]. Au niveau structural, l’effet principal de cette mutation est de rompre (E) ou de renforcer (K) l’homogénéité d’une partie extrêmement basique de la surface du domaine, modulant potentiellement une interaction avec un partenaire de l’hôte [Tarendeau et al., 2008].

C’est donc au domaine 627-NLS que nous avons choisi de nous intéresser lors de cette thèse. Historiquement dans le laboratoire, et de façon à pouvoir étudier l’effet des mutations évolutives, deux souches ont été choisies : A/duck/Shantou/4610/2003 (H5N1) pour la grippe aviaire et A/Victoria/3/1975 (H3N2) pour la grippe humaine. Ces deux souches présentent dix mutations dans le domaine 627-NLS, identifiées lors d’études de l’adaptation inter-espèces [Tamuri et al., 2009, Miotto et al., 2008].

Le domaine 627-NLS

Le domaine 627-NLS est un double domaine de 24 kDa situé en C-terminal de la protéine PB2, des acides aminés 538 à 759. Il est composé du domaine 627 (538-678), d'un *linker* (679-688), et du domaine NLS (689-759). Ce dernier comporte en C-terminal un signal NLS bipartite (736-759), utilisé par PB2 pour l'import au noyau (Figure 2.1A).

1 Structure

La structure cristallographique du domaine 627-NLS a été résolue en 2008 (Figure 2.1B) [Tarendeau et al., 2008]. Elle a révélé que le domaine 627 est composé d'un coeur de 6 hélices α , recouvert sur un côté de 5 courts brins β . Le domaine NLS est, quant à lui, composé de deux courtes hélices α et de trois brins β [Tarendeau et al., 2007]. Le peptide NLS est désordonné. Le *linker* entre les domaines 627 et NLS est, pour sa part, partiellement ordonné, participant à l'interface entre les deux domaines, notamment par la formation d'un pont salin entre E687 et R650 [Tarendeau et al., 2008]. La rupture de ce pont salin par la mutation R650A mène à une perte d'activité totale de la polymérase [Kirui et al., 2014].

Les structures de diverses souches et divers mutants du 627-NLS de la grippe A ont depuis été résolues, présentant toutes des structures similaires (RMSD $< 1 \text{ \AA}$), y compris dans le contexte de la structure de la polymérase entière (Figure 2.1C) [Reich et al., 2014, Yamada et al., 2010, Smith et al., 2011, Kuzuhara et al., 2009, Tarendeau et al., 2008].

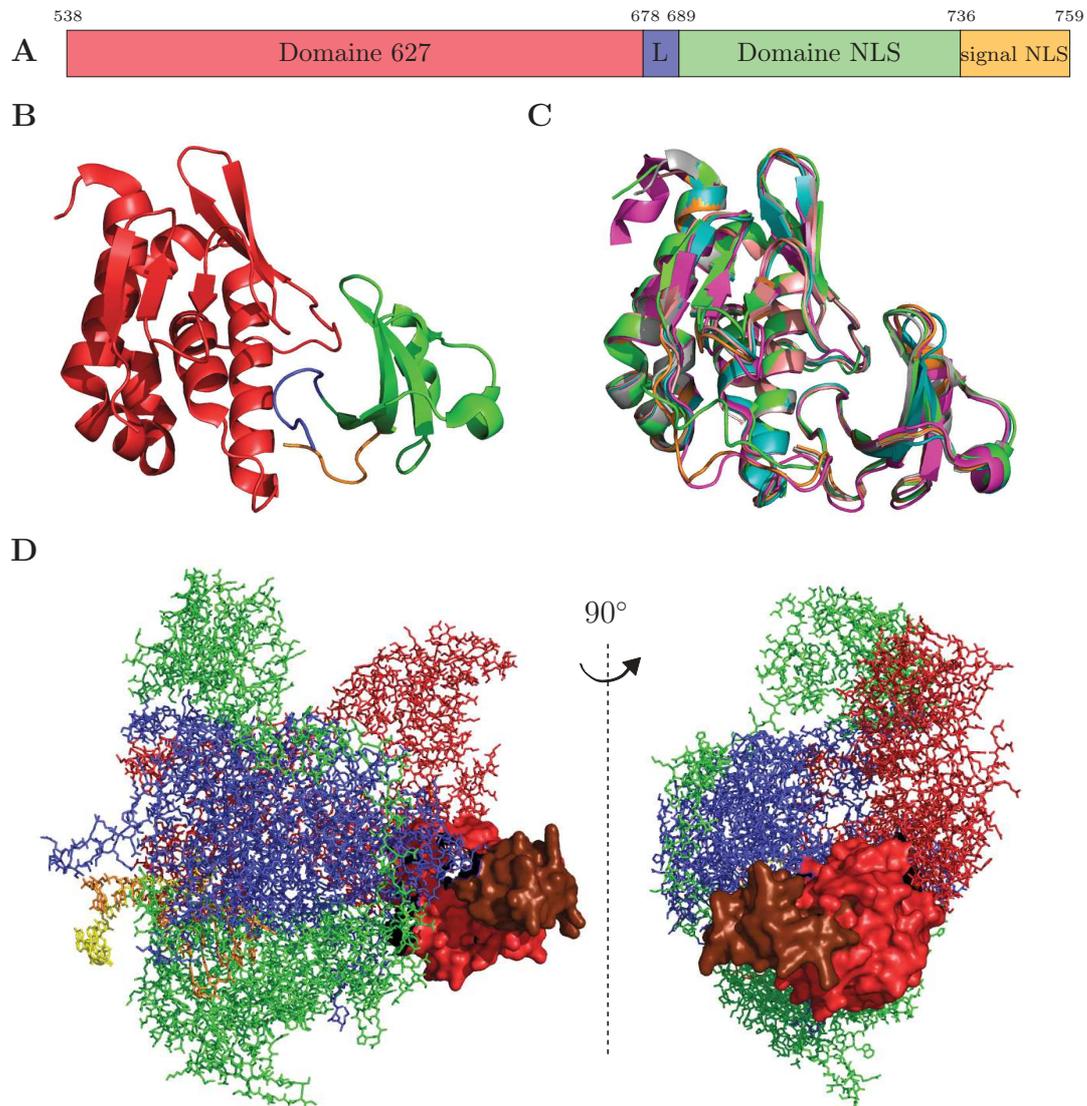


FIGURE 2.1 – Structure du domaine 627-NLS.

A. Représentation de la structure primaire du 627-NLS et de ses domaines (L : *linker*). **B.** Structure cristallographique du 627-NLS, les différentes parties sont colorées de la même façon qu'en A. PDB 2VY6 [Tarendeau et al., 2008] **C.** Superposition des structures du 627-NLS de différentes souches sur 2VY6 en gris. A/little yellow-shouldered bat/Guatemala/060/2010 (H17N10) PDB 4WSB (vert), A/mexico/indre4487/2009 (H1N1) PDB 3KHW (orange), A/vietnam/1203/2004 (H5N1) PDB 3L56 (rose), A/Yokohama/2017/03 PDB 3R2V (H3N2) (bleu), A/PuertoRico/8/1934 (H1N1) PDB 3CW4 (magenta) [Reich et al., 2014, Yamada et al., 2010, Smith et al., 2011, Kuzuhara et al., 2009]. **D.** Structure de la polymérase du virus de la grippe de la souche A/little yellow-shouldered bat/Guatemala/060/2010 (H17N10), PDB 4WSB [Reich et al., 2014]. PA (vert), PB1 (bleu), PB2 (rouge) et l'ARN viral 5' et 3' (orange et jaune respectivement) sont représentés en bâtonnets. Le domaine C-terminal de PB2 est mis en évidence en surface, avec en rouge le domaine 627 et en marron le domaine NLS.

2 Interaction avec l'ARN

L'interaction du domaine 627 ou 627-NLS avec de l'ARN a été caractérisée dans deux études, par résonance des plasmons de surface (SPR) pour l'une, et par anisotropie de fluorescence et RMN pour l'autre [Kuzuhara et al., 2009, Lim et al., 2014].

Par SPR, les auteurs ont pu montrer que l'affinité du domaine 627-NLS humain (souche H3N2) pour le 5' de l'ARNv est de 3,2 μM , et de 12 et 8,8 μM respectivement pour les primers coiffés et non-coiffés. Ces résultats suggèrent l'implication du domaine 627-NLS dans l'initiation de la réplication et de la transcription virale. Ils ont également déterminé les constantes d'affinité du domaine 627-NLS humain portant la mutation K627E, mutant qui rompt l'homogénéité de la surface basique du domaine 627, pour les mêmes ARN (Figure 2.2A). L'affinité observée est diminuée de façon significative, et n'a pu être déterminée pour les ARN coiffés ou non. Pour le 5' de l'ARNv, l'affinité est diminuée d'un facteur 3 ($K_d = 8,9 \mu\text{M}$). Il est proposé que ces différences contribuent à la pathogénicité des virus aviaires portant la mutation E627K chez l'homme [Kuzuhara et al., 2009].

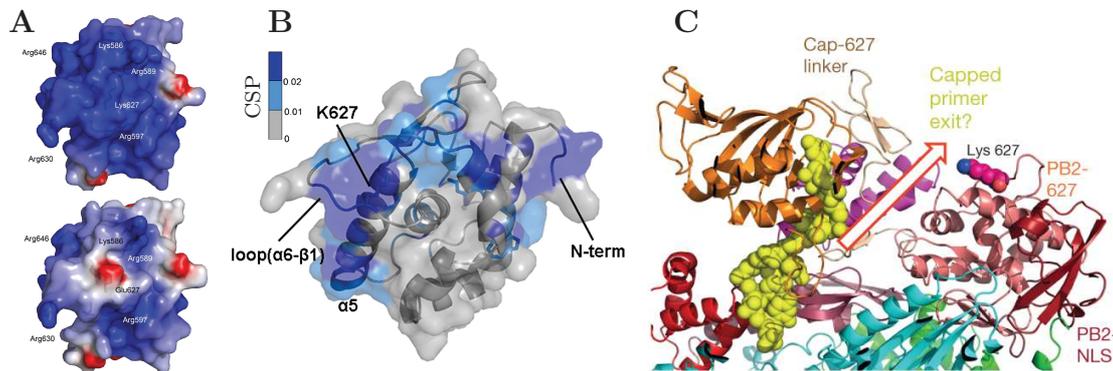


FIGURE 2.2 – Interaction du 627-NLS avec l'ARN.

A. Représentation du domaine 627-NLS en surface colorée selon la charge (de rouge à bleu, de négatif à positif). Lorsque le résidu 627 est une lysine (en haut), la surface est basique, lorsque le résidu 627 est un acide glutamique (en bas), l'homogénéité de la surface est rompue. Adapté de [Tarendeau et al., 2008]. **B.** Zone d'interaction de l'ARNv 5' déterminé par RMN. Les couleurs représentent l'intensité de la perturbation du déplacement chimique lors de l'interaction, de gris (faible) à bleu foncé (forte). Adapté de [Lim et al., 2014] **C.** Position du 627-NLS dans le contexte de la polymérase entière. La flèche rouge dénote le canal de sortie putatif du primer coiffé, et plus généralement des ARNs synthétisés. Adapté de [Reich et al., 2014].

Une récente étude a permis une meilleure caractérisation de l'interaction entre le domaine 627 humain (souche H1N1) et le 5' de l'ARNv. Les auteurs ont déterminé

une constante d'affinité de 10,5 μM par anisotropie de fluorescence et ont pu localiser l'interaction de l'ARN avec le domaine 627 par RMN (Figure 2.2B). Ils ont d'autre part montré qu'*in vivo*, les mutants rompant l'homogénéité de la surface basique (E576A, K627E, R630A, R630D), et donc causant une perte d'affinité pour l'ARN, réduisent l'activité de la polymérase d'environ 90 %. Les auteurs concluent à l'implication du domaine 627 dans l'initiation de la réplication par l'interaction avec l'ARNv 5' [Lim et al., 2014].

Cependant, ces deux publications ne présentent aucun contrôle concernant l'interaction avec un ARN quelconque. En effet, le domaine 627 est une surface positive (Figure 2.2A). Le squelette de l'ARN étant chargé négativement, une interaction est attendue, quelle que soit la séquence de l'ARN. Depuis, la structure atomique de la polymérase entière a été résolue, et place le domaine 627 à proximité du canal de sortie des ARNs synthétisés, formant un canal basique avec le domaine d'interaction avec la coiffe, à proximité du résidu 627 (Figure 2.2C). Ce canal est le canal putatif de sortie du primer coiffé [Reich et al., 2014]. Cette configuration est cohérente avec les résultats décrits ci-dessus, et les affinités trouvées, assez faibles, compatibles avec une interaction transitoire lors de la sortie de l'ARN néo-synthétisé. De plus, il a été montré que NP peut interagir avec PB2 au niveau du domaine 627 [Ng et al., 2012]. Cette interaction couplée à l'interaction transitoire de l'ARN sur le domaine 627 suggère que l'assemblage des nouvelles vRNPs et cRNPs a lieu à proximité et avec l'aide du domaine 627. Cependant, ces données ne permettent pas d'expliquer que la mutation E627K permette au virus de s'adapter à la température de réplication chez un hôte humain [Massin et al., 2001]. L'intérêt évolutif de l'acide glutamique à la position 627 pour les virus aviaires, diminuant fortement l'affinité du domaine 627 pour l'ARN, reste donc à être élucidé.

3 Interaction avec l'importine α

3.1 Importine α et import au noyau canonique

L'importine α est une protéine de la famille des karyophérines. Son rôle est de transporter au noyau des cargos, reconnus par le biais d'une séquence peptidique basique spécifique, KRxR, le signal de localisation nucléaire (NLS). Ce dernier peut être bipartite, avec un NLS majeur et un NLS mineur ; sa séquence est alors KR_{x12-15}KRxR [Dingwall and Laskey, 1991]. L'importine α est constituée de répétitions de domaines armadillo en tandem, et possède en N-terminal un peptide d'interaction avec l'importine β , l'IBB. Ce dernier ressemble à un NLS et possède également un rôle d'auto-inhibition, se fixant sur l'importine α en l'absence de cargo [Chook and Blobel, 2001]. Brièvement, l'interaction d'un cargo avec l'importine α permet la libération de l'IBB, et l'interaction de ce dernier avec l'importine β permet

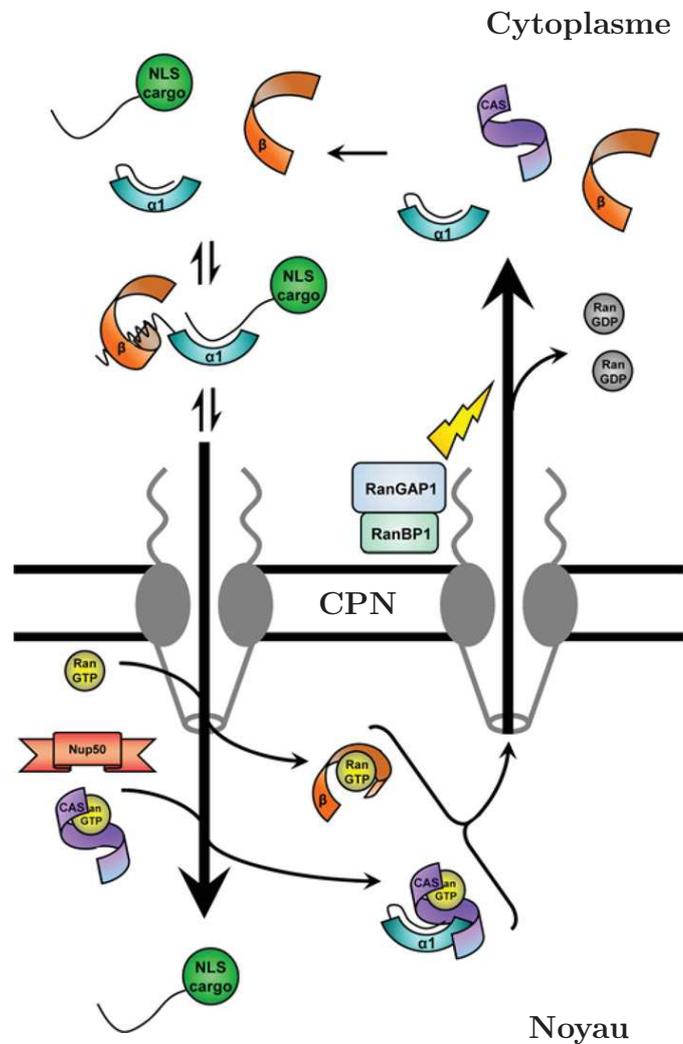


FIGURE 2.3 – Import nucléaire via la voie des importines.

Dans le cytoplasme, le cargo interagit avec l'importine α qui à son tour interagit avec l'importine β . Le complexe est importé au noyau à travers le complexe du pore nucléaire (CPN). Dans le noyau, l'importine β se dissocie du complexe grâce à Ran-GTP, et l'importine α se dissocie du cargo grâce au domaine IBB, Nup50 et Ran-GTP. Enfin, les importines sont transportées au cytoplasme et libérées pour un nouveau cycle d'import après hydrolyse de Ran-GTP en Ran-GDP. Adapté de [Pumroy and Cingolani, 2015].

la formation d'un hétérodimère, portant le cargo. Grâce à l'interaction de l'importine β avec le pore nucléaire par le biais de liaisons hydrophobes avec les répétitions FG des nucléoporines (Nup), l'ensemble est importé au noyau. Une fois dans le noyau, l'importine β subit un changement de conformation grâce à son interaction avec Ran-GTP, ce qui cause sa dissociation de l'importine α . Le cargo est ensuite relâché de l'importine α par compétition avec son IBB et avec l'aide de Nup50 et d'un facteur de recyclage, CAS, associé à Ran-GTP. Les importines α et β sont alors transportées dans le cytoplasme, et dissociées de Ran-GTP après son hydrolyse en Ran-GDP (Figure 2.3) [Pumroy and Cingolani, 2015].

Il existe 7 isoformes de l'importine α chez les vertébrés, répartis en 3 sous-familles [Pumroy and Cingolani, 2015]. Dans ce manuscrit seront étudiés les isoformes 1 (sous-famille $\alpha 1$), 3 (sous-famille $\alpha 2$), et 5 et 7 (sous-famille $\alpha 3$). Ces isoformes sont diversifiés, avec 26 % d'identité et 42 % de conservation de leur séquence en acides aminés. Différents cargos, cellulaires ou viraux, reposent spécifiquement sur différents isoformes d'importine α pour leur transport au noyau. Les différences de régulation de ces isoformes commencent seulement à être comprises [Alvisi and Jans, 2015].

3.2 Interaction du domaine NLS et de l'importine α

3.2.1 Structure

La structure du domaine NLS humain en complexe avec l'importine $\alpha 5$ a été résolue par cristallographie aux rayons X [Tarendeau et al., 2007]. Elle a révélé que le domaine NLS interagit avec l'importine $\alpha 5$ via son peptide NLS bipartite (Figure 2.4A). Trois des quatre acides aminés du signal NLS mineur ($^{736}\text{KRKR}^{739}$) et le NLS majeur ($^{752}\text{KRIR}^{755}$) interagissent fortement via leur chaîne latérale avec l'importine α , dans des poches présentes à la surface de cette dernière. Des liaisons hydrogènes entre les azotes des chaînes latérales de tryptophanes de l'importine α et les groupements carboxyles de la chaîne principale du NLS viennent compléter l'interaction. Enfin, la chaîne latérale de K718 est impliquée dans trois liaisons hydrogènes avec les résidus G284 (squelette), N286 et T325 de l'importine $\alpha 5$ [Tarendeau et al., 2007].

Les acides aminés 742 à 747 et 758 et 759 ne sont pas visibles dans la structure cristallographique, suggérant qu'ils sont flexibles dans le contexte du complexe. La chaîne latérale K736 est engagée dans des liaisons hydrogènes intra-domaine fortes, corrélant l'observation selon laquelle cet acide aminé n'est pas indispensable à l'import nucléaire [Mukaigawa and Nayak, 1991]. Enfin, le domaine replié NLS ne subit pas de changement conformationnel entre sa forme libre et sa forme liée à l'importine α . Cependant, dans la forme libre, le peptide NLS est replié sur la surface du

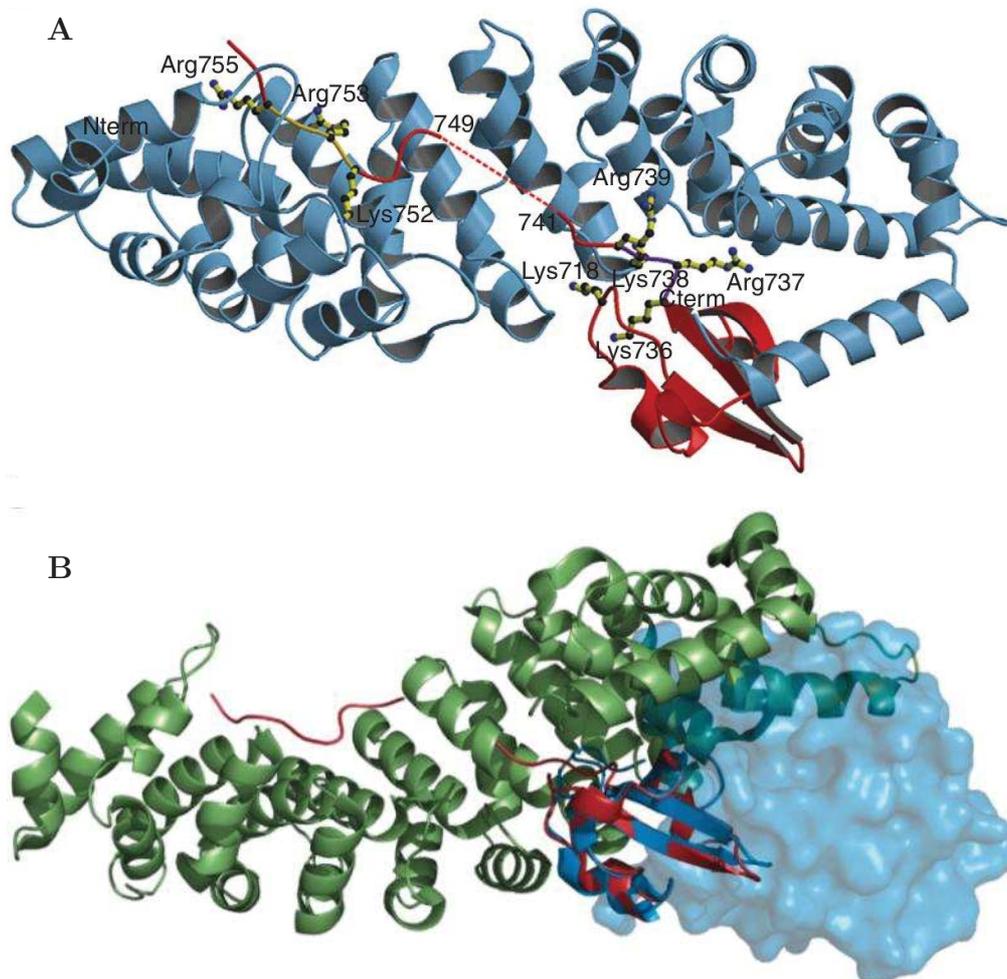


FIGURE 2.4 – Structure cristallographique de l’importine $\alpha 5$ en complexe avec le domaine NLS du virus de la grippe A (souche humaine).

A. Le domaine NLS (rouge) interagit avec l’importine $\alpha 5$ (bleu) via son signal NLS bipartite (acides aminés représentés en bâtons et boules jaunes et noirs) [Tarendeau et al., 2007] **B.** Superposition de la structure du complexe importine $\alpha 5$ /NLS (vert et rouge respectivement, PDB 2JDQ) et du double domaine 627-NLS (bleu, domaine 627 en surface, PDB 2VY6) via le domaine NLS commun.

NLS via un pont salin entre D701 et R753 [Tarendeau et al., 2007]. Ce dernier doit être rompu pour permettre l'interaction avec l'importine α [Boivin and Hart, 2011].

Cette structure est cependant incompatible avec la structure du 627-NLS complet résolue précédemment (Figure 2.4B). En effet, la superposition des deux structures via l'alignement du domaine NLS met en évidence un problème d'encombrement stérique entre le domaine 627 et le C-terminal de l'importine α , suggérant l'existence d'une autre forme du 627-NLS permettant l'interaction avec l'importine α [Boivin and Hart, 2011, Tarendeau et al., 2008].

3.2.2 Affinité

Les affinités des domaines NLS et 627-NLS de la grippe aviaire ou humaine pour les importines humaines $\alpha 1$, $\alpha 3$, $\alpha 5$, $\alpha 7$ ou aviaire $\alpha 1$ ont été mesurées par SPR [Boivin and Hart, 2011]. Toutes les affinités mesurées sont comprises entre 1 et 10 nM, cependant certaines différences sont notables.

Les auteurs ont tout d'abord étudié les différences d'affinité du domaine NLS seul aux importines $\alpha 1$. Le NLS humain a une affinité de 3,8 nM pour l'importine $\alpha 1$ humaine, mais une affinité de 1 nM pour l'importine $\alpha 1$ aviaire. Les mutations évolutives R702K et S714R semblent augmenter la vitesse d'association du NLS à l'importine $\alpha 1$, humaine ou aviaire. La mutation adaptative D701N, acide aminé impliqué dans un pont salin avec le peptide NLS, présente dans les deux cas une association plus rapide mais également une dissociation plus rapide, et donc une affinité finalement moins forte.

Les auteurs se sont ensuite intéressés à l'interaction du double domaine 627-NLS aviaire ou humain avec les importines $\alpha 1$. Ici, les affinités trouvées ne dépendent pas de l'espèce d'origine de l'importine $\alpha 1$. Cependant, le 627-NLS humain présente une affinité légèrement moins forte pour les importines α que le 627-NLS aviaire, avec 6,9 et 7,1 nM contre 1,4 et 3,2 nM pour l'importine $\alpha 1$ de la même espèce ou non respectivement.

Enfin, les auteurs ont identifié une interaction préférentielle (3 à 4 fois plus forte) des 627-NLS aviaire ou humain avec l'importine $\alpha 3$ humaine et $\alpha 1$ aviaire comparé aux importines $\alpha 1$, $\alpha 5$ et $\alpha 7$ humaines.

De façon générale, les résultats obtenus par les auteurs suggèrent que lorsque le domaine NLS seul est testé, l'affinité est maximale (comparable à un signal NLS seul). Cependant, pour le double domaine 627-NLS humain, les affinités sont plus faibles que pour le domaine NLS seul, suggérant un effet négatif du domaine 627 sur

l'interaction. Cependant, cet effet est faible, suggérant que l'encombrement stérique prédit à partir de la superposition des structures cristallographiques disponibles ne nuit pas de façon importante à l'interaction en solution. D'autre part, cet effet n'est pas visible pour le 627-NLS aviaire qui présente une affinité aux importines α similaire, que le domaine 627 soit présent ou non.

Une récente étude structurale du complexe formé par l'importine α humaine et le domaine NLS a confirmé que la présence du domaine IBB inhibe totalement ou partiellement l'interaction à haute affinité du domaine NLS avec les importines $\alpha 1$ ou $\alpha 7$ respectivement [Kobe, 1999, Pumroy et al., 2015]. De façon surprenante, cette inhibition n'est pas observée pour l'importine $\alpha 3$. En effet, les auteurs ont pu montrer grâce à des expériences de co-immunoprécipitation semi-quantitative, que dans le cas de cet isoforme, l'interaction aurait lieu indépendamment de la présence de l'IBB [Pumroy et al., 2015]. D'autre part, leurs résultats suggèrent que l'importine $\alpha 3$ est l'isoforme d'interaction préférentiel lorsque le domaine globulaire NLS est présent, ce dernier apportant un avantage topologique comparé au peptide NLS seul [Pumroy et al., 2015]. D'autres préférences du 627-NLS pour certains isoformes de l'importine α ont été identifiées dans de nombreuses études *in vivo*, et dépendent des souches utilisées et des mutations évolutives présentes (Revue par [Gabriel and Fodor, 2014]). De façon générale, la protéine PB2 issue de virus adaptés à l'homme a une préférence pour l'importine $\alpha 7$ tandis que celle issue de virus aviaires a une préférence pour l'importine $\alpha 3$. PB2 issue de virus H1N1, potentiellement en cours d'adaptation, est capable d'interagir avec les deux importines [Gabriel et al., 2011].

4 Mutations adaptatives

Un nombre important de mutations adaptatives aux mammifères de la polymérase a été recensé depuis les premières identifications de virus aviaires virulents infectant l'humain. Je ne développerai ici que les mutations faisant partie du domaine 627-NLS. Les domaines 627-NLS aviaire et humain utilisés lors de cette thèse sont des protéines existant naturellement, qui ont été sélectionnées pour présenter 10 mutations adaptative simultanément, majoritairement situées en surface du 627-NLS (Figure 2.5). Certaines ont été identifiées expérimentalement (A588I, E627K, A661T, T676I, A684S et R702K), tandis que d'autres proviennent d'études bio-informatiques analysant les mutations présentes dans des virus aviaires s'étant adaptés à l'homme (D567N, A588I, V613T, A661T, A674T et G682S) [Miotto et al., 2008, Tamuri et al., 2009]. Cependant, la comparaison de souches humaines et aviaires révèle généralement une majorité mais pas la totalité de ces mutations.

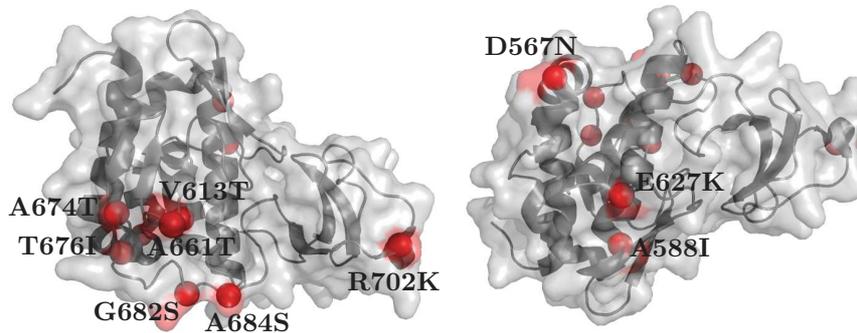


FIGURE 2.5 – Mutations entre le 627-NLS aviaire et humain. La structure cristallographique du 627-NLS est représentée en surface et en ruban (gris). Les mutations existant entre les deux souches utilisées sont mises en évidence en rouge et annotées. Vue de dessus (gauche) ou de dessous (droite). PDB 4WSB.

4.1 Résidu 627

Ce résidu a été le premier identifié comme responsable de la sensibilité à la température des virus humains, capables de se répliquer efficacement à 33°C mais pas à 37°C [Massin et al., 2001], et est considéré comme le résidu déterminant de l’hôte du virus. En effet, à quelques rares exceptions près, le résidu 627 d’un virus aviaire est toujours un acide glutamique, tandis que celui d’un virus mammifère est toujours une lysine. Deux exceptions sont notables, où une lysine persiste à cette position dans un virus infectant les autruches, ainsi que dans un virus épidémique en 2005 (virus du lac Quinghai) [Shinya et al., 2009, Long et al., 2013]. Ces souches ont été considérées comme inquiétantes car la présence d’une lysine à la position 627 peut être considérée comme une pré-adaptation aux mammifères, y compris à l’homme.

La mutation E627K a été observée dans des isolats de virus H1, H3, H5, H9 et H10 de virus aviaires infectant des humains et est la signature d’un virus adapté à l’homme [Gabriel and Fodor, 2014]. Lorsqu’un virus H5N1 contenant une nucléoprotéine aviaire et PB2 627K est passé dans des souris, l’acide aminé 627 est rapidement muté en acide glutamique, revenant à sa signature aviaire. Ce n’est pas le cas pour un virus contenant une nucléoprotéine humaine, suggérant que la répllication requiert une nucléoprotéine de la même origine que PB2 [Bogs et al., 2011]. La mutation E627K augmente également la mobilité de la polymérase dans le noyau [Foeglein et al., 2011], ainsi que l’interaction avec la nucléoprotéine [Labadie et al., 2007]. Elle augmente l’activité de la polymérase chez les mammifères pour les virus H1, H3, H5, H7 et H9 et le recrutement d’ARN promoteur [Gabriel and Fodor, 2014, Paterson et al., 2014]. Elle augmente la virulence pour les virus de type H5 chez les souris [Fornek et al., 2009, Shinya et al., 2004]. Enfin, elle améliore également la transmission du virus par contact chez les cochons d’Inde et par voies aériennes chez les cochons d’Inde et les furets [Steel et al., 2009, Herfst et al., 2010].

Malgré d’extensives recherches, le rôle de la mutation E627K dans l’adaptation d’un virus aviaire à un hôte humain est encore mal compris. En effet, un certain nombre d’effets lui sont attribués, mais les mécanismes moléculaires sous-jacents restent inconnus, et deux hypothèses s’opposent dans la littérature. Un groupe suggère que la présence d’un acide glutamique à la position 627 dans les virus humains, et donc l’existence d’une surface basique avec seulement le résidu 627 chargé négativement, permette l’interaction du domaine 627 avec un cofacteur humain, inhibant ainsi l’activité de la polymérase aviaire [Mehle and Doudna, 2008]. Un autre groupe propose au contraire que la présence d’un acide glutamique à la position 627 d’un virus aviaire permette à PB2 d’interagir avec un cofacteur augmentant l’activité de la polymérase dans les cellules aviaires. Dans les cellules humaines, ce facteur n’existerait pas, et la mutation du résidu 627 en lysine serait sélectionnée car elle permettrait de compenser l’absence du cofacteur aviaire en créant une interaction avec un cofacteur humain, permettant lui aussi d’augmenter l’activité de la polymérase [Moncorgé et al., 2010].

Au niveau de l’interaction avec l’importine α , la mutation E627K augmente l’interaction des vRNPs avec l’importine $\alpha 1$ et $\alpha 7$ chez les mammifères et réduit l’activité de la polymérase et la réplication dans les cellules n’exprimant pas ces deux isoformes d’importines α [Hudjetz and Gabriel, 2012]. La virulence des virus portant cette mutation, mais pas des virus aviaires, est réduite chez les souris double KO pour le gène de l’importine $\alpha 7$ [Gabriel et al., 2011]. Ces études ont mené à suggérer l’utilisation de l’importine $\alpha 7$ comme cible thérapeutique dans le cadre d’une pandémie [Resa Infante et al., 2015].

4.2 Résidu 701

La mutation D701N est décrite comme l’une des mutations essentielles d’adaptation aux hôtes mammifères et a été identifiée dans des isolats de virus H5 et H7 [Gabriel and Fodor, 2014]. Comme la mutation E627K, la mutation D701N augmente l’activité de la polymérase ainsi que la mobilité de la polymérase dans le noyau [Zhou et al., 2013, Foeglein et al., 2011]. Elle améliore également l’import au noyau des vRNPs [Sediri et al., 2015]. Dans le cas d’un virus portant un acide glutamique à la position 627 (signature aviaire), la présence de la mutation D701N permet d’augmenter la transmission du virus par contact chez les cochons d’Inde et par voie aérienne chez les cochons d’Inde et les furets [Steel et al., 2009, Gao et al., 2009, Herfst et al., 2010].

Au niveau de l’interaction avec les importines, la mutation D701N diminue légèrement l’affinité à l’importine $\alpha 1$ humaine, augmentant sa vitesse d’association mais également sa vitesse de dissociation [Boivin and Hart, 2011]. *In vivo*, D701N aug-

mente l'interaction de PB2 avec l'importine $\alpha 1$ dans les cellules de mammifères mais pas dans les cellules aviaires [Gabriel et al., 2008]. Comme D701N fait partie de la signature d'adaptation aux mammifères, comme pour la mutation E627K, l'effet observé pour l'interaction avec les importines α est une interaction préférentielle avec l'importine $\alpha 7$, ainsi qu'une virulence réduite des virus portant cette mutation chez les souris double KO pour le gène de l'importine $\alpha 7$ [Gabriel et al., 2011]. Une hypothèse de mécanisme moléculaire a été décrite, suggérant que le pont salin existant entre D701 (domaine NLS) et R753 (peptide NLS) est rompu lorsque l'acide aminé 701 devient une asparagine, facilitant l'interaction avec l'importine α [Tarendeau et al., 2007].

4.3 Résidu 588

La mutation A588I fait partie des premières à apparaître lors de la transmission d'un hôte aviaire à un hôte humain [Miotto et al., 2008]. Cette mutation a été observée dans un grand nombre de virus aviaires H1N1 infectant l'homme, cependant, dans la souche H1N1 pandémique de 2009, il s'agissait d'une thréonine, ce qui est rare. Lorsque cet acide aminé est muté en isoleucine, l'efficacité de la réplication en cellules de mammifères et chez la souris est augmentée. La présence d'une isoleucine à cette position augmente également l'interaction avec MAVS (*mitochondrial antiviral signaling protein*), réduisant l'expression d'interférons β des cellules infectées [Zhao et al., 2014]. D'autres études ont montré que la combinaison de 588T avec 147T et 339T ou 636F et 701N permet d'améliorer l'activité de la polymérase aussi efficacement que la seule mutation E627K [Fan et al., 2014, Foeglein et al., 2011]. Enfin, des expériences de FRAP (*fluorescence recovery after photobleaching*) ont permis de démontrer que la mutation 588I augmente la mobilité de la polymérase dans le noyau [Foeglein et al., 2011].

4.4 Résidus 590/591

La double mutation 590S/591R est extrêmement rare. Elle a été identifiée dans le virus H1N1 pandémique de 2009. En effet, dans ce virus, l'acide aminé 627 est un acide glutamique (signature aviaire) mais il se réplique de façon extrêmement efficace chez l'homme [Mehle and Doudna, 2009]. Comme expliqué précédemment, la présence d'une lysine à la position 627 (signature humaine) permet à la surface du domaine 627 d'être entièrement basique. Ce caractère est rompu par la présence d'un acide glutamique. Cependant, la présence de la double mutation 590S/591R permet de limiter l'effet de l'acide glutamique à la position 627 et de conserver une surface majoritairement basique, expliquant l'efficacité de réplication de H1N1 pandémique chez l'homme. L'addition de la mutation E627K au double mutant 590S/591R permet d'augmenter encore l'activité de la polymérase puisque la surface du domaine 627 est rendue de nouveau entièrement basique [Mehle and Doudna,

2009]. La mutation 591R seule serait suffisante pour obtenir une réplication efficace et une virulence importante chez la souris [Yamada et al., 2010].

4.5 Résidu 613

Ce résidu a été identifié comme étant le second à muter lors de l'adaptation d'un virus d'un hôte aviaire à un hôte humain. Il s'agit d'une valine qui mute en thréonine, avec comme intermédiaire une alanine ou parfois une isoleucine [Miotto et al., 2008, Tamuri et al., 2009]. Cependant, aucune fonction n'a jamais été proposée pour cette mutation dans des études *in vivo*. Il est intéressant de noter que, de toutes les mutations discutées ici et incluses dans les domaines étudiés lors de cette thèse, V613T est la seule à ne pas être en surface de la protéine. L'acide aminé 613 est situé dans le domaine 627, sous ses brins β , sa mutation d'un acide aminé hydrophobe en acide aminé polaire déstabilisant potentiellement le domaine. Il a été montré au laboratoire que la température de fusion du double domaine humain est 4 degrés plus basse que celle du double domaine aviaire. D'autre part, la seule mutation V613T sur le 627-NLS aviaire permet d'obtenir une température de fusion proche de celle du 627-NLS humain, suggérant que l'effet de cette mutation est bien la déstabilisation du domaine 627-NLS (résultats non publiés). Cette mutation pourrait potentiellement apporter un avantage adaptatif chez l'homme en augmentant la flexibilité et la plasticité du domaine.

4.6 Résidu 676

La mutation T676I fait partie des premières à apparaître lors de la transmission du virus d'un hôte aviaire à un hôte humain [Miotto et al., 2008]. Une autre mutation de cet acide aminé, la mutation A676T fait partie d'un ensemble de sept mutations permettant de compenser l'absence de la mutation E627K lors de l'adaptation chez la souris (virus H5N1) [Li et al., 2009].

4.7 Résidus 661 et 684

L'étude de la capacité d'un virus aviaire contenant une protéine originaire d'un virus aviaire adaptée à l'homme (ici, une souche H1N1 de 2009) à se répliquer à température humaine a permis de mettre en évidence quatre mutations augmentant fortement l'activité réplication du virus : V661A, V683T/A684S et T271A [Hayashi et al., 2015]. Deux de ces résidus, 661 et 684, ont également été identifiés comme impliqués dans l'adaptation à un hôte humain via des études bio-informatiques [Miotto et al., 2008].

4.8 Résidu 702

La mutation R702K a tout d'abord été identifiée par une étude bio-informatique [Miotto et al., 2008]. Bien que l'acide aminé reste basique, cette mutation augmente l'affinité du domaine NLS à l'importine α en augmentant la constante d'association de l'interaction [Boivin and Hart, 2011].

4.9 Résidu 714

La mutation S714R, en combinaison avec D701N, permet l'augmentation de l'activité de la polymérase grâce à l'augmentation de l'interaction avec les ARN coiffés *in vitro* [Zhang et al., 2012]. Un autre mécanisme d'action de la mutation S714R a récemment été proposé, suggérant que son rôle soit de faciliter la séparation du peptide NLS du domaine NLS, événement nécessaire à l'interaction avec l'importine α . En effet, l'acide aminé 714 est situé à proximité de R737. La mutation S714R pourrait donc, via une interaction électrostatique, repousser la chaîne latérale de l'acide aminé R737 et ainsi permettre une meilleure exposition du peptide NLS [Czudai Matwich et al., 2014]. Cette hypothèse est soutenue par l'étude montrant que la mutation S714R augmente l'association du domaine NLS à l'importine α *in vitro* [Boivin and Hart, 2011].

5 Objectifs

La structure du double domaine 627-NLS a été extensivement caractérisée par cristallographie aux rayons X. Cependant, les descriptions existantes ne sont pas suffisantes pour comprendre comment cette protéine effectue sa fonction *in vivo*. En effet, l'existence d'une conformation différente de celle cristallisée est nécessaire pour expliquer son association à l'importine α . De nombreux essais ont été réalisés au laboratoire pour cristalliser le complexe 627-NLS/importine α , sans jamais donner de résultats, suggérant une flexibilité et une dynamique trop importantes. C'est pourquoi nous avons entrepris une étude du domaine 627-NLS en solution, par des techniques basse et haute résolution, telles que la RMN, le SAS et le FRET. Un premier objectif de ma thèse est de déterminer la dynamique du 627-NLS en solution et d'obtenir une description structurale des différents états présents. Un second objectif est de déterminer la façon dont cette dynamique permet l'interaction avec l'importine α et de décrire la structure du complexe.

Par ailleurs, de nombreuses mutations dans le domaine 627-NLS ont lieu lors de l'adaptation inter-espèces du virus. Les souches humaines et aviaires choisies pour cette études comportent 10 de ces mutations dans le double domaine. Nous avons cherché à observer leur effet d'une part sur la dynamique du 627-NLS en solution,

et d'autre part sur son interaction avec l'importine α . En effet, des interactions préférentielles du 627-NLS provenant de virus aviaires ou humains avec certains isoformes de l'importine α ont été mises en évidence dans la littérature, mais les bases moléculaires de celles-ci restent floues. C'est pourquoi nous avons choisi d'étudier non seulement l'affinité de l'interaction de l'importine α avec le 627-NLS humain et aviaire, mais également avec des simples mutants et des combinaisons de mutants définis selon l'évolution naturelle du virus. De plus, l'interaction 627-NLS/importine α s'effectue à plus haute température chez les oiseaux (40°C) que chez l'homme (33°C). Pour cette partie, le troisième objectif de ma thèse est de déterminer par SPR si l'affinité de l'importine α avec le 627-NLS est influencée par l'isoforme d'importine α utilisé combiné à la température à laquelle est effectuée l'interaction. Enfin, une technique permettant la détermination rapide d'affinités et peu gourmande en matériel biologique, le MST (*Microscale thermophoresis*), permettra de déterminer si un effet de certaines mutations évolutives du 627-NLS sur son interaction avec l'importine α peut être identifié.

Les résultats seront présentés en deux chapitres.

Le premier correspond à l'étude de la dynamique ainsi qu'à la description structurale du 627-NLS aviaire et de son complexe avec l'importine α en solution. Il commencera par la publication rédigée lors de ma thèse, puis je présenterai certains résultats complémentaires à propos du 627-NLS aviaire, mais aussi humain. Je présenterai dans une seconde partie les résultats concernant l'interaction du 627-NLS à différentes températures avec différents isoformes de l'importine α , ainsi que l'interaction de l'importine α avec les mutants évolutifs du 627-NLS.

Le second chapitre présente quant à lui la résolution par cristallographie aux rayons X du domaine 627 issu d'une souche de la grippe B et sa comparaison avec les structures du même domaine issu d'autres souches. Ce dernier travail a été effectué lors d'une collaboration dans le cadre de la résolution de la structure cristallographique de la polymérase entière.

Résonance Magnétique Nucléaire

1 Généralités

La résonance magnétique nucléaire (RMN) est une puissante technique d'analyse à résolution atomique des molécules en solution, basée sur la mesure des propriétés magnétiques de certains noyaux soumis à un champ magnétique intense.

Une des caractéristiques des noyaux est le spin, représenté par le vecteur \vec{I} , dont la norme est exprimée en fonction du nombre quantique de spin I et de la constante de Planck h .

$$\|\vec{I}\| = \frac{h}{2\pi} \times \sqrt{I(I+1)} \quad (3.1)$$

Les noyaux possédant un spin nucléaire non nul sont visibles en RMN. Cependant, les noyaux ayant un spin supérieur à $1/2$ possèdent un moment quadripolaire qui induit des couplages complexes et un élargissement des raies. Il est donc beaucoup plus difficile d'avoir une bonne résolution du signal. C'est pourquoi, en RMN des protéines, on utilise préférentiellement les noyaux possédant un spin $1/2$, tels que le ^{15}N ou ^1H . Cependant, à l'exception des protons, la plupart des isotopes de spin $1/2$ intéressants pour l'observation des molécules biologiques ont une abondance naturelle très faible. Des techniques ont donc été développées pour enrichir les protéines d'intérêt en ^{15}N ou en ^{13}C , permettant une acquisition de données dans un temps raisonnable.

Une autre caractéristique intrinsèque des noyaux importante pour la RMN est le moment gyromagnétique, γ . Il peut être positif ou négatif, est fixe pour un type de noyaux donné (^1H , ^{15}N , ...), et influe sur le moment magnétique $\vec{\mu}$ associé au spin selon l'équation suivante :

$$\vec{\mu} = \gamma \times \vec{I} \quad (3.2)$$

2 Résonance des noyaux à spin demi-entier

Lorsqu'un champ magnétique \vec{B}_0 est appliqué (par convention selon l'axe z) à un échantillon, il interagit avec le moment magnétique des spins à une énergie E telle que :

$$E = -\gamma I_z B_0 \quad (3.3)$$

Avec I_z , la projection de \vec{I} sur l'axe z , qui est exprimée en fonction de m_I , le nombre quantique magnétique :

$$I_z = \frac{m_I \hbar}{2\pi}, m_I \in [-I, -I + 1, \dots, I - 1, I] \quad (3.4)$$

Dans le cas d'un noyau de spin $1/2$, m_I ne peut prendre que deux valeurs : $+1/2$ (état β) ou $-1/2$ (état α) :

$$E_\alpha = -\frac{1}{2}\gamma B_0 \hbar \quad \text{et} \quad E_\beta = \frac{1}{2}\gamma B_0 \hbar \quad \text{avec} \quad \hbar = \frac{h}{2\pi} \quad (3.5)$$

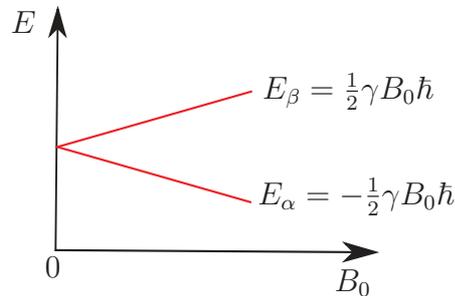


FIGURE 3.1 – Levée de dégénérescence des états de spin.

Le champ magnétique \vec{B}_0 provoque une séparation énergétique des états de spin α et β .

Comme la différence d'énergie entre les états de spin est directement proportionnelle au champ magnétique appliqué, plus le champ \vec{B}_0 est intense, plus grande est la différence d'énergie entre l'état α et l'état β (Figure 3.1). De plus, le champ magnétique « ressenti » par les atomes de l'échantillon n'est pas tout à fait B_0 mais $B_{eff} = B_0(1 - \sigma)$. En effet, il est modulé par la constante d'écran σ due aux électrons du nuage électronique, aux atomes vicinaux, aux liaisons, etc. C'est grâce à cet effet que l'étude des protéines en RMN est possible, chaque atome étant ainsi distinguable dans le spectre enregistré.

Les niveaux α et β étant dégénérés, il est possible d'induire une transition grâce à l'énergie :

$$\Delta E = \gamma B_0 \hbar = h\nu_0 \quad (3.6)$$

On peut exprimer la fréquence de la radiation nécessaire, ν_0 , en fréquence angulaire, ω_0 :

$$\nu_0 = \frac{\gamma B_0}{2\pi} \quad \text{donc} \quad \omega_0 = \gamma B_0 \quad (3.7)$$

La fréquence ω_0 est appelée fréquence de Larmor et satisfait la "condition de résonance", c'est-à-dire qu'elle permet d'induire la transition entre état α et β . Par exemple, pour un champ magnétique de 14,4 T, la fréquence de Larmor des protons est de 600 MHz.

3 La spectroscopie par RMN

Une fois l'échantillon soumis à un champ magnétique \vec{B}_0 , un équilibre est atteint. On peut calculer la différence de population entre les deux états obtenus par l'équation de Boltzmann :

$$\frac{N_\alpha}{N_\beta} = \exp\left(\frac{\Delta E}{k_B T}\right) \quad (3.8)$$

Il n'y a qu'un très faible excès de la population α par rapport à la population β , et c'est cette différence qui est utilisée par la spectroscopie par RMN : c'est pourquoi cette technique manque de sensibilité. C'est pour pallier ce problème que des champs \vec{B}_0 toujours plus élevés sont recherchés, augmentant la différence de population obtenue à l'équilibre et donc la sensibilité de la technique. Le spectromètre RMN de plus haut champ fabriqué actuellement a un champ de 23,5 T (1 GHz).

Au niveau macroscopique, l'excès de population α crée une magnétisation nette selon l'axe z , \vec{M}_0 (Figure 3.2A,B). Il est possible de perturber cette dernière grâce à l'application d'un champ électromagnétique à la fréquence de Larmor dans le plan xOy , nommée le champ \vec{B}_1 . Le retour des spins à l'équilibre (relaxation) génère un courant dans une bobine placée dans le plan xOy . C'est sur ce phénomène que repose le principe de la spectroscopie par RMN (Figure 3.2C).

Expérimentalement, une impulsion d'une courte durée à une fréquence donnée (porteuse) va permettre d'exciter une certaine gamme de fréquences autour de cette porteuse, soit tous les noyaux contenus dans l'échantillon. Le signal collecté contient donc plusieurs fréquences et est appelé le *free induction decay* (FID). Une transformée de Fourier est ensuite utilisée pour convertir ce FID en fréquences (Figure 3.3).

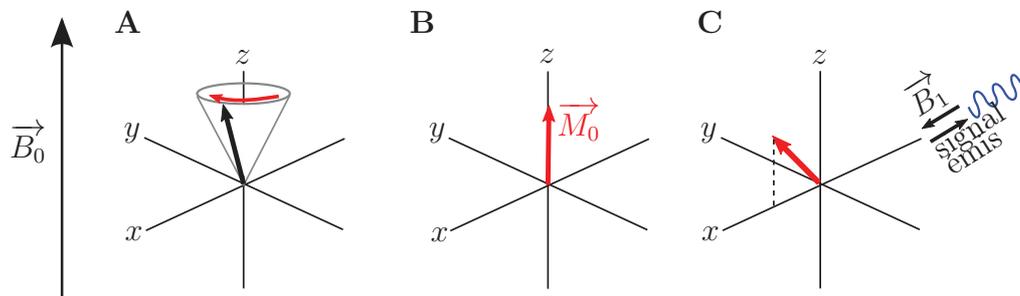


FIGURE 3.2 – Représentation géométrique des spins dans le champ B_0 dans le référentiel du laboratoire (**A** et **B**) ou rotatif, à la fréquence de Larmor (**C**). **A.** A l'équilibre, les spins précessent à la fréquence de Larmor (rouge, noyau possédant un moment gyromagnétique γ positif). **B.** Sous l'effet de \vec{B}_0 , un excès de spins est aligné de la même façon, résultant en une magnétisation macroscopique nette M_0 , en rouge. **C.** Lors de la perturbation du système par une radiofréquence \vec{B}_1 à la fréquence de Larmor dans le plan xOy , les spins sont perturbés et la magnétisation moyenne (rouge) induit un courant détectable, ici en bleu.

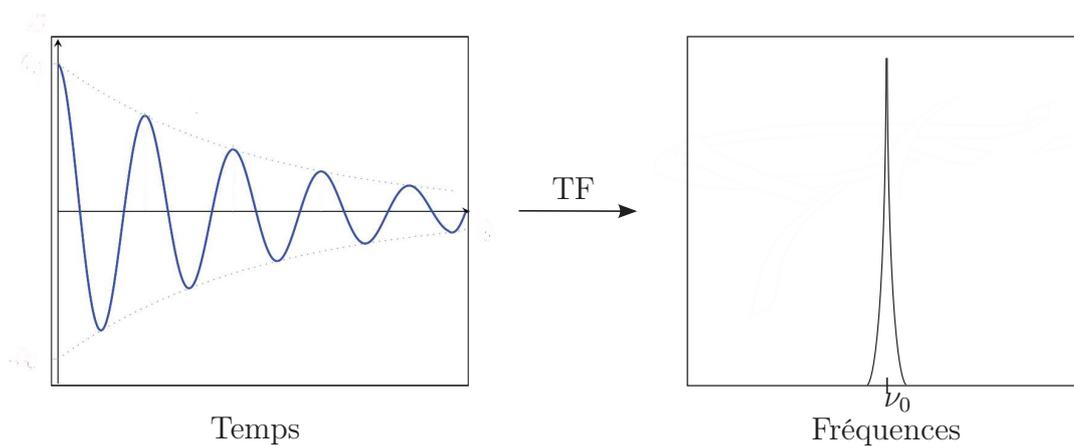


FIGURE 3.3 – Déconvolution de spectre utilisant la transformée de Fourier. Passage du domaine du temps (à gauche) au domaine des fréquences (à droite) grâce à la transformée de Fourier.

4 Les spectres multidimensionnels

Lors de l'étude des macromolécules biologiques par RMN, les spectres obtenus sont très complexes car ce sont des échantillons contenant un très grand nombre de noyaux. Il est donc très difficile de résoudre les signaux individuels observés sur un spectre à une seule dimension. C'est pourquoi en RMN des protéines, en sus du proton, l'azote 15 et le carbone 13 sont utilisés pour permettre l'acquisition de spectre contenant 2, voire 3 dimensions ou plus (Figure 3.4).

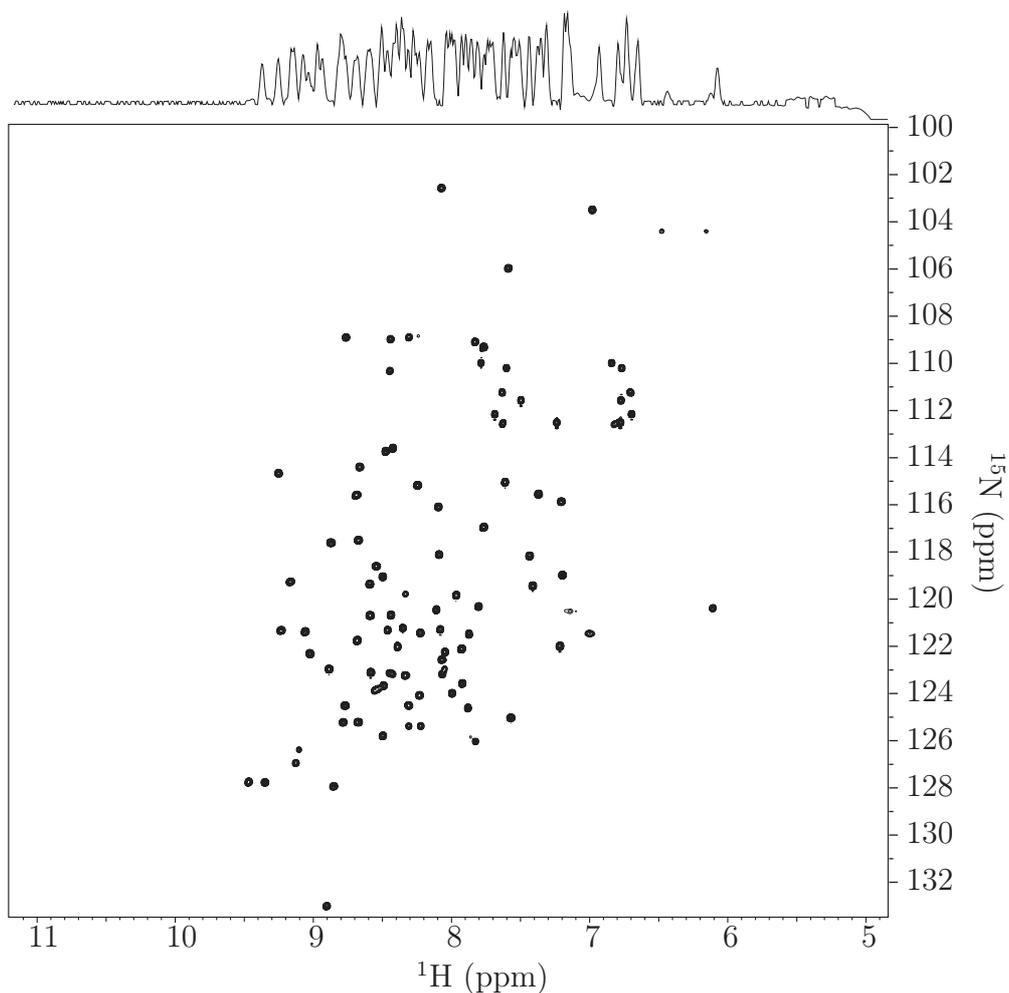


FIGURE 3.4 – Spectres à une et deux dimensions.

En haut, un spectre à une dimension (^1H) de l'ubiquitine. En bas, un spectre à deux dimensions (^1H - ^{15}N HSQC) du même échantillon. Le spectre 2D permet de résoudre tous les signaux observés là où le spectre 1D ne permet que des observations d'ordre général.

Par exemple, l'HSQC pour *heteronuclear single quantum coherence* est l'une des expériences les plus couramment utilisées en RMN des protéines en solution. Elle permet d'obtenir une corrélation entre le proton et l'azote, lorsque leur couplage scalaire est suffisant. Dans une protéine, on trouve ce couple d'atomes une fois par liaison amide (sauf pour les prolines), et dans la chaîne latérale de certains acides aminés (arginines, lysines, glutamines, asparagines, tryptophanes et histidines). Dans le spectre HSQC d'une protéine, on a donc des signaux correspondant aux chaînes latérales (à des positions caractéristiques, et donc facilement identifiables), et un signal par acide aminé. Cette expérience simple permet d'obtenir des informations directes sur le repliement de la protéine (par exemple, les protéines désordonnées ont une dispersion spectrale faible dans la dimension proton) mais elle constitue surtout la base de la plupart des expériences de RMN des protéines en solution.

5 Les paramètres observés

5.1 Les échelles de temps en RMN

La RMN permet d'obtenir des informations à résolution atomique de phénomènes ayant lieu à différentes échelles de temps, de la picoseconde à la seconde (Figure 3.5). Pour cela, différentes expériences ont été conçues et les paragraphes qui vont suivre décriront les paramètres mesurés lors de cette thèse ainsi que l'information qu'ils apportent.

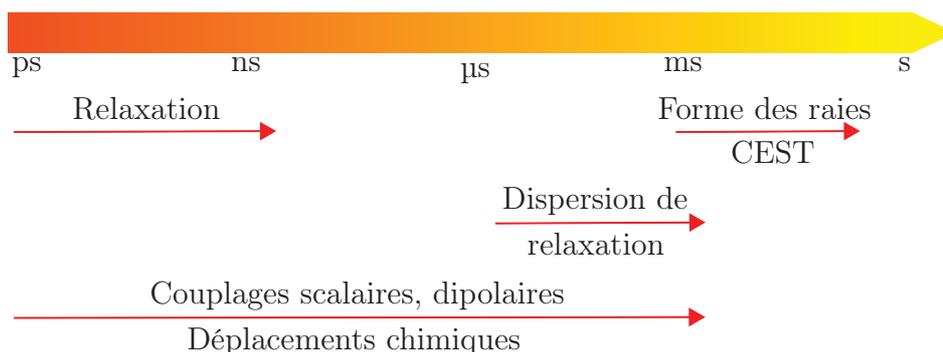


FIGURE 3.5 – Echelles de temps observables par RMN. Les expériences sensibles à une échelle de temps donnée sont représentées par une flèche.

5.2 Les couplages

5.2.1 Les couplages scalaires

Les couplages scalaires correspondent à une interaction entre deux spins liés de façon covalente. Dans un système à deux spins, 4 transitions sont possibles. Lorsque les spins ne sont pas couplés, ces transitions sont équivalentes deux à deux, et ne provoquent donc l'apparition que de deux signaux au total. Mais lorsque les spins sont couplés, les niveaux d'énergie des deux spins changent légèrement, les transitions ne sont plus équivalentes, et on a un dédoublement des raies spectrales (Figure 3.6A). Les couplages se notent ${}^N J$ avec N le nombre de liaison entre deux noyaux. Il est possible de découpler les noyaux pendant la séquence d'impulsion, ce qui

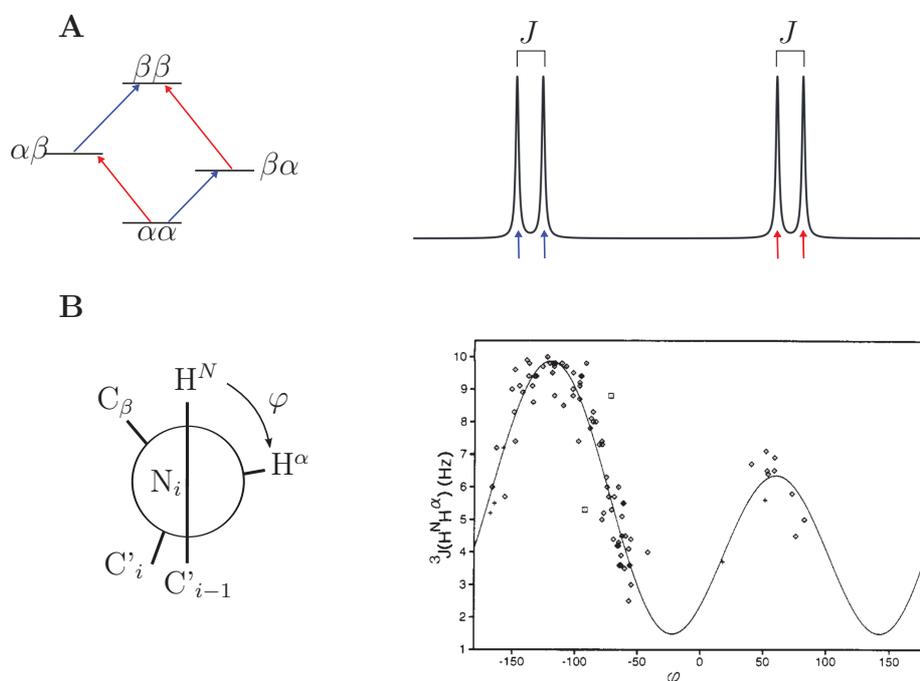


FIGURE 3.6 – **A.** Couplage scalaire entre deux spins. A gauche, représentation schématique des transitions possibles. Les transitions équivalentes dans un système non couplé sont de même couleur. A droite, le spectre obtenu. **B.** A gauche, définition de l'angle dièdre φ de la liaison peptidique (représentation de Newman). A droite, le couplage scalaire ${}^3J(\text{H}^N\text{H}^\alpha)$ en fonction de l'angle dièdre de la liaison peptidique φ en degrés. Les points représentent les couplages ${}^3J(\text{H}^N\text{H}^\alpha)$ mesurés à partir du spectre HNCA de la nucléase du staphylocoque. La ligne montre l'ajustement de la courbe de Karplus obtenue pour $A = 6,5$, $B = -1,76$ et $C = 1,6$ ($RMSD = 0,73$). Adapté de [Vuister and Bax, 1993].

permet par exemple de n'avoir qu'une raie au lieu de 4 dans une HSQC. Par ailleurs, les couplages 1J sont utilisés pour transférer la magnétisation lors d'expérience à trois ou quatre dimensions, et sont donc extrêmement utiles à la technique. Les

couplages ${}^3J(\text{H}^N\text{H}^\alpha)$ se révèlent extrêmement utiles dans le cadre de la détermination de structure, puisqu'ils dépendent de l'angle dièdre φ de la liaison peptidique (Figure 3.6B), selon l'équation de Karplus [Karplus, 1963], avec A, B et C des constantes :

$${}^3J(\varphi) = A \cos^2 \varphi + B \cos \varphi + C \quad (3.9)$$

Pour paramétrer cette fonction, on enregistre des couplages sur une protéine dont la structure à résolution atomique est connue, ce qui permet d'obtenir une courbe des couplages ${}^3J(\text{H}^N\text{H}^\alpha)$ en fonction de l'angle de la liaison peptidique φ . L'ajustement de la fonction de Karplus à la courbe obtenue permet de déterminer les paramètres A, B et C qui peuvent ensuite être utilisés pour la détermination d'une structure inconnue.

5.2.2 Les couplages dipolaires

Les couplages dipolaires correspondent à une interaction à travers l'espace de deux spins : c'est une interaction dipôle-dipôle qui s'exprime ainsi :

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3} \left\langle \frac{3 \cos^2 \theta(t) - 1}{2} \right\rangle \quad (3.10)$$

Avec θ l'angle du vecteur internucléaire avec \vec{B}_0 , r la distance entre les deux spins considérés et μ_0 la permittivité du vide. Les parenthèses angulaire représentent la moyenne sur tous les couples ij dans le temps de la mesure. Dans les solutions isotropes, le couplage dipolaire se moyenne à 0 du fait de la rotation des molécules.

5.3 Le déplacement chimique

Le déplacement chimique correspond à la fréquence de résonance du noyau observé. Comme cette fréquence dépend du champ magnétique auquel est effectuée la mesure, on utilise le déplacement chimique, δ , exprimé en partie par million (ppm), référencé comme suit :

$$\delta(\text{ppm}) = 10^6 \times \frac{\nu_0 - \nu_{ref}}{\nu_{ref}} \quad (3.11)$$

Avec ν_{ref} la fréquence de Larmor de la référence et ν_0 la fréquence de Larmor du noyau observé. La référence classiquement utilisée en RMN des protéines est l'acide 4,4-diméthyl-4-silapentane-1-sulfonique (DSS), dont le déplacement chimique est considéré comme 0. Cependant, comme le solvant utilisé en RMN des protéines est aqueux et que le déplacement chimique de l'eau en fonction de la température est connu, il permet de référencer le spectre de façon simple et a été utilisé comme référence lors de cette thèse.

Comme le déplacement chimique est extrêmement sensible à l'environnement chimique du noyau observé, c'est un paramètre très utile pour déterminer par exemple les zones d'interaction entre une molécule observée et un partenaire lors d'une expérience de titration, ou repérer un changement conformationnel. Une formule classiquement utilisée pour quantifier l'amplitude des changements observés lors d'une interaction utilise les déplacements chimiques pondérés de l'azote et du proton :

$$\text{CSP} = \sqrt{(c \cdot \Delta\delta_N)^2 + (\Delta\delta_H)^2} \quad (3.12)$$

avec CSP la perturbation des déplacements chimiques (*Chemical shift perturbation*, CSP), $\Delta\delta_N$ et $\Delta\delta_H$ les différences de déplacements chimiques entre les deux états observés pour l'azote et le proton respectivement, et c une constante permettant d'équilibrer les différences observées dans les deux dimensions compte tenu de leur différence de largeur spectrale, généralement comprise entre 0,1 et 0,2.

5.3.1 L'attribution des résonances

Lors de l'étude d'une protéine par RMN, une étape essentielle pour l'interprétation de futures mesures est d'identifier à quels atomes correspond chaque signal. Pour ce faire, on utilise un ensemble d'expériences tridimensionnelles lors desquelles la magnétisation est transférée grâce aux couplages scalaires sur les différents noyaux d'intérêt et leurs déplacements chimiques enregistrés (Table 3.7).

Couplage	Valeur (Hz)
$^1J(^1H^N, ^{15}N)$	93
$^1J(^{15}N, ^{13}C')$	14
$^2J(^{15}N, ^{13}C')$	2
$^1J(^{15}N, ^{13}C^\alpha)$	9
$^2J(^{15}N, ^{13}C^\alpha)$	7
$^1J(^{13}C^\alpha, ^{13}C')$	53
$^1J(^{13}C^\alpha, ^{13}C^\beta)$	37

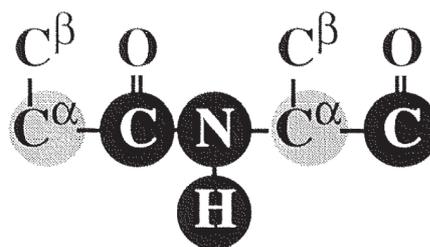


FIGURE 3.7 – A gauche, tableau des couplages scalaires de la liaison peptidique. A droite, représentation de la liaison peptidique permettant de visualiser le “chemin” de la magnétisation lors d’une séquence d’impulsions. Adapté de [Salzmann, 1999].

Par exemple, dans le cas d'une HNCO (Figure 3.8), après une première impulsion sur le proton, la magnétisation est transférée à l'azote via le couplage scalaire. Une impulsion ramène la magnétisation du proton en z tandis que la magnétisation de l'azote est dans le plan xOy . Un premier délai t_1 permet de laisser évoluer le déplacement chimique de l'azote, avec au milieu une impulsion sur les trois canaux permettant d'empêcher l'évolution des couplages scalaires N-H et N-C et de transférer la magnétisation sur le carbonyle grâce au couplage scalaire N-C. A l'issue de cette période t_1 , la magnétisation de l'azote est ramenée en z et celle du carbonyle est basculée en xOy . Une période t_2 permet de laisser évoluer son déplacement chimique tandis qu'une impulsion au milieu de ce délai sur l'azote et le proton permet d'éviter l'évolution des couplages scalaires C-H et C-N. Une fois cette période t_2 terminée, la magnétisation est ramenée sur l'azote, puis sur le proton grâce aux couplages scalaires C-N et N-H respectivement. Enfin, la magnétisation du proton est amenée dans le plan xOy , et les données (FID) sont enregistrées pendant la période t_3 , période d'évolution du déplacement chimique du proton, durant laquelle un train d'impulsions de découplage est appliqué sur le canal de l'azote, permettant d'éviter l'évolution du couplage scalaire N-H pendant la mesure.

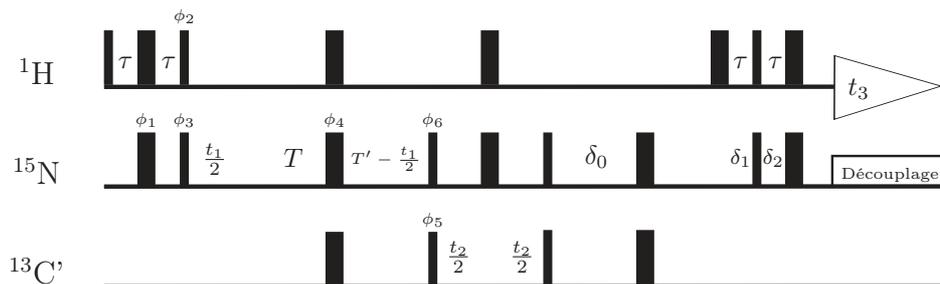


FIGURE 3.8 – Séquence d'impulsion HNCO à temps constant simplifiée. Les rectangles noirs fins et larges représentent des impulsions 90° et 180° respectivement. Les impulsions dont les phases ne sont pas indiquées sont appliquées selon x , et $\phi_1 = x, -x$, $\phi_2 = y, -y$, $\phi_3 = x$, $\phi_4 = 4(x), 4(y), 4(-x), 4(-y)$, $\phi_5 = x, x, -x, -x$, $\phi_6 = x, -x$. $\tau = 2,25$ ms, $T = 13,5$ ms, $T' = T + \tau_{180^\circ(^{13}C')}$, $\delta_1 = 7,25$ ms, $\delta_2 = 2,75$ ms, $\delta_0 = \delta_1 + \delta_2 + \tau_{180^\circ(^1H)}$. Adapté de [Grzesiek and Bax, 1992].

Ces différentes expériences vont permettre de corrélérer les déplacements chimiques d'un résidu donné avec ceux du résidu précédent :

- HNCO et HNcaCO : Donnent les déplacements chimiques de ^NH , N et C' du résidu $i - 1$ (HNCO) ou i et $i - 1$ (HNcaCO)
- HNCA, HNCACB, HNcoCA et HNcoCACB : Donnent les déplacements chimiques de ^NH , N et C_α du résidu i et $i - 1$ (HNCA) ou $i - 1$ (HNcoCA) et C_β

du résidu i et $i - 1$ (HNCACB) ou $i - 1$ (HNcoCACB).

Comme certains spectres permettent d'obtenir les déplacements chimiques de l'acide aminé i et $i - 1$ ou uniquement $i - 1$ pour une résonance ^NH , N donnée, il est possible d'attribuer de façon séquentielle les signaux obtenus (Figure 3.9). L'attribution peut être effectuée de façon semi-automatique, grâce à des logiciels tels que MARS (algorithme d'optimisation stochastique) [Jung and Zweckstetter, 2004]. Ce dernier utilise les systèmes de spins définis par l'utilisateur, contenant les différents déplacements chimiques i et $i - 1$ recensés pour une résonance donnée, et, compte tenu de la séquence de la protéine et des prédictions de structures secondaires, établit une attribution probable par comparaison des valeurs pour les différents noyaux. D'autre part, des tables statistiques établies à partir des déplacements chimiques des protéines attribuées existent [BMRB, 2015, Wang and Jardetzky, 2002]. Ces dernières permettent d'identifier facilement certains résidus du fait de leurs déplacements chimiques caractéristiques, comme la glycine ($\delta C_\alpha \approx 45$ ppm), les sérines et thréonine ($\delta C_\beta \approx 64$ et 70 ppm respectivement) ou l'alanine ($\delta C_\beta \approx 18$ ppm), facilitant ainsi l'attribution.

Dans le cas de l'étude de protéines désordonnées et/ou contenant de nombreux acides aminés, les résonances se superposent souvent. Ce phénomène est problématique car s'il est souvent évident que plusieurs pics sont superposés grâce au nombre de résonances obtenues dans la dimension carbone, il est en général difficile de déterminer quelles résonances appartiennent au même système de spin. D'autre part, MARS ne possède pas de paramètres permettant de prendre en compte la présence d'un échange entre plusieurs états. Dans ce cas, si deux listes correspondant à chaque état peuvent être établies, MARS sera en général capable de réaliser l'attribution. Si ce n'est pas le cas, il faudra réaliser l'attribution manuellement. Quoiqu'il en soit, il reste important de vérifier (et généralement de terminer) l'attribution manuellement, certaines erreurs étant courantes, et certaines situations n'étant discernables que par l'oeil humain.

5.3.2 Les déplacements chimiques secondaires

Du fait de leur sensibilité à l'environnement du spin observé, les déplacements chimiques peuvent nous renseigner sur la présence de structures secondaires. En effet, le déplacement chimique du ^NH , N , C' , C_α , ou C_β d'un spin situé dans un feuillet β ou dans une hélice α ne sera pas le même [Wang, 2004]. Pour identifier ces structures secondaires, on utilise un paramètre nommé déplacement chimique secondaire, qui correspond à la différence entre le déplacement chimique mesuré et le déplacement chimique du même acide aminé dans un état complètement déplié [Dalgarno et al., 1983, Wishart et al., 1995]. Cette dernière valeur était tout d'abord issue de l'index des déplacements chimiques, développé dans les années 1990 [Wishart et al., 1992, Wishart and Sykes, 1994]. Plus récemment, des bases de données ont

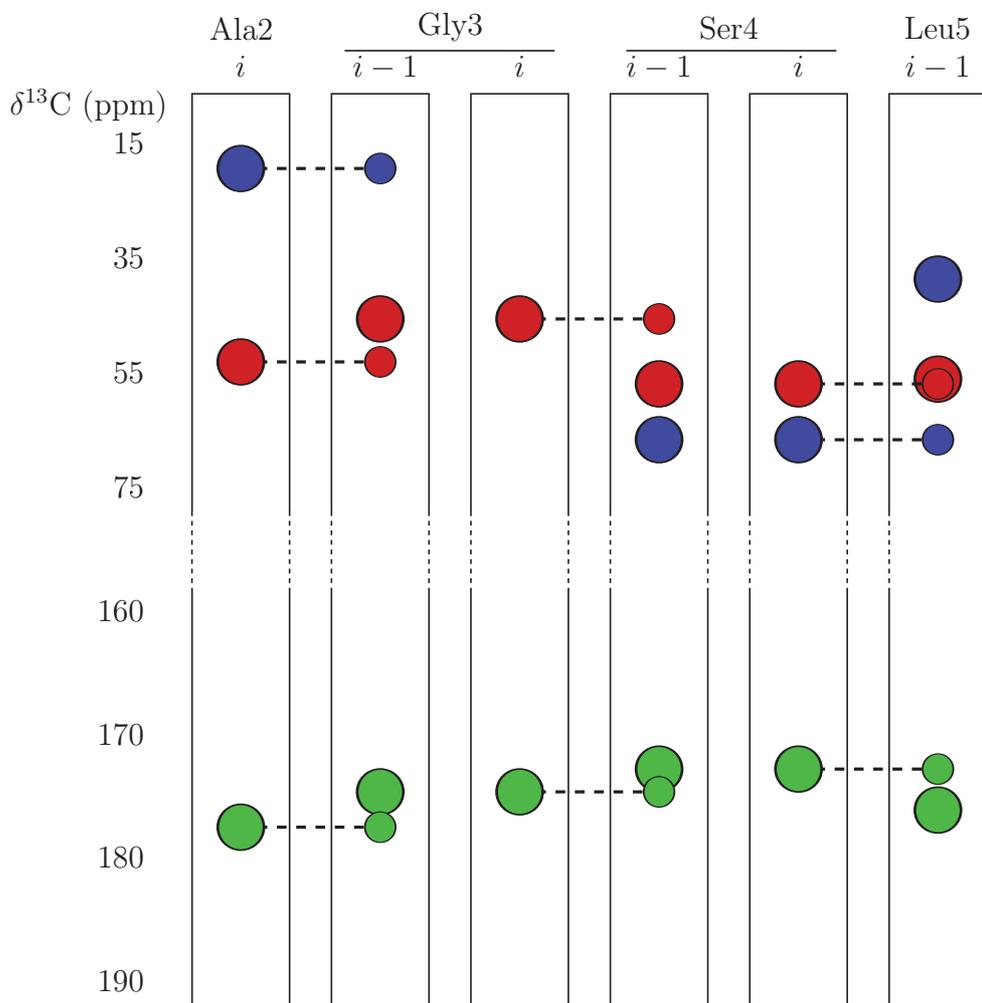


FIGURE 3.9 – Principe de l’attribution séquentielle des résonances.

Exemple pour un peptide de séquence XAGSL. Chaque bande correspond aux déplacements chimiques des carbones α (rouge), β (bleu) et carbonyles (vert) observés dans les expériences tridimensionnelles, aux coordonnées N et H de chaque acide aminé. Les résonances de l’acide aminé i et $i - 1$ sont représentées par un gros et petit rond respectivement pour les résonances C_α et C_β , petit rond et gros rond respectivement pour les résonances C’. La différence d’intensité observée est due aux couplages scalaires plus ou moins grand qui doivent être passés pour le transfert de la magnétisation vers l’acide aminé i ou vers l’acide aminé $i - 1$ (Figure 3.7). L’attribution séquentielle est mise en évidence par des lignes pointillées.

été construites à partir des déplacements chimiques de peptides désordonnés ainsi que des parties désordonnées de protéines attribuées, et sont utilisées pour le calcul des déplacements chimiques secondaires [Tamiola et al., 2010, De Simone et al., 2009]. La déviation par rapport à l'état déplié, ou déplacement chimique secondaire, permet de mettre en évidence les structures secondaires présentes, même partiellement. Le logiciel SSP permet de produire un score spécifique à chaque résidu à partir des déplacements chimiques des différents noyaux, pondérés par leur sensibilité aux différentes structures secondaires possibles [Marsh et al., 2006]. Le score obtenu, ou propension de structure secondaire, représente la fraction d'hélice α ou de brin β attendu.

5.4 La relaxation

La relaxation correspond au retour à l'équilibre des spins après perturbation du système. Les mécanismes de relaxation les plus souvent rencontrés peuvent être décrits comme deux phénomènes séparés :

- Le retour à l'équilibre de la magnétisation le long de l'axe z (direction de \vec{B}_0). T_1 est le temps de relaxation longitudinal et R_1 la vitesse de relaxation longitudinale avec $R_1 = 1/T_1$.
- La disparition de la composante dans le plan xOy . T_2 est le temps de relaxation transversal et R_2 est la vitesse de relaxation transversale avec $R_2 = 1/T_2$.

S'il est possible d'enregistrer des expériences qui nous informent sur la relaxation d'un certain nombre de noyaux, il ne sera présenté dans ce manuscrit que la relaxation ^{15}N . Cette dernière nous apporte des informations sur la dynamique de la protéine, de la picoseconde à la nanoseconde.

En plus du temps de corrélation, paramètre dépendant de la protéine d'intérêt, deux principaux phénomènes contribuent à la relaxation R_1 et R_2 du noyau N dans le couple N-H : l'anisotropie de déplacement chimique (CSA) R^{CSA} et l'interaction dipolaire avec les protons directement liés au noyau observé R^D . La relaxation transversale est quant à elle également sensible à l'échange chimique. On a donc :

$$\begin{aligned} R_1 &= R_1^D + R_1^{CSA} \\ R_2 &= R_2^D + R_2^{CSA} + R_2^{ech} = R_2^0 + R_2^{ech} \end{aligned} \quad (3.13)$$

L'anisotropie de déplacement chimique affecte de façon importante le noyau ^{15}N . Elle est causée par l'anisotropie de l'environnement électronique du noyau.

L'interaction dipolaire correspond à l'interaction d'un spin avec un autre à travers l'espace (voir paragraphe 5.2.2). Dans le cas de l'azote de la liaison amide,

l'interaction dipolaire dominante est celle avec l'hydrogène ^1H , qui a un effet important du fait de son rapport gyromagnétique élevé et de la courte distance H-N. Les effets des carbones de la chaîne principale, C' et C_α , sont dans ce cas négligeables.

Le temps de corrélation rotationnel d'une protéine domine ses mouvements. Noté τ_c , il correspond au temps moyen nécessaire à une molécule se comportant comme une sphère pour effectuer une rotation d'un radian. Il est essentiellement fonction de la taille de la molécule et de la viscosité du solvant. Il est de l'ordre de quelques nanosecondes pour une petite protéine en solution aqueuse (Figure 3.10). Cependant, plus la protéine est de taille importante, plus τ_c est grand et donc plus le temps de relaxation T_2 est rapide, ce qui complique considérablement l'acquisition de données sur de grosses protéines. Différentes techniques ont été mises au point pour pallier ce problème et seront développées au paragraphe 6.

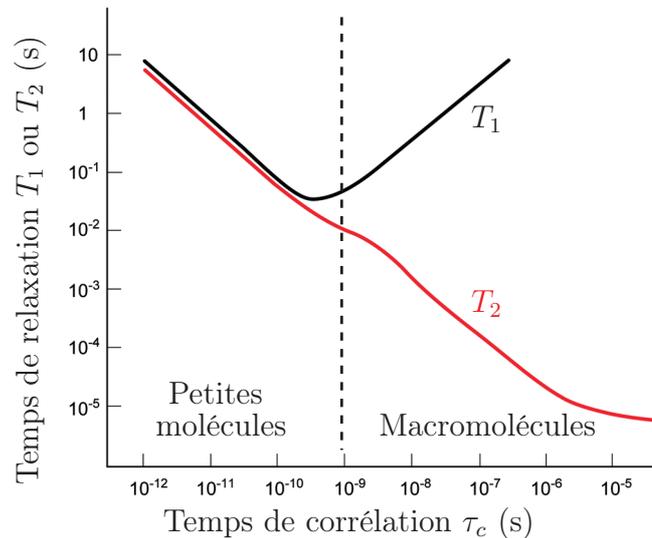


FIGURE 3.10 – Effet du temps de corrélation τ_c sur la relaxation du proton amide.

5.4.1 Mesure de la relaxation

La relaxation R_1 est mesurée grâce à une expérience de type *inversion recovery* [Kay et al., 1989], dans laquelle un cyclage de phase permet d'obtenir une décroissance de magnétisation qui peut être exprimée selon l'équation suivante :

$$M_z(t) = M_{z,eq}(1 - 2 \exp(-t/T_1)) \quad \text{avec} \quad M_z(0) = -M_{z,eq} \quad (3.14)$$

La relaxation R_2 peut quant à elle être déduite directement de la largeur de raie (largeur à mi-hauteur), cependant cette méthode est peu précise à cause de l'inhomogénéité du champ B_0 et de la rotation des molécules dans ce champ. Une séquence

mise au point par Carr, Purcell, Meiboom et Gill [Carr and Purcell, 1954, Meiboom and Gill, 1958] est utilisée intégrée dans une HSQC pour mesurer la relaxation R_2 . Ensuite, comme pour R_1 , la décroissance d'intensité du signal observée lorsque le délai augmente suit une décroissance exponentielle :

$$M_{xy}(t) = M_{xy}(0) \exp(-t/T_2) \quad (3.15)$$

5.4.2 La relaxation par effet Overhauser nucléaire

Lorsque deux noyaux sont couplés de façon dipolaire, une relaxation croisée entre eux influence leur relaxation. C'est l'effet Overhauser nucléaire (*nuclear Overhauser effect*, nOe). Expérimentalement, on mesure cet effet, à l'état stationnaire, sur le couple N-H en saturant l'hydrogène par de faibles impulsions afin d'augmenter l'effet sur le signal observé du couplage dipolaire entre N et ^1H [Kay et al., 1989]. Deux expériences sont mesurées, une saturée, et l'autre non. Le rapport des intensités entre les signaux obtenus lors des deux expériences représente l'augmentation de cet effet et nous informe sur la flexibilité à l'échelle de temps de la picoseconde à la nanoseconde (Figure 3.5).

5.4.3 La relaxation cross-corrélée

Comme on l'a vu au paragraphe 5.2.1, les états α et β d'un spin ne sont en fait pas équivalents. Dans le cas d'un couple N-H, on a deux raies dans la dimension azote et deux dans la dimension proton, soit 4 au total. Or, ces 4 raies ne relaxent pas de la même façon. La relaxation transversale de l'état α ou β peut être exprimée de la façon suivante :

$$\begin{aligned} R_2^\alpha &= R_2^D + R_2^{CSA} + \eta_{xy} \\ R_2^\beta &= R_2^D + R_2^{CSA} - \eta_{xy} \end{aligned} \quad (3.16)$$

Le premier terme correspond à la relaxation dipolaire, R^D , le second à l'anisotropie de déplacement chimique, R^{CSA} , et le dernier à la relaxation cross corrélée, η_{xy} , qui s'ajoute ou se soustrait suivant que le noyau N observé est couplé à un H en état α ou β . Comme pour les mesures d'auto-relaxation, un délai de relaxation τ est appliqué, et la valeur de η_{xy} est extraite [Wang et al., 2003]. Comme l'intensité des signaux obtenus est proportionnelle à $\exp(-R_2^\alpha \tau)$ et $\exp(-R_2^\beta \tau)$, grâce à la formule suivante, il est possible d'extraire la valeur de η_{xy} :

$$\frac{I_\alpha}{I_\beta} = \exp(R_2^\beta - R_2^\alpha)\tau = \exp(\eta_{xy}\tau) \quad (3.17)$$

La valeur de relaxation cross-corrélée obtenue, η_{xy} , nous permet ensuite de calculer R_{ech} selon l'équation suivante, où κ est la moyenne tronquée de R_2/η_{xy} :

$$R_{ech} = R_2 - \kappa\eta_{xy} \quad (3.18)$$

5.4.4 Fonctions d'auto-corrélation et de densité spectrale

La fonction d'auto-corrélation $G(t)$ est la fonction qui décrit la perte de signal au cours du temps causée par la fluctuation aléatoire du système. Pour la diffusion isotrope d'une molécule rigide, elle peut être exprimée ainsi :

$$G(t) = \frac{1}{5} \exp\left(-\frac{t}{\tau_c}\right) \quad (3.19)$$

La transformée de Fourier de la fonction d'auto-corrélation est la fonction de densité spectrale, $J(\omega)$, qui représente l'échantillonnage des fréquences par la molécule considérée, telle que :

$$J(\omega) = \frac{2}{5} \left(\frac{\tau_c}{1 + \omega^2 \tau_c^2} \right) \quad (3.20)$$

Dans le cas de molécules réelles, le modèle libre permet d'exprimer la fonction de densité spectrale en fonction d'un second temps de corrélation, interne, τ_i , ainsi que d'un paramètre d'ordre S^2 , compris entre 0 (mouvement libre) et 1 (mouvement contraint) [Lipari and Szabo, 1982]. Lorsque $\tau_i \ll \tau_c$, la fonction de densité spectrale est alors exprimée selon l'équation suivante :

$$J(\omega) = \frac{2}{5} \left(\frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau_i'}{1 + \omega^2 \tau_i'^2} \right) \quad \text{où} \quad \frac{1}{\tau_i'} = \frac{1}{\tau_c} + \frac{1}{\tau_i} \quad (3.21)$$

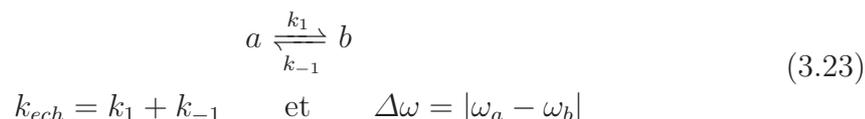
Il est possible d'exprimer cette fonction de densité spectrale pour les différentes expériences d'auto-relaxation et de relaxation croisée discutées précédemment [Peng and Wagner, 1992] :

$$\begin{aligned} R_1 &= d \cdot [J(\omega_{HN} - \omega_N) + 3J(\omega_N) + 6J(\omega_{HN} + \omega_N)] + c \cdot J(\omega_N) \\ R_2 &= \frac{d}{2} \cdot [4J(0) + J(\omega_{HN} - \omega_N) + 3J(\omega_N) \\ &\quad + 6J(\omega_{HN}) + 6J(\omega_{HN} + \omega_N)] + c \cdot \left(\frac{2}{3}J(0) + \frac{1}{2}J(\omega_N) \right) \\ NOE &= d \cdot [6J(\omega_{HN} + \omega_N) - J(\omega_{HN} - \omega_N)] \\ \text{Avec} \quad d &= \frac{\gamma_{HN}^2 \gamma_N^2 \hbar^2}{4r_{NH}^6} \quad \text{et} \quad c = \frac{\Delta^2 \omega_N^2}{3} \end{aligned} \quad (3.22)$$

Avec r la distance entre les deux spins considérés, Δ l'anisotropie de déplacement chimique et ω_H , ω_N , $\omega_H - \omega_N$ et $\omega_H + \omega_N$ les fréquences définies par les transitions du système de spins considéré.

5.5 L'échange chimique

L'échange entre plusieurs environnements chimiques d'un noyau apporte également une contribution à la relaxation transversale. L'échange peut avoir lieu par exemple entre deux conformations. Si on prend l'exemple d'un échange entre deux formes a et b avec une différence de fréquence de résonance entre elles $\Delta\omega$, on a :



En fonction des valeurs de k_{ech} et de $\Delta\omega$, on peut distinguer 3 régimes d'échange (Figure 3.11).

- $\Delta\omega \gg k_{ech}$, l'échange dit "lent", pour lequel on observe un signal distinct pour les deux formes présentes
- $\Delta\omega \approx k_{ech}$, l'échange dit "intermédiaire", pour lequel le signal est extrêmement élargi et presque invisible : c'est la coalescence.
- $\Delta\omega \ll k_{ech}$, l'échange dit "rapide", pour lequel un seul signal est observé, au déplacement chimique pondéré par les populations des états a et b .

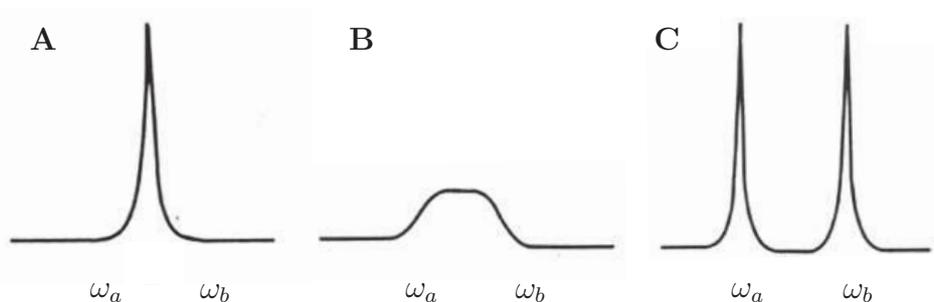


FIGURE 3.11 – Echange chimique entre deux états.

Evolution du spectre RMN en fonction de la vitesse d'échange k_{ech} entre deux états a et b de populations égales et de déplacements chimiques respectifs ω_a et ω_b . **A.** $\Delta\omega \ll k_{ech}$, **B.** $\Delta\omega \approx k_{ech}$, **C.** $\Delta\omega \gg k_{ech}$.

L'échange chimique peut être caractérisé par l'étude de la forme des raies, qui nous informe sur la dynamique de la protéine de la centaine de microsecondes à la seconde. D'autre part, l'échange va apporter une contribution non négligeable à la relaxation transversale R_2 qui va dépendre de la différence de déplacement chimique entre les deux états $\Delta\omega$, des populations des deux états, p_A et p_B , et de la vitesse d'échange k_{ech} .

Dans le cas limite d'un échange lent ou rapide, on aura respectivement :

$$\begin{aligned}
 R_2^{ech} &= p_B k_{ech} & \text{avec} & \frac{k_{ech}}{\Delta\omega} \rightarrow 0 \\
 R_2^{ech} &= \frac{p_A p_B \Delta\omega^2}{k_{ech}} & \text{avec} & \frac{k_{ech}}{\Delta\omega} \rightarrow \infty
 \end{aligned}
 \tag{3.24}$$

Pour pouvoir observer les états présents en solution, différentes techniques RMN ont été mises au point comme la dispersion de relaxation Carr-Purcell Meiboom-Gill (CPMG) [Carr and Purcell, 1954, Meiboom and Gill, 1958] et la relaxation dans le référentiel rotatif ($T_{1\rho}$) [Blackledge et al., 1993, Akke and Palmer, 1996] pour les échanges plutôt rapides (de la microseconde à la milliseconde) ou l'échange ZZ [Montelione and Wagner, 1989, Farrow et al., 1994] et le transfert de saturation de l'échange chimique (CEST) pour les échanges plus lents [Vallurupalli et al., 2012].

5.5.1 Transfert de saturation de l'échange chimique

Le CEST (*chemical exchange saturation transfer*) est une méthode qui permet d'obtenir des informations sur des régimes d'échange un peu plus lents que la dispersion de relaxation CPMG, entre 10 et 200 s⁻¹ [Vallurupalli et al., 2012]. En théorie, elle est particulièrement adaptée aux systèmes avec un état "invisible" ($p_b \ll p_a$), cependant, comme nous le verrons dans la suite de ce manuscrit, elle est également applicable à des systèmes dont les populations a et b sont plus équilibrées.

Le principe de cette méthode est d'observer l'état majoritaire (a) pour obtenir des informations sur l'état minoritaire (b). Pour ce faire, un champ \vec{B}_1 faible est appliqué à la fréquence de l'état minoritaire, perturbant sa magnétisation (Figure 3.12A,B). Cette dernière est transférée à l'état majoritaire durant le processus d'échange chimique, et provoque un changement d'intensité du signal de l'état majoritaire, pouvant donc être détecté dans le spectre (Figure 3.12C).

En pratique, le CEST est enregistré comme une série d'HSQC, et le champ \vec{B}_1 est appliqué successivement à travers toute la dimension azote. On obtient donc une série de plans. En représentant la variation d'intensité du signal de l'état a à travers tous ces plans (rapport I/I_0 où I_0 est l'intensité du signal dans un spectre de référence), on peut déterminer la différence de déplacement chimique entre les formes a et b (Figure 3.12D).

Dans le cas d'un échange entre deux états, l'analyse des profils de CEST nous permet d'obtenir les populations présentes, p_a et p_b , la constante d'échange k_{ech} , ainsi que la différence de relaxation entre les états a et b . Pour déterminer ces paramètres,

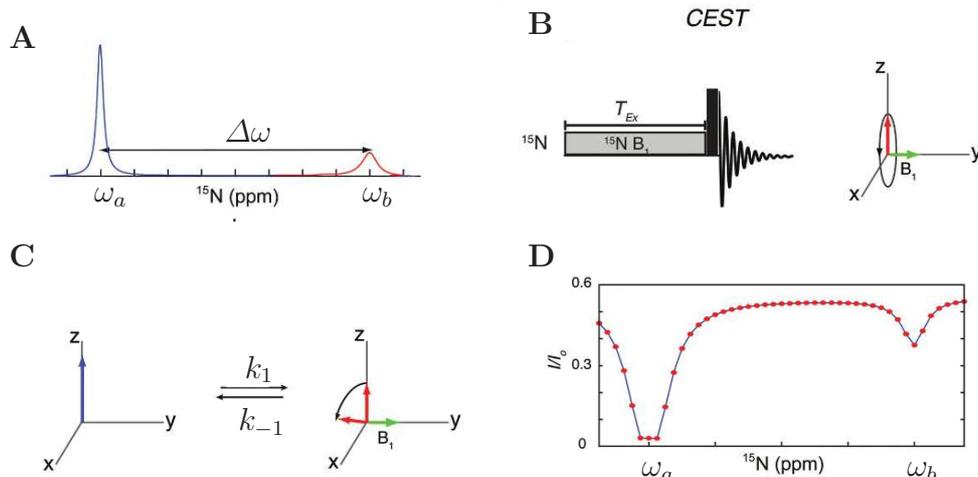


FIGURE 3.12 – Observation du phénomène d’échange chimique par CEST. **A.** Spectre ^{15}N d’un noyau en échange entre deux conformations, a majoritaire et b minoritaire. **B.** Séquence d’impulsion de l’expérience CEST et application d’un champ \vec{B}_1 faible. **C.** Evolution de la magnétisation d’un noyau en échange entre deux états perturbé par un champ \vec{B}_1 faible. **D.** Profil obtenu comme résultat d’une expérience CEST. Le rapport I/I_0 est représenté en fonction du décalage du champ \vec{B}_1 appliqué. Adapté de [Vallurupalli et al., 2012].

les intensités observées peuvent être simulées par intégration des équations de Bloch-McConnell [McConnell, 1958]. Au laboratoire, nous utilisons le logiciel ChemEx [Vallurupalli et al., 2012], et les paramètres sont ajustés jusqu’à obtenir le meilleur accord possible entre les données expérimentales et les données simulées, c’est à dire la minimisation de la fonction χ^2 suivante :

$$\chi^2(\xi) = \sum \left(\frac{I_i^{exp} - I_i^{calc}(\xi)}{\sigma_i^{exp}} \right)^2 \quad (3.25)$$

Avec I_i^{exp} et I_i^{calc} les intensités du pic i expérimentales et calculées respectivement, ξ les paramètres utilisés et σ_i^{exp} l’erreur expérimentale sur l’intensité du pic i .

5.6 L’augmentation paramagnétique de la relaxation

Si on attache un électron libre, dont le moment gyromagnétique est environ 650 fois celui du proton, à un groupement de notre protéine, on obtient une augmentation de la relaxation dipolaire dépendante de la distance à l’électron ajouté (en r^{-6}) et qui peut être détectée jusqu’à une distance d’environ 25 Å (Figure 3.13A). L’effet est environ 100 fois plus important sur le proton que sur l’azote (équation 3.27), c’est pourquoi l’effet sur l’azote est généralement négligé, et c’est l’effet sur le proton qui

est étudié. L'augmentation de relaxation produite touche R_1 et R_2 , cependant, l'augmentation de R_1 n'est que très faible et est donc négligée. Comme l'augmentation de la relaxation R_2 se traduit par un élargissement des raies, il est possible d'utiliser l'intensité pour approximer la valeur de R_2 , selon l'équation suivante [Battiste and Wagner, 2000] :

$$\frac{I_{ox}}{I_{red}} = \frac{R_2 \exp(-R_2^{sp}t)}{R_2 + R_2^{sp}} \quad (3.26)$$

Avec I_{ox} et I_{red} l'intensité d'une résonance oxydée ou réduite respectivement, R_2 la relaxation transversale de l'amide réduit respectivement, t le délai d'évolution de la magnétisation, et R_2^{sp} la contribution à la relaxation du spin paramagnétique.

Une fois R_2^{sp} obtenue, la distance électron-spin nucléaire r peut être extraite en utilisant l'équation suivante, dérivée de l'équation de Solomon-Bloembergen [Battiste and Wagner, 2000] :

$$r = \left[\frac{K}{R_2^{sp}} \left(4\tau_c + \frac{3\tau_c}{1 + \omega_H^2 \cdot \tau_c^2} \right) \right]^{1/6} \quad \text{avec} \quad K = \frac{1}{15} S(S+1) \gamma^2 g^2 \beta^2 \quad (3.27)$$

Avec τ_c le temps de corrélation de l'interaction électron-spin nucléaire, ω_H la fréquence de Larmor du spin nucléaire, γ le moment gyromagnétique du noyau, g le facteur électronique et β le magnéton de Bohr.

En pratique, on utilise des sondes paramagnétiques telles que le MTSL ou TEMPO qui se couplent sur des cystéines (voir chapitre V paragraphe 1.2) (Figure 3.13B,C). On enregistre un spectre avec la sonde oxydée, et un avec la sonde réduite, qui ne possède donc plus d'électron libre. Il suffit ensuite de faire le rapport des intensités obtenues pour avoir une mesure indirecte de la distance sonde-résidu observé.

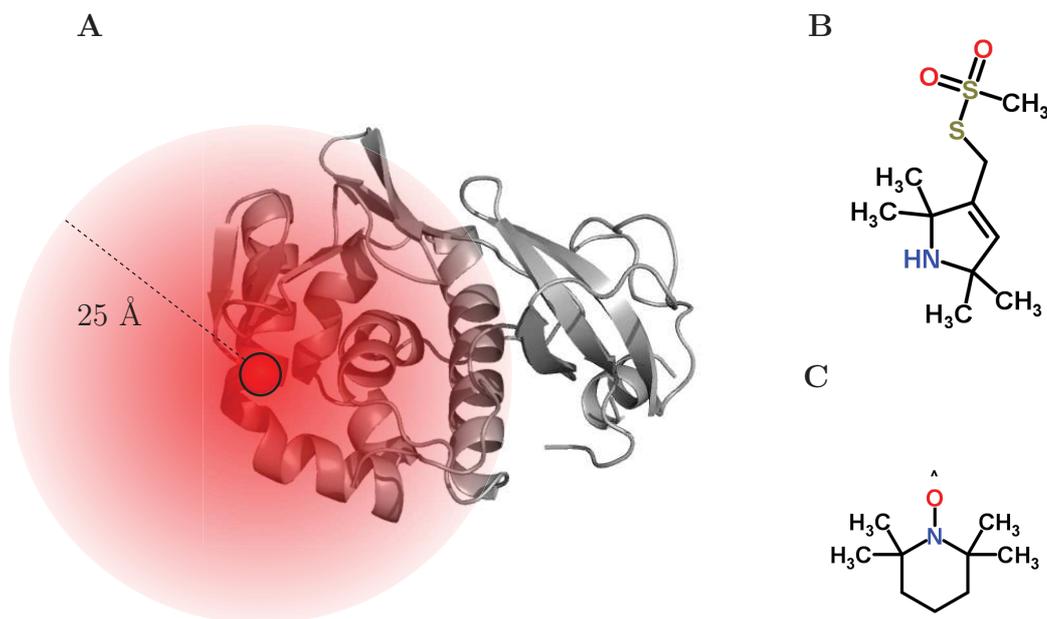


FIGURE 3.13 – Expériences d’augmentation paramagnétique de la relaxation (PRE).

A. Une sonde (rond noir) est attachée à la protéine et a un effet dépendant de la distance (dégradé de rouge) sur la relaxation dans un rayon d’environ 25 Å.

B. MTSL : S-(2,2,5,5-tetraméthyle-2,5-dihydro-1H-pyrrol-3-yl) méthyle méthane-sulfonothioate

C. TEMPO : 2,2,6,6-Tetramethylpiperidine 1-oxyl

6 Limitations de taille

La poids moléculaire est un facteur déterminant pour la qualité des mesures en RMN des protéines en solution. En effet, plus une protéine est grosse, plus sa réorientation en solution sera lente, son temps de corrélation τ_c élevé, et donc son temps de relaxation T_2 court. De plus, le grand nombre d’atomes présents va causer un grand nombre d’interactions dipolaires, réduisant également le temps de relaxation. Pour une protéine globulaire, il existe une corrélation directe entre la longueur de la chaîne d’acides aminés et la relaxation R_2 , c’est à dire que plus la protéine est grosse, plus les raies observées seront larges (Figure 3.10). Différents moyens existent pour remédier à cette limitation. Dans mon cas, j’ai utilisé la deutération couplée à des séquences de type TROSY, comme expliqué ci-après.

6.1 La deutération

La deutération permet de limiter les interactions dipolaires externes entre les protons (à travers l’espace) ainsi que de réduire les interactions dipolaires internes entre le proton et le noyau auquel il est lié (par exemple dans le cas des carbones

α ou β), et donc l'augmentation de la relaxation transversale qui leur est due. Le signal obtenu est donc moins large. Il existe différents schémas de deutération.

La deutération aléatoire partielle (en général jusqu'à 80 %) des chaînes latérales et des liaisons peptidiques, grâce à un milieu de culture minimum enrichi en D_2O . Cette technique crée des isotopomères (variation de l'environnement local) car toutes les molécules ne sont pas deutérées de la même façon. Cette méthode est peu coûteuse et permet d'améliorer le signal de façon significative. De plus, les isotopomères sont utiles lors d'études structurales car ils permettent de distinguer certaines positions normalement équivalentes. Les contributions externes de relaxation sont réduites, T_1 et T_2 augmentent.

La perdeutération grâce à un milieu de culture minimum dans 100 % de D_2O et l'utilisation d'une source de carbone deutérée. Dans ce cas, tous les protons sont remplacés par des deutérium. La protéine d'intérêt est purifiée dans un tampon en H_2O , ce qui permet le contre-échange des deutériums en protons uniquement pour les liaisons amides ($^H N$), les chaînes latérales des asparagines, glutamines et arginines, et les OH labiles. Les contributions externes de relaxation sont fortement réduites, T_1 et T_2 augmentent.

Il est à noter que l'introduction de deutérium a une influence sur le déplacement chimique des atomes qui y sont liés (*isotope shift effect*), la variation observée pouvant atteindre 1,2 ppm pour certains carbones β [Maltsev et al., 2012].

6.2 Les séquences TROSY

Certaines séquences spécifiques, comme les séquences TROSY (*transverse relaxation optimized spectroscopy*) [Pervushin et al., 1997], permettent d'obtenir des raies plus fines pour des protéines de grande taille. En effet, lorsqu'on enregistre la corrélation $^N H-N$ sans appliquer de découplage, on obtient quatre signaux, correspondant aux états $\alpha\alpha$, $\alpha\beta$, $\beta\alpha$ et $\beta\beta$, de relaxations et donc de formes différentes. Comme expliqué au paragraphe 5.4, le temps de relaxation transversale T_2 de l'azote est la somme des phénomènes de relaxation causés par l'anisotropie de déplacement chimique et par l'interaction dipole-dipole. Le premier dépend de la force du champ \vec{B}_0 , mais le second en est indépendant. A haut champ \vec{B}_0 (optimum à 950 MHz), la corrélation croisée de ces deux phénomènes rend la relaxation transversale T_2 minimale pour l'un des quatre pics observés. C'est celui-ci que les séquences TROSY sélectionnent, grâce à un cyclage de phase particulier. Cette sélection cause une perte de signal d'un facteur 4, mais celle-ci est compensée par le gain de résolution offert par la méthode.

Le TROSY est particulièrement efficace pour des échantillons deutérés, car en plus d'obtenir une raie fine dans la dimension azote, l'interaction dipole-dipole entre protons est grandement réduite du fait de la deutération, ce qui diminue la relaxation du proton et donc la largeur de raie dans cette dimension.

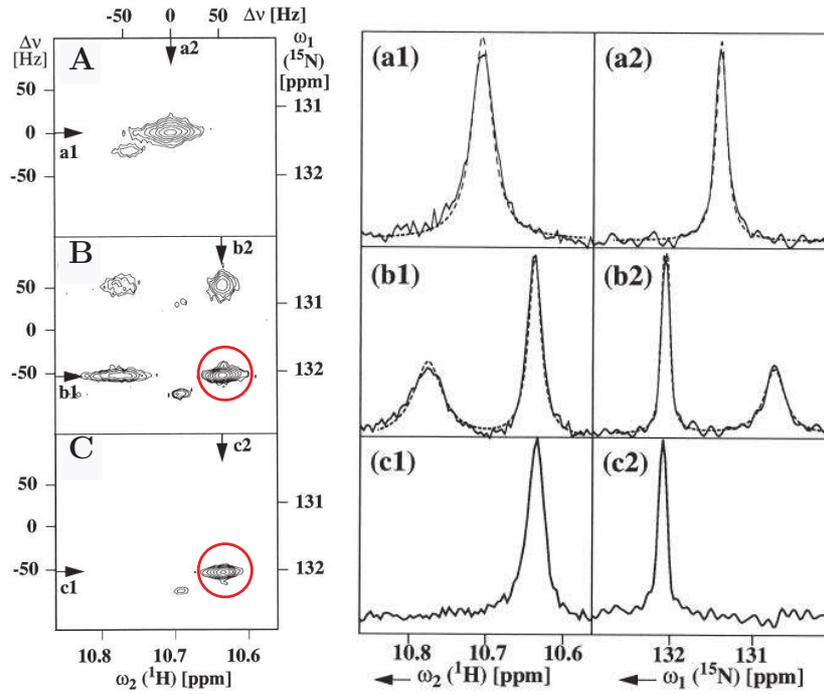


FIGURE 3.14 – Séquence TROSY.

A gauche, spectre de corrélation ^1H -N. A droite, coupe du spectre selon les axes indiqués, permettant d'observer la forme de raie. **A.** Corrélation ^1H -N avec découplage. **B.** Corrélation ^1H -N sans découplage. Les quatre pics présentent des formes différentes. **C.** TROSY sélectionnant directement le pic le plus fin (rouge). Adapté de [Pervushin et al., 1997].

Chapitre IV

Autres techniques

Plus de deux tiers des protéines procaryotes et une encore plus grande proportion des protéines eucaryotes sont composées de deux domaines ou plus [Vogel et al., 2004, Marsden et al., 2006]. La dynamique de ces protéines multi-domaines est essentielle pour leur fonction. En effet, ces protéines subissent généralement plusieurs changements conformationnels, par exemple lors de l'interaction avec un partenaire (protéine, ARN, ADN...) ou sous l'effet de modifications post-traductionnelles (kinases, nucléosomes, protéine G...) [Del Sol et al., 2007, Mackereth and Sattler, 2012, Tokunaga et al., 2014]. Chacune des formes existantes est utile pour la fonction de la protéine et, dans certains cas, les différents domaines peuvent accomplir différentes fonctions complémentaires.

Les techniques de résolution de structure telles que la diffraction des rayons X ou encore la microscopie électronique apportent un grand nombre d'informations quant aux structures possibles de ces protéines multi-domaines, seules ou en complexe. Cependant, elles ne nous renseignent pas sur un élément essentiel à leur fonction : leur dynamique. C'est pourquoi des techniques biophysique d'étude de la structure en solution sont de plus en plus couramment utilisées de façon complémentaire, permettant la résolution de la structure et de la dynamique de protéines multi-domaines et de complexes de plus en plus grands, ainsi qu'une meilleure compréhension des mécanismes de fonctionnement de ces protéines.

PB2 est une protéine composée de 9 domaines : du N-terminal au C-terminal, on trouve le domaine d'interaction avec PB1, le domaine N1, le domaine du couvercle hélical, le domaine N2, le domaine "du milieu", le domaine d'interaction avec la coiffe, un domaine de liaison, le domaine 627 et le domaine NLS [Reich et al., 2014]. C'est au double domaine C-terminal 627-NLS que nous avons choisi de nous intéresser lors de ma thèse. Cependant, il était clair que son étude par cristallographie aux rayons X ne suffisait pas à expliquer comment ce double domaine peut fonctionner

in vivo, les différentes structures précédemment obtenues étant incompatibles entre elles [Tarendeau et al., 2007, Tarendeau et al., 2008]. Des études de microscopie électronique avaient également été effectuées sur la polymérase entière, ne permettant pas de replacer de façon précise les différents domaines contenus dans le complexe, dont la structure a été récemment résolue par cristallographie aux rayons X [Area et al., 2004, Coloma et al., 2009, Reich et al., 2014].

C'est pourquoi, lors de ma thèse, nous avons utilisé diverses techniques biophysiques apportant différentes informations sur le comportement en solution du double domaine 627-NLS. Leur complémentarité nous a permis de décrire pour la première fois la dynamique du 627-NLS en solution, de son interaction avec l'importine α , et de proposer un certain nombre d'implications biologiques *in vivo* de nos découvertes. Dans ce chapitre, je présenterai brièvement les différentes techniques utilisées : leur principe, des considérations expérimentales, le traitement des données et les informations pouvant être obtenues.

1 Diffusion des rayons X aux petits angles

La diffusion des rayons X aux petits angles (*small angle X-ray scattering*, SAXS) apporte une information structurale à basse résolution pour des échantillons en solution [Putnam et al., 2007, Svergun et al., 2001]. Elle est généralement utilisée sur des systèmes ne permettant pas l'accès aux techniques à haute résolution (RMN, diffraction des rayons X), ou en complément de ces dernières.

1.1 Principe

Lorsqu'un faisceau de rayons X rencontre une solution de macromolécules biologiques, il diffuse de façon élastique (sans changement de longueur d'onde) en rencontrant les électrons présents dans la solution. Le SAXS est une méthode de contraste, c'est à dire que le signal de diffusion correspondant à la molécule d'intérêt $\Delta\rho(\vec{r})$ est dérivé de la différence moyenne de la densité d'électrons entre la solution contenant la molécule d'intérêt ($\rho(\vec{r})$) et le tampon seul (blanc, ρ_s) :

$$\Delta\rho(\vec{r}) = \rho(\vec{r}) - \rho_s \quad (4.1)$$

L'intensité diffusée est enregistrée en fonction de l'angle de diffusion 2θ , et dépend du vecteur de diffusion, \vec{q} (Figure 4.1A) selon la formule suivante :

$$\|\vec{q}\| = q = \frac{4\pi \sin \theta}{\lambda} \quad (4.2)$$

Comme les particules en solution s'orientent librement, seul l'angle auquel l'observation est effectuée importe. L'intensité mesurée est donc moyennée sur tous les

rayons observés pour obtenir la fonction de diffusion (Figure 4.1B). Il est à noter qu'en SAXS, comme les molécules diffusent librement en solution, les phases du signal sont moyennées et donc inaccessibles. Ce n'est pas le cas en diffraction des rayons X, bien que l'obtention des phases selon cette dernière technique ne soit pas directe.

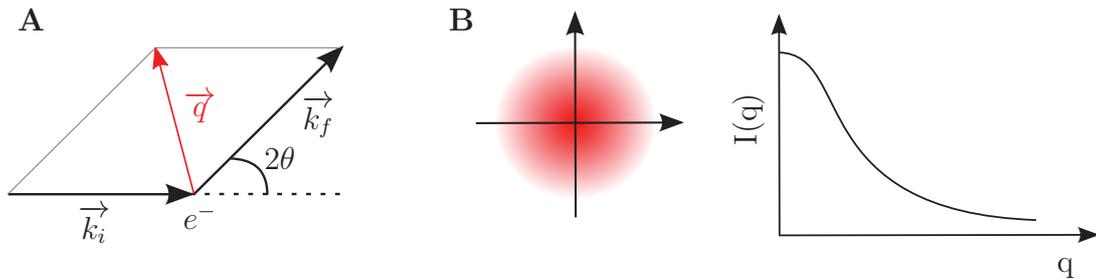


FIGURE 4.1 – Phénomène de diffusion.
A. Lorsque l'onde initiale \vec{k}_i est diffusée par un électron (e^-), elle produit l'onde diffusée \vec{k}_f . La différence de ces deux vecteurs est le vecteur de diffusion, \vec{q} . **B.** Une moyenne sur tous les rayons est effectuée pour passer des intensités enregistrées (à gauche) à la fonction de diffusion (à droite).

1.2 Expériences

Le SAXS nécessite des échantillons monodisperses et idéaux, c'est à dire composés de particules pures qui n'interagissent pas entre elles (dilution infinie). En pratique, cette dernière condition est vérifiée en effectuant plusieurs mesures à différentes concentrations et en vérifiant que la fonction de diffusion obtenue est indépendante de la concentration. De plus, des systèmes de chromatographie d'exclusion de taille couplés au SAXS se développent, permettant un meilleur contrôle de la qualité de l'échantillon pour les systèmes sensibles. La diffusion du tampon est également enregistrée et sera soustraite du signal de l'échantillon puisque les molécules présentes (eau, sel, ...) diffusent également les rayons X (Equation 4.1).

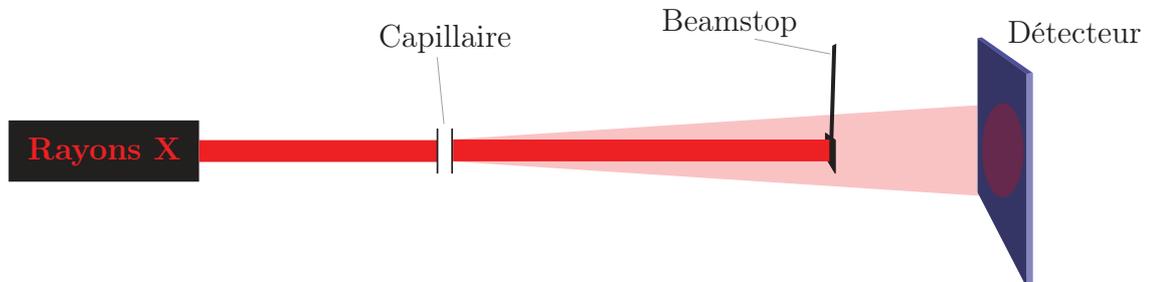


FIGURE 4.2 – Dispositif expérimental pour les expériences de SAXS.

Au niveau du dispositif expérimental, le faisceau de rayons X est collimaté et projeté sur l'échantillon qui passe à travers un capillaire transparent aux rayons X. Comme les rayons X sont destructifs pour les macromolécules biologiques, une série de clichés de diffusion correspondant à l'exposition de l'échantillon pendant un temps court sont enregistrés et seront moyennés par la suite. Un *beamstop* est situé entre l'échantillon et le détecteur pour éviter sa destruction par le faisceau direct (non diffusé) de rayons X (Figure 4.2).

1.3 Analyse des données

1.3.1 Loi de Guinier

La loi de Guinier permet d'approximer la courbe de diffusion à proximité de l'origine, et s'exprime ainsi :

$$I(q) = I(0) \exp\left(\frac{-R_g q^2}{3}\right) \quad \text{d'où} \quad \ln(I(q)) = \ln(I(0)) - \frac{R_g q^2}{3} \quad (4.3)$$

Avec $I(q)$ la fonction de diffusion, $I(0)$ sa valeur extrapolée pour un angle de 0, q l'angle de mesure, et R_g le rayon de giration de la molécule observée. Lorsque la fonction de diffusion est représentée comme $\ln(I)$ en fonction de q^2 , le R_g peut être extrait à partir de la pente ($-R_g/3$) dans la partie linéaire de la courbe, pour q petit (Figure 4.3A). Cette représentation permet également de repérer l'éventuelle inhomogénéité de l'échantillon : en effet, s'il est agrégé, la courbe augmente de façon non linéaire à l'approche de $q = 0$ (Figure 4.3B). Dans ce cas, l'exploitation des données n'est pas possible, la qualité de l'échantillon doit être améliorée. Le rayon de giration est une mesure de la taille générale de la protéine, R_g^2 étant défini comme la moyenne quadratique de la distance au centre de masse de la protéine pondérée par le contraste de densité électronique entre la protéine et le solvant.

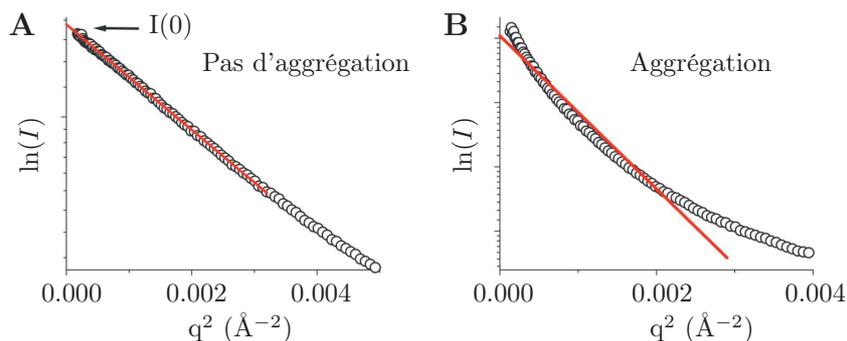


FIGURE 4.3 – Représentation de Guinier des données SAXS.

A. Courbe pour un échantillon non-agrégé. La région de Guinier utilisée pour le calcul de R_g et l'extrapolation de $I(0)$ est représentée en rouge. **B.** Courbe pour un échantillon agrégé. Il n'est pas possible de définir une région de Guinier, l'analyse est impossible. Adapté de [Putnam et al., 2007]

L'intensité à l'origine $I(0)$ normalisée par la concentration de l'échantillon permet d'estimer la masse moléculaire de l'objet présent en solution. C'est également un moyen de vérifier l'idéalité et la monodispersité de l'échantillon.

1.3.2 Fonction de distribution des paires

La fonction de distribution des paires (*pair distribution function*, $P(r)$) est la fonction d'auto-corrélation qui est calculée à partir du signal de diffusion par une transformation de Fourier :

$$P(r) = \frac{r^2}{2\pi^2} \int_0^\infty q^2 I(q) \frac{\sin qr}{qr} dq \quad (4.4)$$

Le résultat informe directement sur les distances entre les électrons des particules diffusant les rayons X en solution, et permet de déterminer une distance, D_{max} , qui correspond à la distance maximale enregistrée dans la molécule observée (Figure 4.4A).

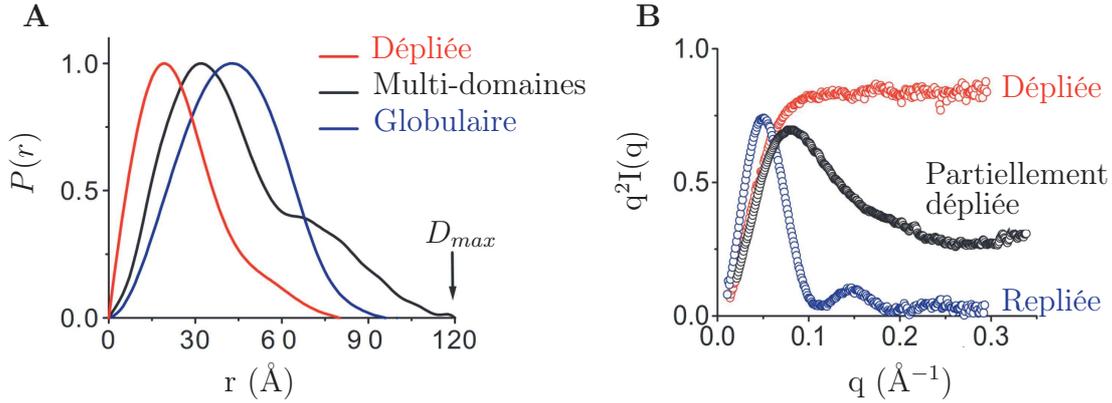


FIGURE 4.4 – Représentations des données.

A. Fonction de distribution de paires de molécules de différentes formes et détermination de la plus grande distance existante, D_{max} . **B.** Représentation de Kratky des données obtenues pour des molécules repliées ou non. Adapté de [Putnam et al., 2007]

La fonction de distribution des paires permet également de repérer la présence d'agrégation (si la partie des grandes distances présente une "queue"), et de calculer R_g et $I(0)$, validant les valeurs calculées grâce à la loi de Guinier, selon les formules suivantes :

$$R_g^2 = \frac{\int_0^{D_{max}} r^2 P(r) dr}{2 \int_0^{D_{max}} P(r) dr} \quad \text{et} \quad I(0) = 4\pi \int_0^{D_{max}} P(r) dr \quad (4.5)$$

1.3.3 Représentation de Kratky

La représentation de Kratky permet d'identifier les protéines dépliées. En effet, selon la loi de Porod, il est possible de calculer le volume V de la molécule observée selon l'équation suivante :

$$V = 2\pi^2 \frac{I(0)}{Q} \quad \text{avec} \quad Q = \int_0^\infty q^2 I(q) dq \quad (4.6)$$

Où Q est l'invariant de Porod [Porod, 1982]. En représentant $q^2 I(q)$ en fonction de q , il est possible d'obtenir des informations sur le repliement de la protéine étudiée les molécules (Figure 4.4B). Par exemple, les molécules repliées ont une courbe en cloche.

1.4 Modélisation

Différentes modélisations sont possible à partir des données de SAXS suivant les informations déjà disponibles à propos du système étudié :

- La reconstruction *ab initio* : dans le cas où aucune structure atomique n'est disponible , il est possible de reconstruire une enveloppe basse résolution avec un logiciel comme GASBOR [Svergun et al., 2001].
- La comparaison des données expérimentales à des données prédites à partir de modèles atomiques grâce à des logiciels tels que CRY SOL ou FOXS [Svergun et al., 1995, Schneidman Duhovny et al., 2010]. C'est cette approche qui sera utilisée dans ce manuscrit : des ensembles de conformations seront générés, leur courbes de SAXS prédites, moyennées, puis comparées aux courbes expérimentales.
- Le "rigid body modeling" couplé à la reconstruction *ab initio* : si une partie de la structure atomique est disponible, comme deux domaines reliés par un "linker" flexible, ou plusieurs domaines dont la structure de l'un est connue, des logiciels tels que BUNCH ou CREDO permettent une reconstruction partielle et une modélisation du système [Petoukhov and Svergun, 2005, Petoukhov et al., 2002].

1.5 Conclusion

Le SAXS est une méthode rapide et simple, en solution, qui permet d'apporter un certain nombre d'informations structurales à basse résolution, sans limite de taille. Combiné avec des méthodes à haute résolution telles que la RMN et la diffraction aux rayons X, le SAXS permet très souvent d'obtenir un modèle du système étudié. C'est une méthode peu gourmande en matériel biologique, qui requiert environ 100 μ L et moins d'un milligramme pour une mesure. Cependant, l'obtention d'échantillons d'assez bonne qualité se révèle parfois être un problème. Par ailleurs, une analyse minutieuse est nécessaire pour ne pas surestimer la quantité d'informations que les données peuvent fournir.

2 Diffusion des neutrons aux petits angles

2.1 Principe

Le principe physique général de la diffusion des neutrons aux petits angles (*small angle neutron scattering*, SANS) est similaire à celui du SAXS, à ceci près qu'à la place d'électrons diffusant des rayons X, ce sont les noyaux qui diffusent des neutrons (Figure 4.1). De plus, et contrairement au SAXS, il est impossible d'avoir un contenant pour l'échantillon qui soit transparent aux neutrons. Les diffusions de la cellule et du solvant seul devront donc être soustraites de la diffusion totale de l'échantillon. Après soustraction, la diffusion obtenue peut être exprimée ainsi :

$$I(q) = \sum |\Delta\bar{\rho} \exp(-i(q \cdot r)) dr|^2 \quad \text{avec} \quad \Delta\bar{\rho} = \bar{\rho}_{\text{molécule}} - \bar{\rho}_{\text{solvant}} \quad (4.7)$$

Avec $I(q)$ l'intensité de la diffusion, q l'angle de la diffusion, r la distance radiale au centre d'un corps diffusant, et $\Delta\bar{\rho}$ la densité de longueur de diffusion (*scattering length density*, SLD) moyenne de la molécule moins celle du solvant. En effet, le pouvoir diffusant d'une particule est pondéré par une constante, la SLD. Cette dernière est une caractéristique intrinsèque de chaque noyau et représente son pouvoir diffusant. La différence entre la SLD du système et celle du solvant est le contraste. Dans le cas du SAXS, tous les électrons ont la même SLD. Dans le cas des noyaux, le proton (^1H) ayant une SLD de diffusion incohérente (destructrice) très élevée par rapport aux autres noyaux, un solvant deutéré est préféré lorsque le système étudié le permet.

2.2 Variation de contraste

Il est possible, en choisissant les noyaux présents dans le solvant et dans les macromolécules étudiées, de contrôler le contraste dans l'échantillon étudié. La différence la plus utilisée en SANS sur des échantillons biologiques est la différence de diffusion cohérente entre le proton (^1H , $SLD = -0,3741 \cdot 10^{-12}$ cm) et le deuterium (^2H , $SLD = 0,6671 \cdot 10^{-12}$ cm) (Figure 4.5A). En jouant sur la deutération des molécules en solution et sur la contenance en deuterium du solvant, il est donc possible de rendre invisible une molécule lors de la mesure (Figure 4.5B). Ce phénomène est particulièrement utile lors de l'étude de complexes. En général, la deutération requise pour faire "disparaître" une protéine dans un tampon 100 % D_2O est de 75 %, et une protéine hydrogénée n'aura aucun signal dans un tampon contenant 40 % de D_2O . Ces valeurs sont calculables en faisant la somme des SLD de tous les atomes de la macromolécule d'intérêt. De façon pratique, des serveurs en ligne sont disponibles, tel que celui de la source ISIS (Angleterre). Les valeurs calculées sont ensuite vérifiées expérimentalement lors de l'acquisition des données.

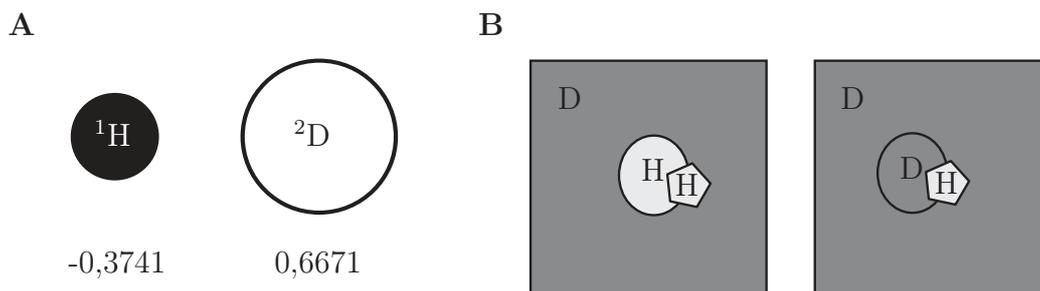


FIGURE 4.5 – Variation de contraste.

A. Représentation de la SLD du proton et du deutérium. Valeurs en 10^{-12} cm. **B.** A gauche, un complexe protonné (H) dans un solvant deutéré (D). Le complexe entier est visible. A droite, un complexe dont un seul des partenaires est protonné (H), l'autre étant deutéré (D), dans un solvant deutéré (D). Dans ce cas, le partenaire deutéré est invisible, car la différence de SLD entre celui-ci et le solvant est nulle.

2.3 Conclusion

Le SANS, et plus particulièrement la variation de contraste, est une technique qui permet d'observer la structure à basse résolution d'une seule molécule d'un complexe, en solution. Cependant, les mesures requièrent de grandes quantités d'échantillon (environ 200 μL et plus d'un milligramme par mesure) et une bonne stabilité, car le SANS est moins sensible que le SAXS du fait de la faible densité en noyaux comparée à celle en électron pour un même échantillon. Ces deux méthodes sont souvent couplées car elles sont complémentaires, et le SANS n'étant pas destructif, il est possible d'enregistrer des mesures SANS puis SAXS sur un même échantillon.

3 Transfert d'énergie par résonance de Förster sur molécules uniques (smFRET)

3.1 Principe

Lorsqu'une molécule organique comportant un système d'électrons π délocalisés (fluorophore) absorbe un photon d'une longueur d'onde appropriée (UV-visible en général), elle est excitée et passe de l'état électronique fondamental S_0 au premier état électronique excité S_1 (spins électroniques anti-parallèles). Elle peut ensuite revenir à l'état fondamental par l'émission d'un photon (phénomène de fluorescence) ou passer à un état triplet non fluorescent T_1 (spin électroniques parallèles, figure 4.6A).

Si deux fluorophores pour lesquels la raie d'émission de l'un recouvre partiellement la raie d'absorption de l'autre se trouvent proches dans l'espace, alors, par une interaction dipole-dipole, le donneur peut transférer son énergie à l'accepteur par

un processus non-radiatif appelé FRET (Figure 4.6B,C). L'accepteur retourne alors à l'état fondamental par l'émission d'un photon ou par tout autre processus de désexcitation possible, comme décrit plus haut. Le FRET dépend de la distance entre les fluorophores impliqués : plus ils sont proches, plus le transfert est efficace. Les molécules donneuses et accepteurs utilisées lors de ce travail sont Alexa488 et Alexa594 respectivement (4.6D).

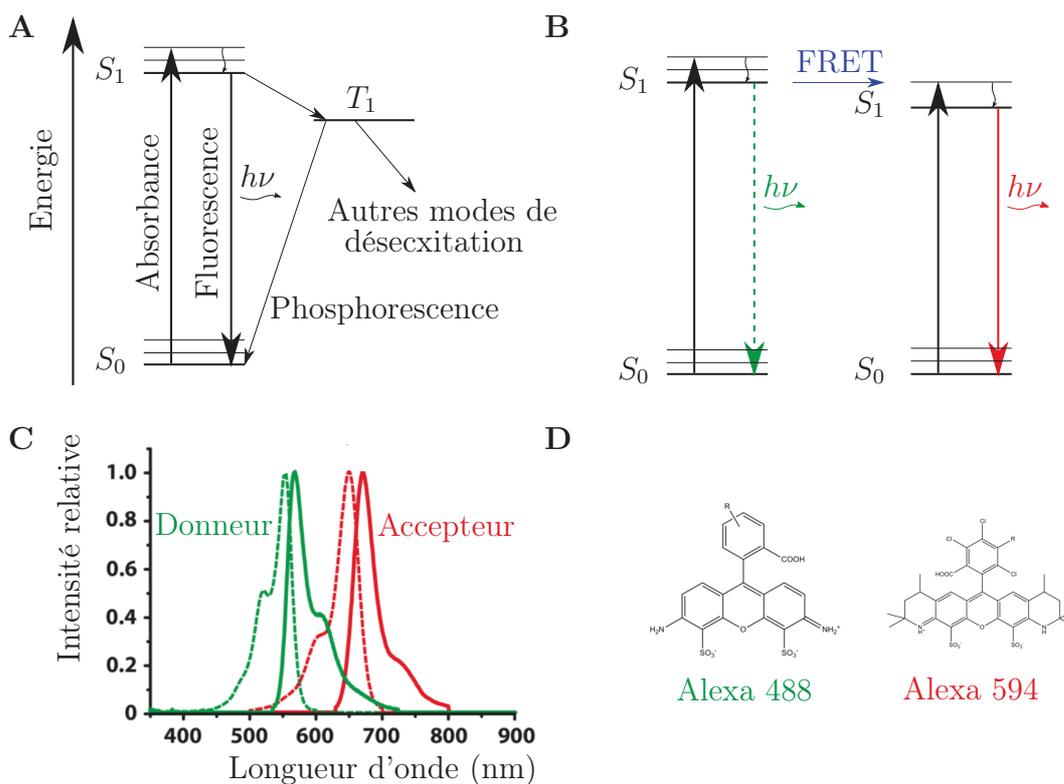


FIGURE 4.6 – Principe du transfert d'énergie par résonance de Förster (FRET). Représentations des niveaux énergétiques d'une molécule et des transitions entre eux (diagrammes de Jablonsky). S_0 et S_1 sont l'état électronique fondamental et premier état excité respectivement, chacun possédant des niveaux d'énergie vibrationnels se désexcitant par conversion interne. T_1 est l'état triplet. CI : croisement inter-système. **A.** Excitation et retour à l'état fondamental d'un fluorophore. **B.** Phénomène de transfert d'énergie par résonance de fluorescence. Le fluorophore donneur (vert) transfère son énergie au fluorophore accepteur (rouge), qui se désexcite ensuite. **C.** Spectre d'absorption (pointillé) et d'émission (trait plein) d'un fluorophore donneur (vert) et d'un accepteur (rouge). **D.** Fluorophores donneur (vert) et accepteur (rouge) utilisés lors de ce travail. Adapté de [Gust et al., 2014]

3.2 Paramètres

L'efficacité (E) du processus de FRET peut être exprimée comme suit :

$$E = \frac{F^D(A)}{F^D(A) + F^D(D)} \quad (4.8)$$

Avec $F(A)$ et $F(D)$ la fluorescence de l'accepteur et du donneur respectivement lorsque le donneur est excité.

La durée de vie de l'état excité correspond au temps caractéristique durant lequel le fluorophore reste excité, avant de revenir à l'état fondamental. Cependant, cette durée est raccourcie lorsqu'un autre processus de retour à l'état fondamental est présent, tel que le FRET. Selon l'efficacité du FRET E , la durée de vie de l'état excité est plus ou moins raccourcie. Ce paramètre permet donc le calcul de l'efficacité de FRET selon la formule suivante :

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad (4.9)$$

Avec $\tau_{D(A)}$ et $\tau_{D(0)}$ la durée de vie de l'état excité du donneur respectivement en présence et en absence de l'accepteur.

Le transfert d'énergie entre la paire de fluorophores choisis dépend de la distance selon l'équation suivante :

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \quad (4.10)$$

Avec R la distance entre les deux fluorophores et R_0 le rayon de Förster, défini tel que pour $R = R_0$, $E = 0,5$, calculé comme suit :

$$R_0 = 9780(n^{-4} \cdot \kappa^2 \cdot \phi_D \cdot J)^{1/6} \text{ \AA} \quad (4.11)$$

Avec n l'indice de réfraction du tampon (typiquement 1,3 pour les solutions aqueuses), κ^2 le facteur d'orientation des dipôles des fluorophores, minimal pour deux dipôles perpendiculaires et maximal pour des dipôles anti-parallèles, ϕ_D le rendement quantique du donneur en absence de l'accepteur et J l'intégrale du recouvrement entre les spectres d'émission du donneur et d'excitation de l'accepteur. κ^2 est approximé à $2/3$ lorsque les fluorophores ont une rotation libre en solution. On utilise donc des fluorophores couplés à un bras espaceur pour permettre leur libre rotation, cette condition étant vérifiée indépendamment. Les distances ainsi mesurées ne correspondent pas à la distance entre les points d'attache des fluorophores (en général cystéines) mais entre les centres de leurs moments dipolaires. Cependant, il n'existe pas pour le moment de méthode simple pour prendre en compte ce problème dans l'analyse des données.

Le rendement quantique est une caractéristique intrinsèque des fluorophores. Il est défini par l'équation 4.12. Cependant, ce rendement peut être modulé lorsque la molécule est attachée à une molécule biologique, par exemple par son environnement immédiat (résidus protéiques, tampons, ...). Il est donc important pour la bonne interprétation des données de le mesurer, une fois le fluorophore lié à la protéine d'intérêt, dans les conditions réelles d'expérience.

$$\phi = \frac{\text{nombre de photons émis}}{\text{nombre de photons absorbés}} \quad (4.12)$$

Enfin, l'intégrale de recouvrement des spectres d'émission du donneur et d'absorption de l'accepteur doit également être déterminée expérimentalement pour le calcul de la distance inter-fluorophores. En effet, les spectres d'absorbance et de fluorescence d'une molécule donnée changent légèrement en fonction du tampon, et ceci doit être pris en compte.

3.3 Mesures sur molécules uniques

La mesure sur molécules uniques permet d'obtenir pour chaque molécule observée une efficacité, un temps de vie de l'état excité ainsi qu'une stoechiométrie de mesure. Les avantages de ce type de mesure par rapport à une mesure de FRET classique, moyennant directement les observables, sont nombreux. En effet, en enregistrant des données individuelles sur plusieurs centaines de molécules, il devient notamment possible de distinguer les différences de marquages existantes ainsi que la présence de plusieurs populations. C'est pourquoi ces mesures requièrent l'utilisation d'un microscope, un spectromètre classique ne possédant pas la résolution requise.

Dans le cadre de cette thèse, les mesures ont été effectuées sur un microscope confocal à deux couleurs construit sur mesure (laboratoire Lemke, Heidelberg), similaire au schéma présenté dans la figure 4.7. Brièvement, deux lasers d'excitation, rouge et vert, sont combinés et alternés pour exciter une molécule plusieurs fois avec chaque laser pendant sa diffusion à travers le volume confocal. La fluorescence observée est limitée par un diaphragme, puis séparée par un filtre dichroïque sur deux canaux de détection, un pour le donneur (vert) et un pour l'accepteur (rouge). Chaque photon enregistré peut être attribué à quatre canaux différents, chaque type d'évènement possible étant ainsi séparé :

- excitation de l'accepteur/émission du donneur (bleu)
- excitation de l'accepteur/émission de l'accepteur (rouge)
- excitation du donneur/émission du donneur (vert)
- excitation du donneur/émission de l'accepteur (FRET, noir)

Pour obtenir une bonne statistique et ainsi permettre une analyse fiable, les données sont enregistrées pendant une à plusieurs heures.

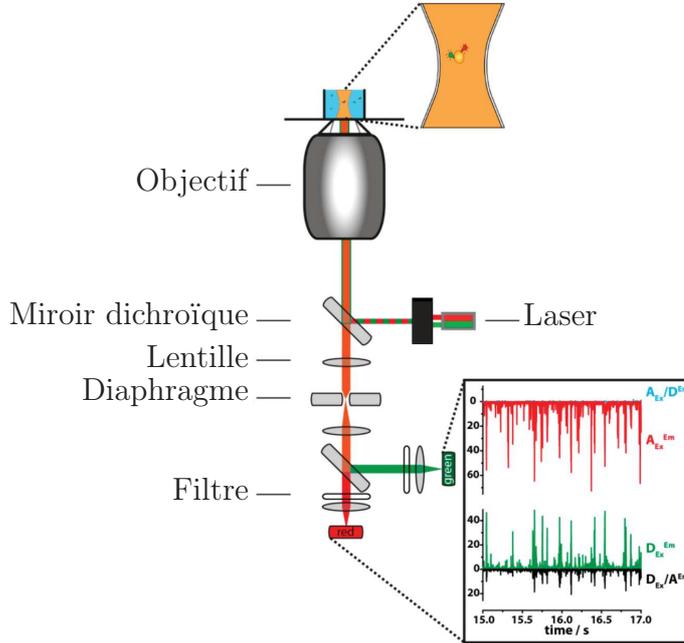


FIGURE 4.7 – Exemple de schéma d’un spectromètre de détection de molécules individuelles à géométrie de détection confocale à deux couleurs.

Le faisceau laser (vert/rouge) est concentré sur l’échantillon. Le miroir dichroïque sépare la lumière d’excitation et d’émission (orange). La lentille concentre la lumière et le diaphragme limite la détection à la lumière venant du plan confocal. Ensuite, une lentille permet de paralléliser le faisceau, et un miroir dichroïque de séparer la lumière verte et la lumière rouge. Les filtres sélectionnent la longueur d’onde désirée, et une dernière lentille permet de focaliser la lumière reçue sur les détecteurs. Adapté de [Gust et al., 2014]

3.4 Analyse des données

La première étape de l’analyse des données est de repérer les émissions de fluorescence enregistrées. Lors de son passage dans le volume confocal (quelques milli-secondes), une molécule aura le temps d’être excitée et d’émettre plusieurs photons. Les “groupes” de photons ainsi enregistrés (Figure 4.7), appelés *bursts*, sont utilisés pour distinguer le signal du bruit. Cette partie de l’analyse est réalisée grâce à un algorithme de recherche de *bursts* dont un des paramètres est un nombre minimal de photon par *burst* [Eggeling et al., 2001].

Lors d’expériences sur des molécules individuelles, il est possible de distinguer les inhomogénéités de marquage grâce au rapport stoechiométrique, S , donné par la formule suivante :

$$S = \frac{I_{D_{Ex}}^{D_{Em}} + I_{D_{Ex}}^{A_{Em}}}{I_{D_{Ex}}^{D_{Em}} + I_{D_{Ex}}^{A_{Em}} + I_{A_{Ex}}^{A_{Em}}} \quad (4.13)$$

Les molécules dont la stoechiométrie est inférieure à 0,2 ou supérieure à 0,8 correspondent au pic de l'accepteur seul ou du donneur seul respectivement. Cette propriété peut être utilisée pour filtrer les données utilisées si besoin est (Figure 4.8A).

En théorie, l'efficacité de FRET peut être calculée à partir de la fluorescence du donneur et de l'accepteur comme décrit par l'équation 4.8. L'intensité de fluorescence mesurée pour chaque fluorophore doit cependant être corrigée pour deux phénomènes principaux :

- Le *cross-talk*, c'est à dire la fuite de l'émission du donneur dans le canal de l'accepteur (LK) ;
- L'excitation directe de l'accepteur lors de l'excitation du donneur (DIR).

On utilise donc les corrections suivantes :

$$F^D(A) = I_{D_{Ex}}^{AEm} - LK - DIR \quad ; \quad F^D(D) = I_{D_{Ex}}^{DEm} + LK$$

$$\text{et } \gamma = \frac{\phi_A \cdot \eta_{A_{Ex}}^{AEm}}{\phi_D \cdot \eta_{D_{Ex}}^{DEm}} \quad \text{d'où} \quad E = \frac{I_{D_{Ex}}^{*AEm}}{I_{D_{Ex}}^{*AEm} + \gamma \cdot I_{D_{Ex}}^{DEm}} \quad (4.14)$$

γ est le facteur de correction, calculé à partir de ϕ le rendement quantique et η l'efficacité de détection avec le microscope utilisé pour la mesure. Les efficacités de FRET obtenues pour chacune des molécules mesurées sont alors présentées sous forme d'un histogramme (Figure 4.8B), et la distribution des efficacités de FRET approximée par une gaussienne. C'est le centre de cette distribution qui est utilisé pour calculer les distances selon la formule 4.10. D'autre part, il est possible d'extraire la durée de vie de l'état excité du donneur ($\tau_{D(A)}$) pour chaque *burst*. L'intensité détectée en fonction du temps est ajustée à une fonction de décroissance exponentielle convoluée avec la fonction qui décrit la vitesse de réponse du spectromètre (Figure 4.8C). Lorsqu'on représente l'efficacité de FRET en fonction de la durée de vie de l'état excité du donneur, on ajoute une dimension d'analyse et donc on améliore la résolution. Il est alors plus facile de distinguer les populations présentes. De plus, ce type d'histogramme permet de faciliter l'étude de la dynamique du système observé (Figure 4.8D). En effet, dans le cas de plusieurs populations présentes, si la molécule est statique à l'échelle de temps de la mesure FRET (ns), tous les pics se trouveront sur une seule droite, appelée la ligne statique de FRET (Equation 4.9 avec $\tau_{D(0)}$ constant, extrait à partir de l'efficacité du pic du donneur seul ($E=0$)). Dans le cas d'une molécule "non-statique", un traitement particulier pour ce qui est du calcul de distance à partir des efficacités est requis.

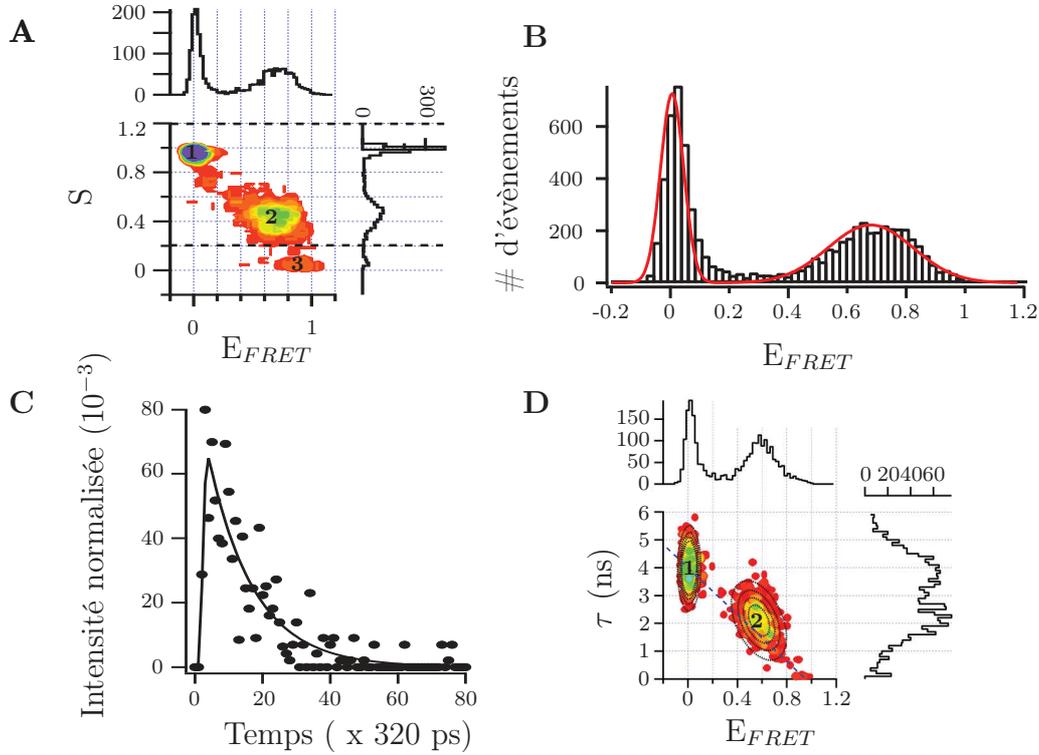


FIGURE 4.8 – Représentation des données de smFRET.

A. Stoechiométrie de marquage en fonction de l'efficacité de FRET. Le pic 1 correspond au pic du donneur seul : $I_{D_{Ex}}^{D_{Em}} > 0$, $I_{D_{Ex}}^{A_{Em}} = 0$ et $I_{A_{Ex}}^{A_{Em}} = 0$. Le pic 2 correspond au pic de FRET : $I_{D_{Ex}}^{A_{Em}} > 0$, $I_{D_{Ex}}^{D_{Em}} > 0$ et $I_{A_{Ex}}^{A_{Em}} > 0$. Enfin, le pic 3 correspond au pic de l'accepteur seul : $I_{D_{Ex}}^{A_{Em}} \sim 0$, $I_{D_{Ex}}^{D_{Em}} \sim 0$ et $I_{A_{Ex}}^{A_{Em}} > 0$. En pointillé, les limites de sélection des données choisies, $[0,2 : 1,2]$. **B.** Histogramme de l'efficacité de FRET. En rouge, l'ajustement par deux gaussiennes. **C.** Détermination de $\tau_{D(A)}$ par ajustement de l'intensité d'un *burst* en fonction du temps. **D.** Histogramme de l'efficacité de FRET en fonction de la durée de vie de l'état excité du donneur (τ), ajusté par 2 gaussiennes (pointillé noir). 1. Pic du donneur seul, $\tau_{D(0)}$. 2. Pic donneur-accepteur, $\tau_{D(A)}$. En bleu pointillé, la ligne de FRET statique.

3.5 Conclusion

Le FRET est une technique qui ne demande pas de grandes quantités d'échantillon et n'est limitée ni par la taille du système ni par sa dynamique. Il s'effectue en solution mais requiert un marquage par deux fluorophores ou plus, ce qui peut amener à introduire un grand nombre de mutations dans le système étudié (suppression des cystéines existantes et introduction de celles servant au marquage). Cependant, cette technique permet d'obtenir de nombreuses informations à propos de la dynamique du système, de sa structure ainsi que de ses interactions.

4 Résonance des plasmons de surface

4.1 Principe

Un plasmon de surface est une onde électromagnétique à décroissance exponentielle qui se crée à l'interface d'un métal (matériau riche en électrons) et d'un milieu diélectrique sans pertes (milieu aqueux par exemple), des deux côtés de l'interface, et qui se propage parallèlement à cette dernière. Il est possible d'exciter ce plasmon en appliquant un laser polarisé parallèlement à l'interface à travers un prisme de verre, de façon à ce que la lumière soit totalement réfléchi à l'interface prisme/métal (condition de réflexion totale, figure 4.9A.). A un angle donné, le laser est capable d'exciter les plasmons de surface, ce qui cause une légère perte d'intensité dans le faisceau réfléchi (Figure 4.9B.). Comme les plasmons sont situés à l'interface des deux milieux, ils sont extrêmement sensibles à toutes les variations de cette surface, comme par exemple l'adsorption de macromolécules biologiques. Ce changement est ensuite reflété dans l'angle auquel on observe une perte d'intensité du faisceau réfléchi. Il suffit donc d'appliquer le laser sur une gamme d'angles incidents pour repérer ces changements par rapport à une référence, et ainsi en déduire les changements ayant lieu à la surface.

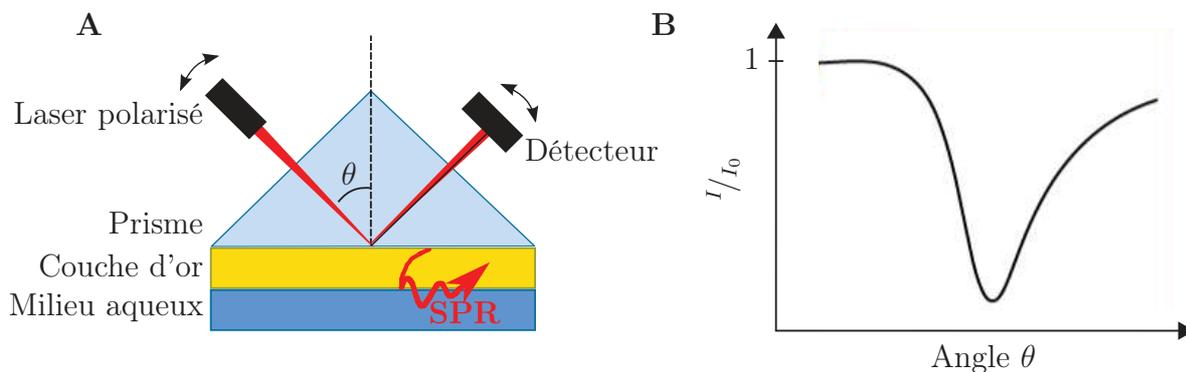


FIGURE 4.9 – Principe de la résonance des plasmons de surface.

A. Représentation schématique du montage expérimental. Le faisceau laser (rouge) présente une baisse d'intensité après réflexion (noir). Le laser et le détecteur bougent simultanément de façon à balayer une gamme d'angles θ . **B.** Rapport d'intensité (intensité réfléchie / intensité appliquée) du laser après réflexion en fonction de l'angle θ . Le minimum de la courbe est l'angle auquel les plasmons entrent en résonance.

4.2 SPR en biologie

La SPR permet de détecter en temps réel l'adsorption de macromolécules biologiques sur une surface. Ainsi, si on immobilise une molécule sur la surface (le ligand),

il devient possible d'étudier la cinétique de ses interactions avec un partenaire (l'analyte). L'expérience consiste alors en trois phases :

- La phase d'association, durant laquelle l'analyte est injecté à flux constant sur la surface, permet de déterminer la constante de vitesse d'association k_a
- La phase de dissociation, durant laquelle seulement du tampon est injecté, permet de déterminer la constante de vitesse de dissociation k_d
- La phase de régénération, durant laquelle tout l'analyte restant est décroché (glycine pH 2, NaOH 10 mM) pour permettre à un nouveau cycle d'avoir lieu.

Le sensorgramme obtenu est ensuite traité par le logiciel du fabricant (*BiaEval*), utilisant les équations classiques des cinétiques d'association et de dissociation (Figure 4.10A). Il est donc possible de déterminer K_D , la constante apparente de dissociation à l'équilibre, selon la formule $K_D = k_d/k_a$.

Généralement, il est nécessaire d'effectuer une série d'expériences à différentes concentrations d'analyte pour déterminer le K_D . Cependant, un nouveau type d'expériences a été introduit récemment. Il s'agit des cinétiques mono-cycle (*single cycle kinetics*, SCK) [Karlsson et al., 2006]. Dans ce cas, plusieurs phases d'association et de dissociation se suivent, sans étape de régénération entre elles, à concentration croissante en analyte, avec seulement à la fin une longue phase de dissociation (Figure 4.10B). C'est de cette façon qu'ont été réalisées les expériences décrites dans ce manuscrit.

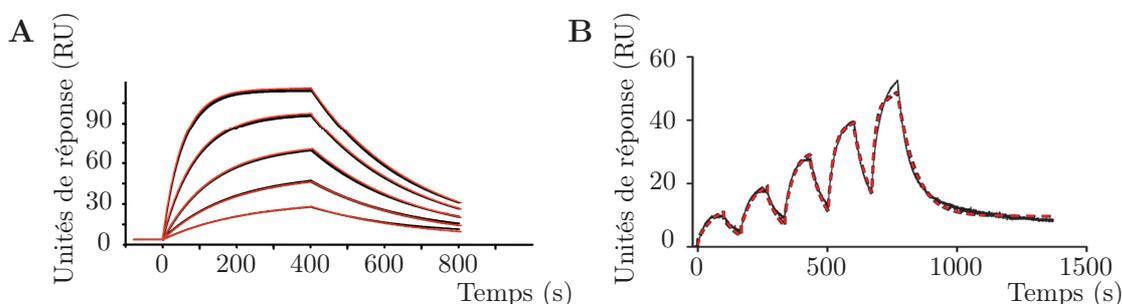


FIGURE 4.10 – Exemples de sensorgramme (noir) et de leur ajustement (rouge).

A. Cinétique multi-cycle classique. L'étape de régénération entre les cycles n'est pas montrée. Adapté de [SPRPages, 2006]. **B.** Cinétique mono-cycle. Adapté de [Adamson and Watts, 2014].

4.3 Conclusion

La résonance des plasmons de surface est une technique qui ne nécessite pas de marquage et qui permet d'étudier la cinétique de l'interaction entre macromolécules. Cependant, c'est une technique qui requiert l'immobilisation d'un des partenaires,

ce qui peut potentiellement affecter sa capacité à se lier. De plus, pour cette même raison, les affinités observées en SPR sont souvent supérieures à celles obtenues par d'autres techniques. Ces limitations doivent être prises en compte lors de l'analyse des résultats.

5 Thermophorèse à échelle microscopique

5.1 Principe

La thermophorèse à échelle microscopique (*microscale thermophoresis*, MST) est une technique qui n'a commencé que très récemment à être appliquée aux interactions protéiques [Wienken et al., 2010]. Le principe de cette technique est de chauffer l'échantillon, constitué du complexe à étudier dont l'un des partenaires est marqué par une sonde fluorescente, contenu dans un capillaire, grâce à un laser infrarouge (Figure 4.11A). La différence de température induite va de 2 à 6 K à l'équilibre, et un gradient microscopique de température se forme alors. Ce gradient provoque un mouvement des protéines présentes, diminuant leur concentration dans la partie la plus chaude (Figure 4.11B). Cette migration est dépendante, entre autres facteurs, du poids moléculaire. Ainsi, la migration observée sera différente suivant si la protéine fluorescente est en complexe ou non (déplétion de la zone la plus chaude). C'est grâce à ce phénomène qu'il est possible d'étudier les interactions par MST, et de déterminer des constantes d'affinité. En effet, si une telle mesure est effectuée à différents rapports de concentration des partenaires, il est possible d'obtenir une courbe de titration (Figure 4.11C), d'exprimer la fraction liée (FL), et de déterminer le K_d de l'interaction par l'ajustement de la formule suivante :

$$K_d = \frac{[A]_{libre}[B]_{libre}}{[AB]} \quad \text{et} \quad [A] = [A]_{libre} + [AB] \quad \text{et} \quad [B] = [B]_{libre} + [AB]$$

$$\text{d'où} \quad FL = \frac{[AB]}{[B]} = \frac{[A] + [B] + K_d - \sqrt{([A] + [B] + K_d)^2 - 4[AB]}}{2[B]}$$
(4.15)

5.2 Considérations expérimentales

Le marquage des échantillons est réalisé grâce à un kit fourni par le fabricant. Le protocole comporte une étape de marquage et une étape de purification de la protéine marquée. Des mesures d'absorbance à différentes longueurs d'onde permettent ensuite de vérifier que la protéine est bien marquée (rapport $\frac{[fluorophore]}{[proteine]}$). Cependant, le nombre de fluorophores par molécule n'est pas contrôlé. L'influence de ce paramètre sur la mesure est limitée puisqu'on observe une déplétion et donc un pourcentage, cependant, une grande inhomogénéité de marquage serait gênante.

Par ailleurs, l'échantillon doit être monodisperse. En effet, comme la technique est basée sur un mouvement en fonction du poids moléculaire, la présence d'agrégats est très délétère pour la mesure.

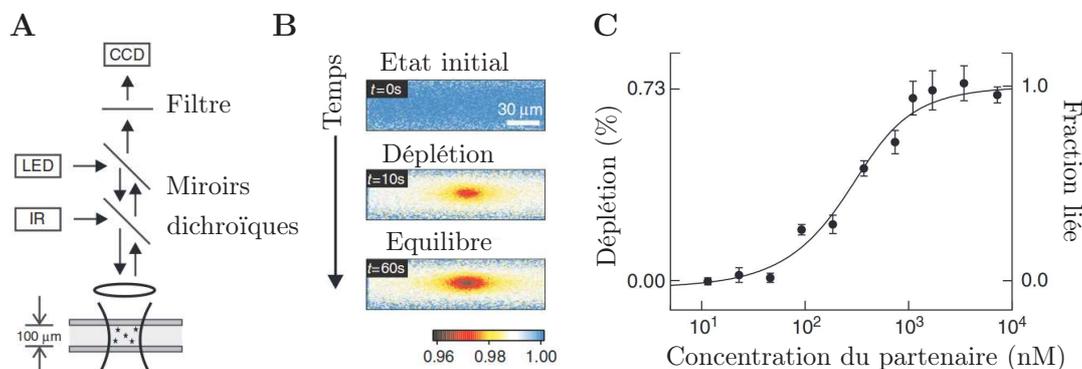


FIGURE 4.11 – Principe du MST et résultats.

A. Montage expérimental composé d'un microscope à fluorescence et d'un laser infrarouge, permettant l'observation et le chauffage simultané du capillaire. **B.** Déplacement des molécules marquées par un fluorophore selon le gradient de température au cours du temps. **C.** Exemple de courbe de titration obtenue par MST. Adapté de [Wienken et al., 2010]

5.3 Conclusion

Le thermophorèse à échelle microscopique est une technique très peu gourmande en matériel biologique, et qui donne des résultats reproductibles. Les expériences sont rapides, ce qui permet d'analyser un grand nombre d'interactions en un temps réduit. C'est cependant une technique qui demande un marquage - sur les amines - et qui peut donc modifier l'interaction dans certains cas.

Matériel et Méthodes

1 Préparation des échantillons

1.1 Expression et purification : 627-NLS et importine α

Le protocole d'expression et de purification était pré-établi au laboratoire et a été le même pour tous les mutants du 627-NLS. Toutes les constructions utilisées ont été sous-clonées dans un dérivé de vecteur pET9a, avec une étiquette de six histidines et un site de clivage par la TEV en N-terminal de la séquence de la protéine (voir annexe A). Le plasmide d'intérêt est transformé dans la souche BL21 d'*Escherichia coli*, puis un clone est cultivé en milieu LB jusqu'à atteindre une densité optique à 600 nm de 0,6. La production de protéine est induite par l'ajout de 1 mM final d'isopropyl β -D-thiogalactopyranoside (IPTG). Les bactéries sont ensuite maintenues à 16°C pendant 16 heures. Après récolte par centrifugation à 6000 g et 4°C, les bactéries sont conservées à -80°C jusqu'à la purification. Pour les échantillons isotopiquement enrichis, une préculture en LB est utilisée pour ensemercer une préculture en milieu M9 (table 5.1), qui est elle même utilisée après lavage (centrifugation, resuspension dans un petit volume de M9, centrifugation) pour ensemercer à une densité optique à 600 nm comprise entre 0,15 et 0,2 la quantité désirée de milieu M9 pour la production de protéine. Le reste du protocole est identique.

Toutes les constructions sont purifiées grâce à une chromatographie d'affinité Nickel (IMAC) suivie d'une chromatographie d'exclusion de taille. Les cellules sont lysées par sonication dans le tampon A (50 mM Tris pH 7,6, 200 mM NaCl) complété par un cocktail d'inhibiteur cOmplete[®] (Roche) durant 7 minutes, par cycles de 5 secondes on/off, à 70 % de puissance. La fraction soluble est séparée des débris cellulaires par centrifugation à 20 000 g pendant 45 minutes et est chargée sur la colonne IMAC équilibrée dans le tampon A. Après un lavage à haute concentration de sel (Tampon A, 500 mM NaCl), la protéine est éluée (Tampon A plus 300 mM

imidazole). L'étiquette histidine est clivée par digestion par la TEV sur la nuit à 4°C, et l'échantillon subit en même temps une dialyse contre 50 mM Tris pH 7, 200 mM NaCl. Après un second passage sur IMAC pour retirer l'étiquette clivée et la fraction non clivée, une chromatographie d'exclusion de taille est ensuite effectuée sur une colonne Superdex 75[®] (GE Healthcare) équilibrée dans le tampon RMN (50 mM Tris pH 6.5, 200 mM NaCl).

Pour les mutants contenant au moins une cystéine, 10 mM de dithiothréitol (DTT) sont ajoutés après la deuxième IMAC. L'étape de marquage est ensuite effectuée directement, le DTT étant retiré du tampon par dialyse juste avant le marquage.

Pour les isoformes de l'importine α , le même protocole a été suivi, à l'exception près que l'élution de l'IMAC a été réalisée à 150 mM imidazole.

Composant	Quantité (g ou mL)	Marquage	Composant	Quantité (g)
KH ₂ PO ₄	3	15N	15NH ₄ Cl	1
Na ₂ HPO ₄	6,75		Glucose	2
NaCl	0,5		H ₂ O	qsp 1 L
MgSO ₄ 1 M	2	15N 13C	15NH ₄ Cl	1
CaCl ₂ 0,1 M	1		U-13C-Glucose	2
ZnSO ₄ 0,05	1		H ₂ O	qsp 1 L
MnSO ₄ 0,1 M	0,1	2H 15N 13C	15NH ₄ Cl	1
Vitamines MEM 100X	10		U-13C-2H-Glucose	2
Antibiotique(s) 1000X	1		D ₂ O	qsp 1 L

TABLE 5.1 – Composition du milieu M9 utilisé en fonction du marquage souhaité. Les deux premières colonnes indiquent les éléments communs tandis que celles de droite indiquent les composants à ajouter en fonction du marquage voulu.

1.2 Marquage par des sondes

Toutes les sondes utilisées dans ce travail, que ce soit pour des expériences de RMN ou de FRET, se lient de façon covalente sur une cystéine de la protéine par un pont disulfure ou une liaison thioéther dans le cas des conjugués maléimides.

Le marquage par MTSL

Le marquage par S-(2,2,5,5-tetraméthyle-2,5-dihydro-1H-pyrrol-3-yl) méthyle méthanesulfonothioate (MTSL) s'effectue après le retrait du DTT par dialyse. Après mélange du MTSL et de la protéine à un ratio molaire de 7/1, la réaction est conservée dans le noir, 20 minutes à température ambiante, puis sur la nuit et sous agitation

à 4°C. L'excès de MTSL est retiré par chromatographie d'exclusion de taille dans le tampon RMN.

Le marquage par conjugués maléimide

Dans ce cas, une dialyse permet de retirer le DTT mais aussi de se placer à pH 7. Cette fois, le conjugué est ajouté à la protéine rapidement, dans le noir, et sur vortex, à un ratio molaire de 5/1. La réaction est conservée dans le noir, 20 minutes à température ambiante, puis sur la nuit à 4°C. L'excès de conjugué maléimide est retiré par chromatographie d'exclusion de taille. Pour cette thèse, les conjugués maléimide TEMPO (2,2,6,6-Tetramethylpiperidine 1-oxyl), AlexaFluor[®] 594 et AlexaFluor[®] 448 (Life Technologies) ont été utilisés.

2 Biologie moléculaire

2.1 Clonage du 627-NLS de la grippe B

Pour le clonage du 627-NLS de la grippe B, la technique de PCR chevauchante a été utilisée. Une première PCR a permis d'amplifier l'étiquette poly-histidine et le site de clivage par la TEV à partir du vecteur dérivé de pET9a utilisé pour toutes les constructions de ce projet, tout en ajoutant un site de digestion par *HincII* à l'extrémité 5' du fragment synthétisé. En parallèle, une deuxième PCR a permis d'amplifier la séquence du 627-NLS de la grippe B proprement dite, depuis un plasmide synthétique contenant la séquence de la totalité de la polymérase de la grippe B, tout en ajoutant un site de digestion par *SacI* à l'extrémité 3' du fragment synthétisé. Après purification des produits de PCR, ils ont été mélangés et ont subi une troisième PCR utilisant uniquement l'amorce 5' de la première PCR et l'amorce 3' de la seconde. Comme les amorces des premières PCR ont été imaginées de façon à créer une séquence chevauchante entre les deux premiers produits de PCR, la troisième PCR permet d'obtenir un insert complet, contenant à la fois l'étiquette poly-histidine, le site de clivage par la TEV, et la séquence du domaine 627-NLS de la grippe B, encadrés par les sites de digestion par *HincII* et *SacI*.

Une fois l'insert obtenu, 60 ng d'insert et 2 µg du vecteur dérivé de pET9a ont été digérés par *HincII* et *SacI* 1 heure à 37°C. Après inactivation des enzymes de restriction (20 minutes à 70°C), l'insert et le vecteur ont été purifiés sur gel d'agarose 0,5 %. La ligation a été effectuée en utilisant un kit rapide, avec un total de 100 ng d'ADN et un ratio insert/vecteur de 3/1. L'ADN obtenu a été transformé dans des *Escherichia coli* NEB10β commerciales, puis les clones positifs ont été sélectionnés sur milieu solide grâce à la résistance à la kanamycine contenue dans le plasmide d'intérêt.

2.2 Mutagénèse dirigée

Toutes les mutations ponctuelles réalisées lors de ce projet l’ont été selon la technique de mutagénèse dirigée, en utilisant le kit Agilent Quick-Change. Les amorces utilisées contenaient la mutation désirée en leur centre, et avaient une température de fusion d’environ 80°C. Le cycle PCR utilisé est décrit dans la table 5.2.

Etape	Température (°C)	Durée	Répétition
1	90	5 min	
2	90	30 s	
3	55	30 s	
4	68	7 min	retour à l’étape 2, 16 fois
5	68	10 min	

TABLE 5.2 – Cycle PCR utilisé pour la mutagénèse dirigée.

3 Mesures par résonance magnétique nucléaire

Toutes les mesures RMN ont été effectuées dans le tampon RMN décrit précédemment, additionné de 10% D₂O. Tous les spectres obtenus ont été transformés grâce à NMRPipe [Delaglio et al., 1995] et ont été analysés avec le logiciel Sparky [Goddard and Kneller, 2008].

3.1 Expériences d’attribution

Les attributions du 627-NLS aviaire triplement marqué (²H, ¹⁵N, ¹³C) à une concentration de 0,5 mM, du 627-NLS aviaire R650A doublement marqué (¹⁵N, ¹³C) à une concentration de 2 mM et de l’importine α 1 doublement marquée (¹⁵N, ¹³C) à une concentration de 1,5 mM ont été obtenues à 10°C. Ont été enregistrées des expériences 3D de type BEST-TROSY [Solyom et al., 2013], sur un spectromètre Bruker équipé d’une cryo-sonde dont la fréquence ¹H était de 950 MHz pour le 627-NLS aviaire et le 627-NLS R650A ou 850 MHz pour l’importine α 1. L’attribution du domaine NLS aviaire doublement marqué (¹⁵N, ¹³C) à une concentration de 1 mM a été obtenue à 25°C. Ont été enregistrées des expériences 3D de type BEST [Lescop et al., 2007], sur un spectromètre Varian équipé d’une cryo-sonde dont la fréquence ¹H était de 600 MHz. L’attribution du domaine 627 aviaire triplement marqué (²H, ¹⁵N, ¹³C) à une concentration de 0,66 mM a été obtenue à 25°C. Ont été enregistrées des expériences 3D de type BEST [Lescop et al., 2007], sur un spectromètre Varian équipé d’une cryo-sonde dont la fréquence ¹H était de 800 MHz. Le programme MARS [Jung and Zweckstetter, 2004] a fourni une attribution automatique des systèmes de spins qui a ensuite été vérifiée manuellement.

3.2 CEST

Les expériences de transfert de saturation lors de l'échange chimique (*chemical exchange saturation transfer*, CEST) [Vallurupalli et al., 2012] ont été effectuées sur le 627-NLS aviaire (voir publication) et sur le 627-NLS humain à 5, 10 et 15°C en utilisant un champ B_1 de 22 Hz, un pas de 25 Hz, et une période de saturation de 300 ms. Les expériences de CEST sur le 627-NLS aviaire R650A ont été effectuées à 700 MHz à 10 °C en utilisant un champ B_1 de 22 Hz ou 44 Hz, un pas de 25 ou 40 Hz, et une période de saturation de 400 ou 300 ms respectivement. Les données ont été analysées en utilisant ChemEx [Vallurupalli et al., 2012] en supposant un modèle d'échange entre deux états.

3.3 PREs

Les expériences d'augmentation paramagnétique de la relaxation (*paramagnetic relaxation enhancement*, PRE) ont été mesurées pour différents mutants, marqués par différentes sondes, et dans diverses conditions (5.3). Une fois le spectre HSQC paramagnétique acquis, la sonde est réduite par l'acide ascorbique pour l'acquisition du spectre HSQC de référence. Les intensités extraites sont ensuite normalisées par le nombre de scan de chaque expérience.

Protéine	Sonde	Position marquée	Champ (MHz)	Température (°C)
627-NLS aviaire R650A	TEMPO	587	850	10
627-NLS aviaire R650A	TEMPO	675	850	10
627-NLS aviaire R650A	TEMPO	675	850	25

TABLE 5.3 – Paramètres des expériences d'augmentation paramagnétique de la relaxation.

3.4 Relaxation

Des expériences de relaxation longitudinale (R_1 ou $T_{1\rho}$) [Lakomek et al., 2012, Akke and Palmer, 1996], transversale (R_2) [Kay et al., 1989], et cross-corrélées [Wang et al., 2003] ont été enregistrées pour différentes protéines et dans différentes conditions résumées dans la table 5.4.

Protéine	Expérience	Champ (MHz)	Température (°C)
627-NLS aviaire	R_1	600	10
627-NLS aviaire	R_2	600	10
627-NLS aviaire R650A	η_{xy}	850	10
Importine $\alpha 1$	$T_{1\rho}$	600	10

TABLE 5.4 – Paramètres des expériences de relaxation.

4 Diffusion aux petits angles

4.1 Des rayons X

Les expériences de diffusion des rayons X aux petits angles (SAXS) du 627-NLS aviaire à 5, 10, 15, 20, 25 et 30°C ont été enregistrées sur la ligne de lumière BM29 de l'ESRF de Grenoble. L'énergie utilisée était de 12,5 keV pour une longueur d'onde de 0,9918 Å. 50 µL d'échantillon à trois concentrations différentes (4,9, 2,45 et 1,23 mg/mL) et le tampon correspondant ont été exposés aux rayons X (robot passeur d'échantillons). Dix images correspondant chacune à une durée d'exposition d'une seconde ont été collectées pour chaque échantillon par un détecteur Pilatus 1M (Dectris) placé à 2,864 mètres de l'échantillon. Chaque image a ensuite été traitée par le logiciel EDNA, qui utilise les outils d'ATSAS (EMBL-Hambourg), résultant en courbes individuelles radialement moyennées d'intensité normalisée en fonction de l'angle de diffusion $q = 4\pi \sin(\theta)/\lambda$ [Incardona et al., 2009, Petoukhov et al., 2012]. Les images de chaque échantillon sont ensuite combinées, en excluant celles affectées par l'aggrégation induite par les radiations, résultant en la courbe moyenne de diffusion, le tampon étant mesuré avant et après l'échantillon et moyenné pour substraction. La suite de logiciel ATSAS a ensuite été utilisée pour traiter les données.

4.2 Des neutrons

Les expériences de diffusion des neutrons aux petits angles (SANS) du complexe importine $\alpha 1$ /627-NLS aviaire ont été enregistrées à 8°C sur la ligne D33 de l'institut Laue-Langevin (ILL) à Grenoble. Le faisceau de neutron ($\lambda = 6 \pm 10$ %) passait à travers un collimateur rectangulaire juste avant d'atteindre l'échantillon, situés respectivement à 2 ou 14,4 mètres et à 2 ou 14 mètres du détecteur. La gamme d'angles q couverts dans cette expérience allait de 0,004 nm⁻¹ à 0,048 nm⁻¹. La diffusion du complexe à 2,45 mg/mL dans un tampon 100 % D₂O et du tampon correspondant a été mesurée dans une cuvette ronde de 2 mm d'épaisseur.

5 Transfert d'énergie par résonance de Förster

Les expériences de *single molecule Förster resonance energy transfer* (smFRET) ont été effectuées sur des couples de mutants cystéines du 627-NLS aviaire et/ou du 627-NLS aviaire R650A : I539C/A707C, I539C/A717C, L675C/A707C et L675C/A717C. Ce mutant a été marqué de façon aléatoire par les fluorophores Alexa488 et Alexa594 en ratio 1/1 (voir paragraphe 1.2). Juste avant la mesure de smFRET, le tampon a été échangé contre le tampon RMN supplémenté de 2 mM DTT. Les mesures de smFRET ont été effectuées en collaboration avec le laboratoire d'Edward Lemke à l'EMBL-Heidelberg.

Les *bursts* ont été extraits des photons enregistrés grâce à un algorithme de recherche de *burst* [Eggeling et al., 2001]. Ceux dont l'intensité cumulée du donneur et de l'accepteur lors d'une excitation par le laser vert dépassait 50 photons ont été sélectionnés pour l'analyse. Les efficacités de FRET ont ensuite été calculées à partir des fluorescences du donneur et de l'accepteur corrigées selon les équations 4.8 et 4.14, avec le rendement quantique $\phi = 1,14$ qui a été mesuré séparément selon les protocoles décrits dans la littérature et indépendamment validé grâce aux paramètres extraits de la conformation fermée, statique, en combinant les équations 4.9 et 4.14 [Milles et al., 2012]. La stoechiométrie de marquage S a été calculée selon l'équation 4.13, et les *bursts* pour lesquels $S = [0, 2 : 1, 2]$ ont été sélectionnés pour les analyses [Sisamakris et al., 2010]. Les histogrammes représentant E et τ ont ensuite été ajustés par une double ou triple gaussienne suivant le nombre de populations présentes (voir le paragraphe VI.2 pour plus de détail). Enfin, la distance entre les fluorophores a été déterminée selon l'équation 4.10, en prenant en compte le rendement quantique ainsi que l'intégrale de recouvrement des spectres d'émission du donneur et d'excitation de l'accepteur mesurée sur le 627-NLS simplement marqué [Sisamakris et al., 2010].

6 Création et sélection d'ensembles

Les ensembles de structures représentant les formes ouvertes et fermées du 627-NLS aviaire ont été générés en utilisant *flexible-meccano* [Ozenne et al., 2012, Bernadó et al., 2005]. Pour l'ensemble représentant la forme fermée, la structure cristallographique du 627-NLS aviaire a été utilisée, et *flexible-meccano* a été utilisé pour ajouter le peptide NLS flexible en assumant une distribution *statistical coil* pour ces acides aminés (résidus 742 à 759). La forme ouverte a été générée de la même façon, mais le *linker* entre les domaines 627 et NLS a également été construit par *flexible-meccano* (acides aminés 677 à 689), pour obtenir un ensemble dans lequel les deux domaines sont séparés et peuvent s'orienter librement dans la limite de l'espace conformationnel autorisé par le *linker*.

Dans le cas du complexe avec l'importine α , un ensemble de structure a été généré à partir de la structure cristallographique du NLS humain en complexe avec l'importine $\alpha 1$ (PDB 4UAF). *Flexible-meccano* a été utilisé pour construire le *linker* entre le NLS et le 627, résultant en un ensemble de structures du 627-NLS en complexe avec l'importine $\alpha 1$ dans lequel le 627 échantillonne un grand nombre de conformations. Comparé à la structure cristallographique de l'importine α , la construction expérimentale comprend 13 acides aminés supplémentaires en N-terminal et 34 acides aminés supplémentaires en C-terminal qui sont prédits désordonnés et ont été construits grâce à *flexible-meccano* (résidus 60 à 72 et 496 à 529 respectivement).

Les courbes de SAXS des ensembles obtenus sont ensuite prédites structure par structure et moyennées pour permettre une analyse comparative avec les données expérimentales, ou une sélection grâce au logiciel ASTEROIDS [Jensen et al., 2014].

7 Détermination de constantes d'affinité

7.1 Thermophorèse à échelle microscopique

L'importine α a été marquée par le fluorophore commercial *red* de NanoTemper Technologies, qui se couple sur les amines, selon le protocole décrit dans le kit. Les expériences de MST ont été enregistrées à 25°C, une puissance MST de 40 % et une puissance de LED à 90 %, dans le tampon SPR (10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0,05 % Tween20). 16 échantillons ont été préparés par dilution en cascade du 627-NLS, partant de 2,4 μ M, puis par mélange 1/1 avec l'importine α marquée à 13,5 nM. Après centrifugation, chaque échantillon a été aspiré dans un capillaire de type "standard", puis la mesure a été effectuée immédiatement. Un triplicata a été réalisé pour chaque mutant de 627-NLS testé.

7.2 Résonance des plasmons de surface

Les expériences de résonance des plasmons de surface (SPR) ont été effectuées sur un Biacore T200, à 5, 10, 15, 20, 25, 30, 35, 40 et 45°C. Le 627-NLS aviaire ou humain a été immobilisé sur une puce CM5 série S par un couplage via les amines. La surface a été activée par un mélange 1/1 de N-hydroxysuccinimide (NHS) à 0,1 M et d'hydrochlorure de 1-éthyl-3-(3-diméthylamino-propyl)carbodiimide (EDC) à 0,4 M. Le 627-NLS aviaire ou humain a ensuite été dilué à 5 μ g/mL dans 10 mM d'acétate de sodium pH 4.5 puis injecté jusqu'à ce qu'une réponse de 150 unités de réponse (RU) soit atteinte. Les esters activés restants ont été saturés par l'injection d'une solution d'hydrochlorure d'éthanolamine 1 M pH 8,5. Une des cellules de la puce a été utilisée comme référence, pour surveiller la réponse provoquée par le tampon et les interactions non-spécifiques. Elle a subi le même protocole d'immobilisation, excepté l'étape d'injection de protéine où elle n'a reçu que du tampon. Les expériences d'interactions ont été effectuées en mode "single cycle kinetics" (SCK), avec les isoformes $\alpha 1$, $\alpha 3$, $\alpha 5$ et $\alpha 7$ de l'importine (3 - 250 nM). Les phases d'association pour chaque concentration ont été de 120 secondes à un flux de 30 μ L/min avec une phase d'équilibration de 15 secondes entre chaque concentration, et la phase de dissociation finale de 300 secondes. L'interaction non-spécifique avec la matrice a été corrigée en soustrayant au signal obtenu le signal de la piste de référence. La régénération entre chaque expérience a été effectuée par le passage de 10 mM de glycine pH 2, 20 secondes à 30 μ L/min. Les paramètres cinétiques ont été déterminés par le logiciel BIAevaluation en utilisant un modèle d'interaction 1/1.

Chapitre VI

Structure et dynamique du domaine 627-NLS aviaire et de son interaction avec l'importine α

1 Résumé de la publication

L'objectif de cette étude était de déterminer les bases moléculaires de l'interaction du 627-NLS aviaire avec l'importine $\alpha 1$ en solution. Pour cela, nous avons utilisé la RMN, le SAXS, le SANS et le smFRET. Nous avons révélé que deux conformations en équilibre dynamique coexistent en solution. Nous avons déterminé les caractéristiques cinétiques et structurales de cet équilibre et avons construit une description structurale des conformations présentes. Enfin, nous avons également décrit la structure du complexe, mettant en évidence un phénomène de sélection conformationnelle qui pourrait jouer un rôle dans l'adaptation inter-espèces du virus de la grippe.

1.1 Equilibre entre conformations ouvertes et fermées

Comme nous l'avons vu en introduction, les structures cristallographiques des différentes souches de grippe sont toutes similaires (Figure 1A). Cependant, ces structures ne sont pas compatibles avec celle du complexe entre le domaine NLS et l'importine α , car lorsqu'on les superpose, un important problème d'encombrement stérique est mis en évidence (Figure 1C). Pour comprendre la façon dont cette interaction est possible *in vivo*, nous avons entrepris l'étude du 627-NLS aviaire en solution.

Sur le spectre de corrélation TROSY ^1H - ^{15}N enregistré, au lieu des 213 pics attendus (223 acides aminés - 10 prolines), j'ai observé environ 350 pics correspondant

à des amides du squelette du 627-NLS (Figure 2A). Cette observation correspond à la présence d'un échange lent entre deux formes, chacune d'entre elles donnant naissance à un ensemble de signaux sur le spectre. Grâce à des expériences RMN tridimensionnelles de type TROSY, j'ai pu attribuer les deux formes présentes à 10°C (Table S1). Les signaux d'une des deux formes se superposent avec les signaux des deux domaines séparés. Dans ce cas, les deux domaines sont donc suffisamment éloignés pour ne pas interagir mutuellement sur leurs déplacements chimiques : c'est une forme ouverte. L'autre forme, avec des déplacements chimiques différents, correspond à une forme où les deux domaines sont proches : c'est une forme fermée (Figure 2B,C). Pour vérifier la distinction faite lors de l'attribution entre les deux formes et attribuer certains signaux supplémentaires, j'ai également attribué les deux domaines, 627 et NLS, séparément (Figure S1A, S1B, Tables S2 et S3).

Les déplacements chimiques secondaires des carbones (C' , C_α et C_β) calculés à partir de l'attribution ont montré que les deux formes possèdent les mêmes structures secondaires, et que ces dernières sont comparables aux structures secondaires observées dans les structures cristallographiques, sauf pour le *linker* inter-domaines (Figure S1C, S1D). Ceci indique que le seul changement structural entre les deux formes est localisé au niveau du *linker*. Les déplacements chimiques secondaires de l'azote sont quant à eux plus sensibles à l'environnement chimique local, et sont significativement différents entre les deux formes à l'interface cristallographique entre les deux domaines, suggérant que la forme fermée observée en solution correspond à la forme cristallographique (Figure 3).

Le 627-NLS aviaire en solution est en échange lent entre deux conformations : une conformation fermée, qui correspond à la structure cristallographique, et une conformation ouverte et dynamique, pour laquelle le *linker* est flexible, et qui explore le paysage conformationnel possible en échange rapide car elle ne donne lieu qu'à un seul jeu de signaux. J'ai réalisé les mêmes expériences sur le 627-NLS humain qui se comporte de la même façon (Figure S1E).

1.2 La forme fermée est stabilisée par un pont salin

Il est possible d'identifier un pont salin entre les deux domaines, conservé dans toutes les structures cristallographiques disponibles, impliquant trois acides aminés : R650 (domaine 627), D730 (domaine NLS) et E687 (*linker*) (Figure 4A). Le *linker* participe à la structuration de la forme fermée, il est donc stabilisé et par conséquent visible dans les structures cristallographiques. J'ai muté le résidu 650 en alanine, ce qui a permis de rompre cette interaction et d'obtenir un 627-NLS uniquement ouvert. En effet, la forme fermée n'est plus visible en RMN (Figure 4B). Des expériences de relaxation montrent que le *linker* est très dynamique, et les deux

domaines 627 et NLS ont un temps de corrélation τ_c indépendant et corrélé à leur taille (Figure S3). La relaxation du 627-NLS aviaire non muté sera présentée dans les résultats complémentaires.

1.3 L'équilibre dépend de la température

J'ai mesuré une titration en température du 627-NLS aviaire à 5, 10, 15, 20, 25, 30 et 35°C. Lorsque la température augmente, la population de la forme ouverte augmente et celle de la forme fermée diminue (Figure 5A). Pour estimer les populations des deux formes en fonction de la température, j'ai effectué une analyse de la forme de raie d'un groupe de pics représentatifs et bien résolus (Figure 5C, points noirs). A 10°C, le 627-NLS est majoritairement en conformation fermée, alors qu'à 30°C on trouve environ 50 % de chaque forme.

Une analyse plus poussée a été entreprise grâce à des expériences de CEST, qui ont permis de déterminer les populations et constantes d'échanges à chaque température (Figure 5B,C,D, Figure S4). De plus, une analyse simultanée des pics utilisés admettant une relation d'Eyring a permis d'extraire les paramètres thermodynamiques de l'équilibre ($\Delta G = 65$ kJ/mol à 10°C) (Figure 5E). Les contributions au changement de conformation sont majoritairement enthalpiques à basse température, la part d'entropie augmentant avec la température (Table 1).

Ces expériences ont également été effectuées sur le 627-NLS humain et seront présentées dans les résultats complémentaires.

1.4 Confirmation par SAXS

J'ai enregistré les données de SAXS du 627-NLS aviaire à 5, 10, 15, 20, 25, 30 et 35°C (Figure 6A). Les R_g extraits augmentent selon la température, ce qui est en accord avec l'augmentation de la population de conformation ouverte selon la température. Nous avons ensuite utilisé *flexible-meccano* (modèle de *coil* statistique développé au laboratoire) pour créer un ensemble de structures représentant chaque forme. L'ensemble de la forme fermée a été calculé à partir de la structure cristallographique, en ajoutant le peptide NLS C-terminal flexible manquant. L'ensemble de la forme ouverte a été calculé en rendant en plus le *linker* entre les domaines 627 et NLS flexible, et en supposant que tout l'espace conformationnel est exploré. Ensuite, CRYSOLE a été utilisé pour prédire les courbes SAXS de ces deux ensembles. En combinant les deux courbes prédites en proportions ouvertes et fermées telles que déterminées par le CEST, il est possible de reproduire les courbes expérimentales de SAXS à 5 et 30°C (Figure 6B,C).

1.5 La conformation ouverte interagit avec l'importine α

Nous avons utilisé deux méthodes pour montrer que la conformation ouverte du 627-NLS interagit avec l'importine α .

Premièrement, nous avons utilisé un ensemble, construit sur la base de la structure cristallographique du complexe entre l'importine $\alpha 1$ et le domaine NLS, dans lequel le domaine 627 échantillonne tout l'espace conformationnel, en supposant que le NLS ne bouge pas (Figure 7A). Nous avons ensuite prédit la courbe SAXS de chacun des 25000 conformères prédits. La meilleure concordance avec les données expérimentales n'est pas obtenue pour la courbe SAXS moyenne des 25000 conformères, mais pour celle de certaines structures en particulier. L'échantillonnage réel de l'espace conformationnel par le domaine 627 n'est donc pas uniforme. Les meilleurs accords entre données prédites et expérimentales ont été obtenus pour des structures dans lesquelles le domaine 627 est proche du C-terminal de l'importine α (Figure 7B,D). De la même façon, si une sélection est effectuée contre les données SAXS en laissant le *linker* flexible, l'ensemble de solutions retenues contient les mêmes structures (Figure 8). Ces structures permettent une excellente reproduction des données SANS du complexe hydrogéné dans un tampon 100 % D₂O, alors que l'ensemble original de 25000 conformères ne le permet pas (Figure S5). De plus, dans ces conditions, le SANS nous informe sur la forme de la protéine seule, contrairement au SAXS pour lequel il faut prendre en compte une couche d'hydratation.

Deuxièmement, nous avons construit un modèle *ab initio* en utilisant DAMMIF pour construire une enveloppe électronique correspondant au complexe [Franke and Svergun, 2009]. L'accord entre cette enveloppe et les solutions dérivées de l'ensemble de 25000 conformères est bon, même si il reste une petite partie de densité électronique vide (Figure 7C). Ceci peut être dû à la flexibilité soit du 627 (confirmée par smFRET), soit de l'importine α .

1.6 Propriétés dynamiques et mode d'interaction du 627-NLS avec l'importine α par smFRET

J'ai marqué aléatoirement un double mutant cystéine (I539C, A707C) du 627-NLS aviaire par les fluorophores Alexa488 et Alexa594 (Figure 9A). L'analyse quantitative de la largeur de la population FRET avec la durée de vie de l'état excité τ montre l'existence de plus d'une conformation (Figure 6B). Cette observation est en accord avec le k_{ex} obtenu par RMN, plus lent que le temps d'observation de chaque molécule (~ 5 ms). Il est donc probable que la population FRET du 627-NLS libre soit une combinaison de deux états. C'est pourquoi les données ont été analysées avec une gaussienne triple, donnant $E_{FRET} = 0.62$ et 0.77 (Figure 6B). L'efficacité FRET de 0.77 correspond à une distance entre fluorophores de 47 \AA , ce qui correspond bien à la forme cristallographique du 627-NLS (la distance C $_{\alpha}$ -C $_{\alpha}$ entre

les acides aminés 539 et 707 est de 44 Å). Cette population se comporte de façon statique car elle est située sur la ligne statique de FRET. L'autre conformation, pour laquelle $E_{FRET} = 0.62$, présente une efficacité de FRET plus basse que la forme fermée, et donc une distance entre fluorophores plus grande. C'est donc la forme ouverte. Cette population dévie de la ligne statique de FRET, indiquant une interconversion rapide ($< \mu\text{s}$) entre plusieurs conformations. A 23°C, le nombre de molécules dans chaque état est similaire et proche des populations extraites par RMN (rapport ouvert/fermé de 45/55 par FRET contre 40/60 par RMN).

Lors de l'addition de l'importine α , le pic de FRET se déplace vers une efficacité plus faible, c'est à dire une plus grande distance (Figure 9C,D, S6A). On observe une augmentation de la population ouverte et la disparition progressive de la forme fermée lors de l'addition d'importine α , ce qui est en accord avec l'hypothèse selon laquelle seule la forme ouverte est capable d'interagir avec l'importine α . Cependant, une fois liée à l'importine α , le pic FRET du 627-NLS n'est toujours pas situé sur la ligne statique de FRET, ce qui indique qu'il reste flexible une fois en complexe.

Enfin, le mutant R650A a également été analysé par smFRET et on observe que la conformation fermée disparaît presque totalement, au profit d'une grande majorité d'une seule population, encore plus ouverte (E_{FRET} plus faible) que la forme ouverte du 627-NLS non muté (Figure S6B).

1.7 Conclusion

Nous avons utilisé une combinaison d'approches en solution pour caractériser le comportement conformationnel du domaine 627-NLS. Nous avons montré que deux conformations, ouverte et fermée, existent en solution, en équilibre dynamique. La barrière énergétique entre ces deux états correspond à la rupture d'un pont salin très conservé à l'interface entre les deux domaines. La forme fermée correspond à la structure cristallographique, alors que la forme ouverte est hautement flexible et dynamique. L'équilibre lent entre ces deux formes est contrôlé par la température, la vitesse de l'échange et la population ouverte augmentant avec cette dernière. La forme ouverte est la seule permettant l'interaction avec le partenaire physiologique du 627-NLS, l'importine α , et le domaine 627 reste dynamique dans le complexe, échantillonnant préférentiellement la proximité du C-terminal de l'importine α .

Avec cette étude, nous avons démontré l'importance fonctionnelle de la dynamique des domaines, médiée par un *linker* intrinsèquement désordonné. Nous avons également montré que l'interaction du 627-NLS avec l'importine α suit probablement un mécanisme de sélection conformationnelle à partir de l'équilibre existant.

Large-Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin α

Elise Delaforge,^{†,‡,§} Sigrid Milles,^{†,‡,§} Guillaume Bouvignies,^{†,‡,§} Denis Bouvier,^{†,‡,§,||,⊥,#} Stephane Boivin,^{||,⊥,#} Nicola Salvi,^{†,‡,§} Damien Maurin,^{†,‡,§} Anne Martel,[¶] Adam Round,[□] Edward A. Lemke,[△] Malene Ringkjøbing Jensen,^{†,‡,§} Darren J. Hart,^{*,||,⊥,#} and Martin Blackledge^{*,†,‡,§}

[†]Univ. Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38044 Grenoble, France

[‡]CEA, DSV, IBS, F-38044 Grenoble, France

[§]CNRS, IBS, F-38044 Grenoble, France

^{||}Univ. Grenoble Alpes, UVHCI, Grenoble, France

[⊥]CNRS, UVHCI, Grenoble, France

[#]Unit for Virus Host Cell Interactions, Univ. Grenoble Alpes-EMBL-CNRS, Grenoble, France

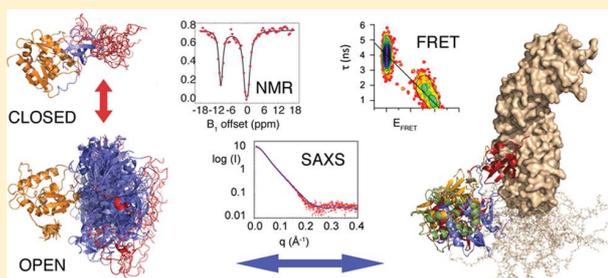
[¶]Institut Laue-Langevin, F-38044 Grenoble, France

[□]European Molecular Biology Laboratory, Grenoble Outstation, 38042 Grenoble, France

[△]European Molecular Biology Laboratory, Structural and Computational Biology Unit, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Supporting Information

ABSTRACT: Influenza A RNA polymerase complex is formed from three components, PA, PB1, and PB2. PB2 is independently imported into the nucleus prior to polymerase reconstitution. All crystallographic structures of the PB2 C-terminus (residues 536–759) reveal two globular domains, 627 and NLS, that form a tightly packed heterodimer. The molecular basis of the affinity of 627-NLS for importins remained unclear from these structures, apparently requiring large-scale conformational changes prior to importin binding. Using a combination of solution-state NMR, small-angle neutron scattering, small-angle X-ray scattering (SAXS), and Förster resonance energy transfer (FRET), we show that 627-NLS populates a temperature-dependent dynamic equilibrium between closed and open states. The closed state is stabilized by a tripartite salt bridge involving the 627-NLS interface and the linker, that becomes flexible in the open state, with 627 and NLS dislocating into a highly dynamic ensemble. Activation enthalpies and entropies associated with the rupture of this interface were derived from simultaneous analysis of temperature-dependent chemical exchange saturation transfer measurements, revealing a strong temperature dependence of both open-state population and exchange rate. Single-molecule FRET and SAXS demonstrate that only the open-form is capable of binding to importin α and that, upon binding, the 627 domain samples a dynamic conformational equilibrium in the vicinity of the C-terminus of importin α . This intrinsic large-scale conformational flexibility therefore enables 627-NLS to bind importin through conformational selection from a temperature-dependent equilibrium comprising both functional forms of the protein.



INTRODUCTION

Influenza A viruses are a major concern for human health, causing both annual epidemics and less frequent pandemics. Influenza A is a member of the Orthomyxoviridae family, with a segmented genome that comprises eight negative-sense RNA strands, encoding 10 major proteins and 5 auxiliary polypeptides.¹ The eight negative-polarity RNA genomic segments are each complexed with many copies of the nucleoprotein (NP) and one copy of the viral polymerase, forming the viral RNP that is packaged into a membrane-enveloped viral particle.^{2–5} The RNA polymerase complex

comprises three subunits: PA (acidic protein), PB1 (basic protein 1), and PB2 (basic protein 2). Following infection of the host cell, the RNP is imported into the nucleus where it performs the first cycle of transcription. The resulting viral mRNAs are translated in the cytoplasm to produce new viral proteins, of which PB2⁶ and the PA–PB1⁷ heterodimer are imported into the nucleus where they assemble into new

Received: July 24, 2015

Published: October 1, 2015

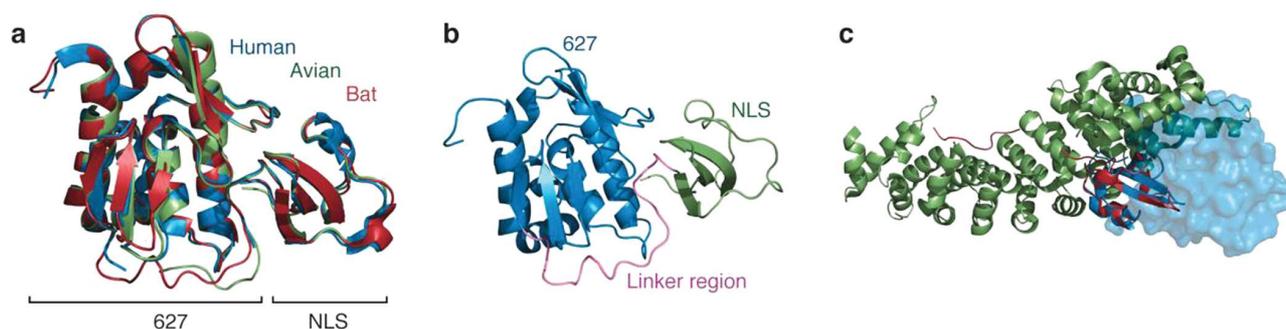


Figure 1. Crystallographic structure of 627-NLS is incompatible with importin binding. (a) Superposition of the crystal structures of 627-NLS of influenza A from human (PDB 2VY6, blue), avian (this work, green) together with the domain from bat influenza virus present in full heterotrimeric polymerase (PDB 4WSB, red). (b) Crystal structure of bat 627-NLS showing the linker peptide (magenta) connecting the 627 and NLS domains. (c) Crystal structure of human importin $\alpha 5$ (green) in complex with NLS (red) (PDB 2JDQ). Superposition of the human 627-NLS structure (blue) on the importin:NLS structure reveals a steric clash between importin and the 627 domain (transparent blue surface).

functional polymerase heterotrimers that catalyze further rounds of viral RNA replication and transcription.

The C-terminus of PB2 comprises two domains, the 627 (residues 538–677) and the NLS (residues 689–759) domains, showing negligible sequence or structural similarities with known proteins.^{6,8} The linear nuclear localization signal (NLS) peptide, located at the C-terminus of 627-NLS interacts with importin α *in vivo*, resulting in transfer of PB2 to the nucleus. Insight into the structural basis of the interaction of 627-NLS with importin α was initially obtained from the structure of the NLS domain free and in complex with importin $\alpha 5$,⁶ showing how the bipartite NLS peptide 736-RKRX₁₂KRIR-755 binds importin α . The structure of a construct comprising both 627 and NLS (residues 538–759) revealed a single well-packed conformation (Figure 1), although both domains are stable and soluble separately.^{6,9} The recent determination of the structure of bat influenza A polymerase complex, comprising subunits PA, PB1, and PB2, bound to its viral RNA promoter, revealed identical quaternary structure of these two PB2 domains within the context of the functional polymerase, where 627-NLS constitutes part of the putative basic exit channel for capped transcript RNA.¹⁰

Adaptation of avian viruses to human hosts is characterized by accumulation of mutations, especially in the PB2 C-terminal region comprising the 627-NLS domain, where surface residues exhibit positively charged clusters.^{9,11} Temperature-dependent replication has been observed¹² in line with adaptation of the viral proteins from the warmer environment of the bird intestine to the cooler mammalian respiratory tract, but the molecular basis of this effect is unknown. There is however evidence that mutations in this region directly modulate importin binding,^{13,14} while variations in the NLS peptide sequence have been shown to disrupt active polymerase assembly suggesting a second chaperone-like role for importins.¹⁵

Studying the interaction of 627-NLS with importin is therefore crucial to understanding the mechanistic role of this domain in nuclear import. The interaction is particularly intriguing since attempts to model the complex on the basis of the 627-NLS and NLS:importin $\alpha 5$ crystal structures predict a severe steric clash between the 627 domain and importin (Figure 1c), although 627-NLS binds importin with similar affinity to the NLS domain alone.¹⁶ These observations suggest that large scale conformational changes in importin or 627-NLS are necessary to allow formation of the complex in solution.

In the current study we investigate the molecular basis of the interaction between 627-NLS from avian influenza and importin $\alpha 1$ in solution using nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS), and single-molecule Förster resonance energy transfer (smFRET). These complementary methods reveal the coexistence of two very different conformations of 627-NLS in solution corresponding to open and closed states. The two forms interchange in a highly dynamic equilibrium with a slow characteristic exchange rate ($k_{\text{ex}} = 30 \text{ s}^{-1}$ at 15 °C). The closed form is stabilized by a tripartite salt bridge in the interface, implicating highly conserved basic and acidic amino acids in the 627 and NLS domains and the inter-domain linker.⁹ Mutation of one of these amino acids results in suppression of the closed form in solution. The population of the open form, in which the NLS domain is free to interact with importin, increases as temperature increases toward the physiological range. Analysis of the 627-NLS:importin complex using SAXS and smFRET shows that 627-NLS adopts an open conformation when binding importin, with the 627 domain remaining flexible on the surface of importin. These observations therefore present a structural, dynamic, and thermodynamic paradigm for understanding how the C-terminal region of PB2 interacts with importin and how the polymerase subunit is shuttled into the nucleus, identifying a role for conformational selection from a pre-existing equilibrium, and suggesting a possible role for temperature regulation of this essential equilibrium during inter-species adaptation.

RESULTS

627-NLS Exchanges between Open and Closed Conformations in Solution. We have determined the crystalline structure of 627-NLS from avian H5N1 influenza virus, which is found to be very similar (backbone rmsd 0.4 Å) to the previously determined human form,⁹ and the recently determined bat influenza B 627-NLS from the complete polymerase complex (Figure 1a)¹⁰ as well as other known 627-NLS structures.^{8,17,18} In most crystal structures it is possible to trace part or all of the peptide linking the two domains (residues 674–693), revealing strong similarity of the visible parts of the linker in human and avian forms (Figure 1b). As shown previously for human 627-NLS,⁹ superposition of the structure of avian 627-NLS with the NLS:importin $\alpha 5$ ⁶ structure via their common NLS domain results in a significant clash between the 627 domain and the C-terminal region of

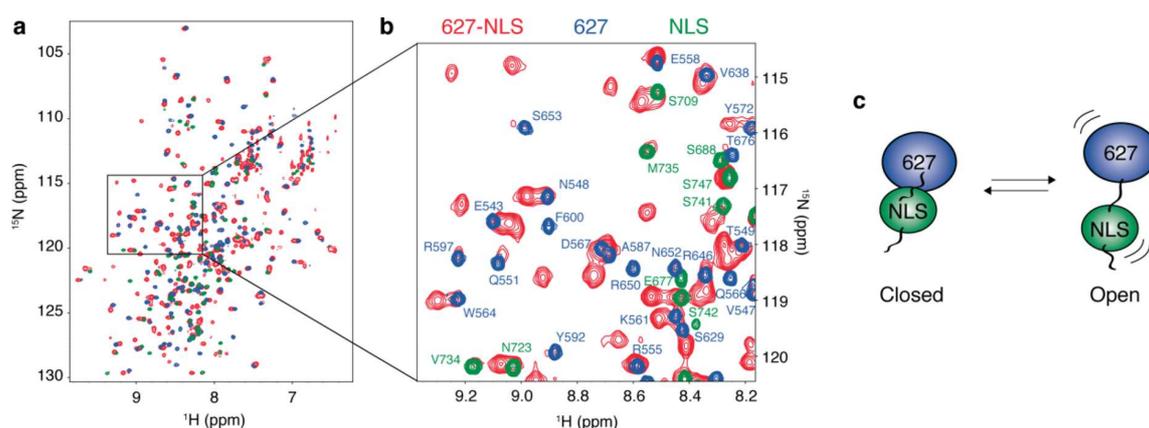


Figure 2. 627-NLS samples an equilibrium between closed and open conformations. (a) Comparison of ^1H - ^{15}N TROSY spectra of avian 627-NLS (red), 627 (blue), and NLS (green) at 25 °C. (b) Expanded region of the spectra displayed in (a) with assignments shown for the two separate domains 627 and NLS. (c) 627-NLS populates a dynamic equilibrium of closed and open conformations in slow exchange on the NMR chemical shift time scale.

importin α ,¹⁶ indicating that the conformation observed in the crystal structure is not binding-competent (Figure 1c). Attempts to crystallize the 627-NLS:importin complex were also unsuccessful (data not shown). Based on highly similar NLS-domain-bound structures of importin $\alpha 1$, $\alpha 3$, $\alpha 7$,¹⁹ and $\alpha 5$,⁹ identical clashes would be predicted with the 627 domain, although all isoforms have been shown to import PB2 *in vivo*,²⁰ including importin $\alpha 1$ studied here.¹³

To resolve this issue, we have investigated the conformational behavior of avian 627-NLS in solution using high-resolution NMR spectroscopy. Remarkably, ^1H - ^{15}N transverse relaxation optimized spectroscopy (TROSY) spectra of 627-NLS at 25 °C show around 350 peaks corresponding to backbone amides (Figure 2a), where only 213 are expected, indicating a conformational exchange of the protein on a slow NMR time scale. Comparison of TROSY spectra of isolated 627 and NLS domains with the spectrum of 627-NLS, shows a near-perfect superposition of the vast majority of resonances of 627 and NLS with one set of the 627-NLS resonances (Figure 2b). Assignment of the 627 and NLS domains and the two sets of peaks in the 627-NLS construct (Tables S1–S3) allows comparison of ^{13}C chemical shifts, demonstrating that the secondary structures of both sets of peaks are highly similar and largely coincident with that of the crystal structure (Figure 3a and Figure S1). Nearly all significant deviations localized to the linker region. ^{15}N chemical shifts are more sensitive to the local environment, and comparison of these shifts in the two forms maps predominantly to the interface between the two domains as well as the linker (Figure 3b,c and Figure S1d).

On the basis of these observations, we propose that 627-NLS undergoes slow exchange between a closed form, most likely the conformation found in the crystal lattice, and an open form in which the linker becomes flexible, and the two sub-domains, while retaining secondary and tertiary structure, dislocate and evolve more or less independently (Figure 2c). The open state is therefore highly dynamic, rapidly sampling the extensive degrees of conformational freedom allowed by the linker peptide, on a time scale that is fast enough to give rise to a single set of narrow NMR resonances ($\tau_{\text{ex}} < 50 \mu\text{s}$). We note that a similar behavior is observed for the human 627-NLS domain, which also displays sets of resonances corresponding to closed and open conformations (Figure S1e).

The Closed Form of 627-NLS Is Stabilized by a Tripartite Salt Bridge. Examination of the crystal structure of avian, human, and bat 627-NLS identifies a tripartite salt bridge in the interface between the two domains (Figure 4a), implicating a basic side chain from the 627 domain (R650) an acidic side chain from the NLS domain (D730), and an acidic side chain from the linker peptide (E687). This configuration is conserved in all available crystal structures of 627-NLS from different strains of influenza A and B,^{8–10,17,18} as is the presence of acidic and basic side chains at these positions (Figure S2). The implication of the linker peptide in this stabilizing interaction, resulting in structuring of the peptide and an uncommon insertion of the linker into the inter-domain interface, explains why this region of the peptide is structured and partially or fully resolved in the different crystal structures. A mutation of R650 to an alanine (R650A) removes all visible evidence of the closed form on the basis of NMR spectroscopy, with the only remaining peaks coinciding with the resonances of the individual domains (and therefore the open form) (Figure 4b). Spin relaxation demonstrates that the linker peptide (residues 676–689) of 627-NLS exhibits increased dynamics (Figure 4c), while the two domains present differential rotational correlation properties, related to their different molecular dimensions (130 compared to 60 amino acids) as determined from transverse auto- and cross-correlated relaxation rates (Figure S3). We note that the specific exchange regime characterizing the interchange between open and closed forms severely complicates interpretation of relaxation rates measured in wild-type protein due to averaging of dynamic characteristics of the open and closed states.

Equilibrium between Open and Closed 627-NLS Conformations Is Temperature-Dependent. We have used NMR spectroscopy to investigate the temperature dependence of the equilibrium between open and closed conformations. TROSY spectra of ^{15}N -labeled avian 627-NLS at 5, 10, 15, 20, 25, 30, and 35 °C were recorded. As temperature increases, the intensity of signals corresponding to the open conformation increases while that of the closed conformation decreases (Figure 5a). A representative sample of residues displaying well-resolved peaks, corresponding to open and closed conformations were quantified using a line shape analysis at the different temperatures, resulting in a population

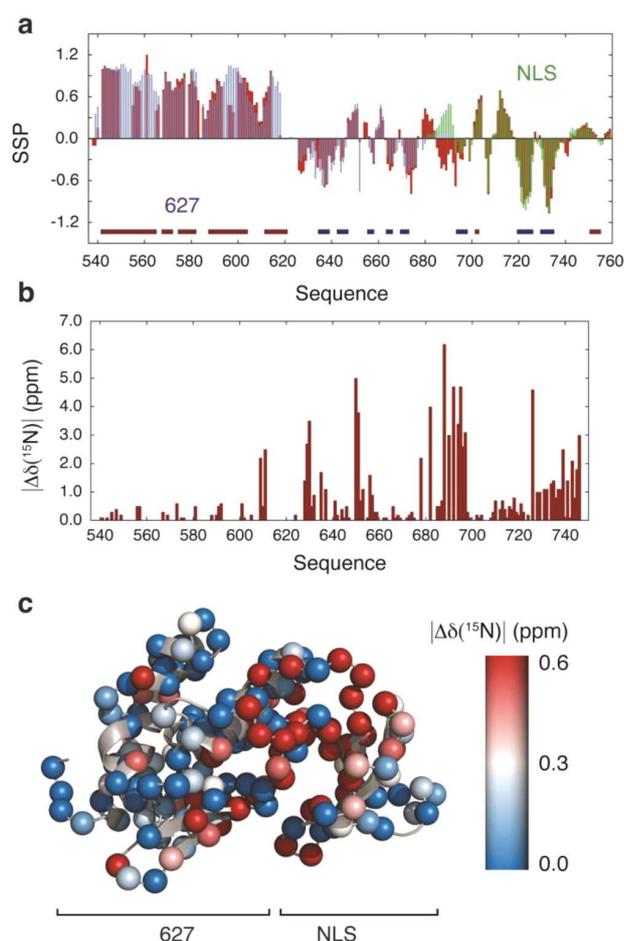


Figure 3. Chemical shift differences between individual domains and 627-NLS. (a) Secondary structure propensities (SSP) of 627 (blue, 25 °C), NLS (green, 25 °C), and 627-NLS (red, 10 °C) calculated from experimental $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts.⁶⁷ The chemical shifts of 627 were obtained using a deuterated sample, while the chemical shifts of 627-NLS were obtained using a perdeuterated sample. Corrections were made for isotope effects on experimental chemical shifts of perdeuterated 627-NLS using the corrections recently published by Maltsev et al.⁶⁸ In the case of 627-NLS, the SSP score is reported for the set of resonances corresponding to the closed form of 627-NLS or, if applicable, the resonances where the open and closed forms are indistinguishable. Horizontal bars in the bottom panel indicate the position of helices (red bars) and sheets (blue bars) as observed in the crystal structure of avian 627-NLS. (b) ^{15}N chemical shift differences between the resonances corresponding to open and closed conformations as a function of the sequence of 627-NLS. (c) ^{15}N chemical shift differences mapped onto the crystallographic structure of 627-NLS showing that the largest shifts correspond to the crystallographic interface between 627 and NLS.

shift from predominantly closed at 10 °C to approximately equal populations at 30 °C (Figure 5c). Chemical exchange saturation transfer (CEST)²¹ experiments were used to examine the conformational equilibrium in more detail, providing a precise determination of the variation of the exchange rate and population with temperature (Figure 5c,d). Activation enthalpies and entropies associated with breaking the stabilizing interface between the domains (Table 1) were derived from simultaneous analysis of CEST data from a series of eight well-resolved resonances corresponding to the closed

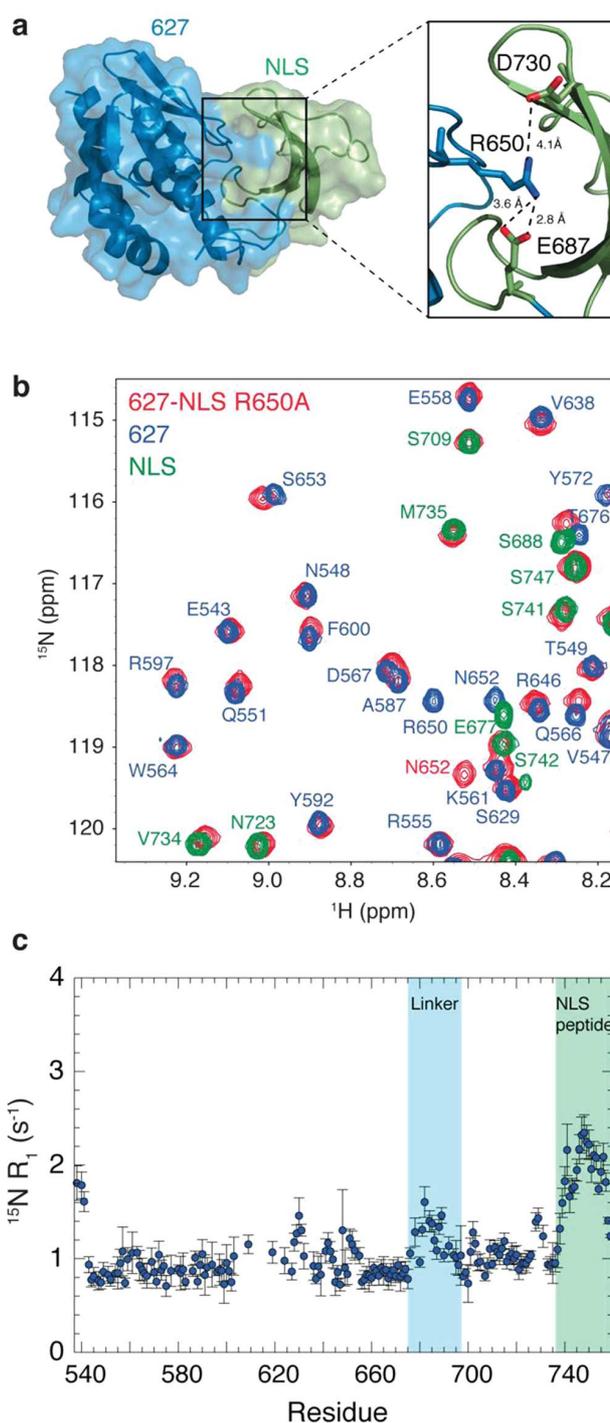


Figure 4. A salt bridge stabilizes the closed conformation of 627-NLS. (a) Crystal structure of avian 627-NLS showing that R650 stabilizes the closed conformation of 627-NLS via interactions with E687 of the linker region and D730 of the NLS domain. (b) Expanded region of the ^1H - ^{15}N TROSY spectra of 627-NLS R650A (red), 627 (blue), and NLS (green) obtained at 25 °C. (c) ^{15}N R_1 relaxation rates of 627-NLS R650A at a ^1H frequency of 600 MHz and 10 °C.

state of 627-NLS at temperatures ranging from 5 to 20 °C (two resonances reporting on residues from either the 627 or NLS domains are shown in Figure 5b) assuming an Eyring relationship (see Methods). All individual CEST traces are

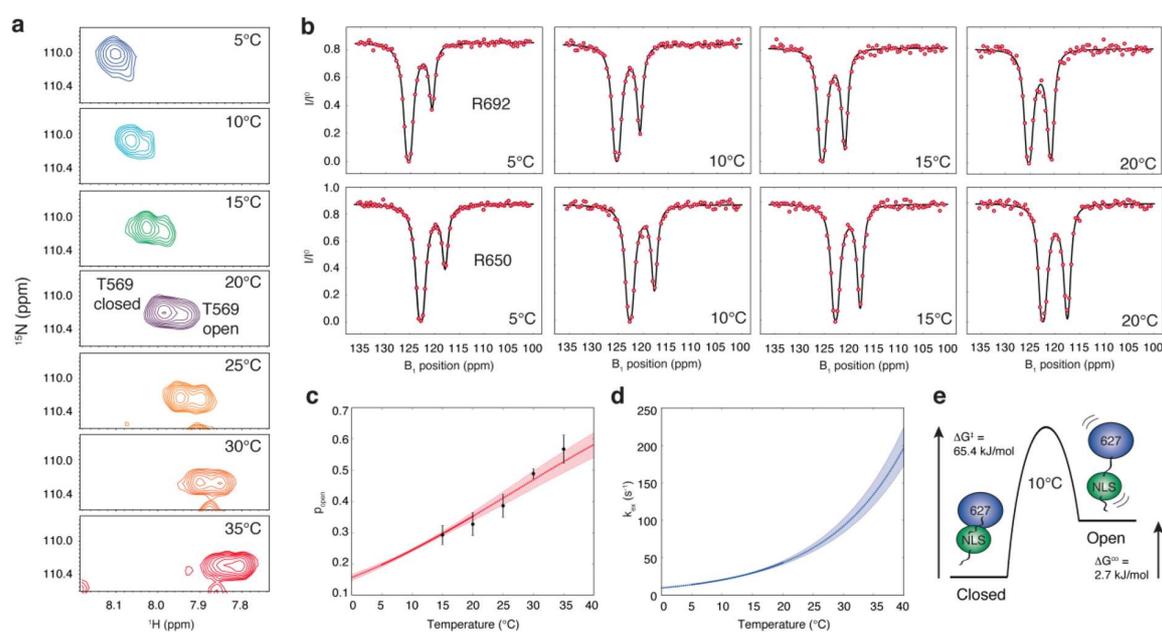


Figure 5. Characterization of the conformational exchange between open and closed conformations of 627-NLS. (a) Expanded regions of the ^1H - ^{15}N TROSY spectra of 627-NLS at multiple temperatures showing two resonances corresponding to open and closed conformations of the amide group of T569. (b) Experimental ^{15}N CEST profiles (red points) for the residues R692 and R650 obtained at multiple temperatures. The black lines correspond to a fit to a two-site exchange model by analyzing simultaneously data for multiple selected residues (Figure S4) at 5, 10, 15, and 20 °C assuming an Eyring relationship. (c) Population of open conformation (p_{open}) as a function of temperature derived from a line shape analysis of TROSY spectra at different temperatures (black points) or extracted from CEST experiments (red line). The full-drawn red line indicates the temperature range where CEST experiments were recorded. The shaded red area shows the 68% confidence interval as obtained by 500 Monte Carlo simulations. (d) Exchange rate (k_{ex}) between open and closed conformations of 627-NLS obtained from CEST experiments (blue line). The full-drawn blue line indicates the temperature range where CEST experiments were recorded. The shaded blue area shows the 68% confidence interval as obtained by 500 Monte Carlo simulations. (e) Schematic representation of the energy diagram of the open and closed states of 627-NLS at 10 °C with ΔG values derived from CEST experiments (see Methods).

Table 1. Thermodynamic Parameters of the Conformational Equilibrium between Open and Closed Conformations of 627-NLS Derived from CEST Experiments Recorded at Four Different Temperatures^a

	ΔH (kJ mol ⁻¹)	ΔS (J mol ⁻¹ ·K ⁻¹)	$\Delta G_{10\text{ °C}}$ (kJ mol ⁻¹)
energies of open state	36 ± 4	117 ± 15	2.66 ± 0.06
energies of activation	74 ± 2	32 ± 7	65.44 ± 0.02

^aRecorded at 5, 10, 15, and 20 °C. The CEST experiments were analyzed simultaneously at all temperatures by including experimental data for selected residues (Figure S4) and assuming an Eyring relationship (see Methods). Energy values are reported relative to the energy of the closed state.

remarkably well reproduced using this simple model (Figure S4). Not surprisingly, the activation energy required to rupture the interactions in the interface (65 kJ/mol at 10 °C), is largely dominated by enthalpic contributions, while the free energy difference of the open compared to the closed state (2.67 kJ/mol at 10 °C) reports on dominant enthalpic contributions at lower temperatures that are matched by entropic effects as temperature increases (Figure 5e). Overall these results provide an unambiguous description of the thermodynamic origin of the increase of the exchange rate and open-state population with temperature. We note that the population dependence coincides very closely with the values derived from line shape analysis (Figure 5c). Additional CEST-based analyses were performed using human 627-NLS, the more common partner of importin $\alpha 1$ but which shows higher levels of instability *in*

vitro compared to the avian form, resulting in population and exchange rate dependences that were identical within experimental uncertainty (data not shown).

SAXS Measurements Confirm the Temperature Dependence of the Equilibrium. Open and closed forms were further characterized using SAXS. Data were collected for avian 627-NLS at concentrations of 4.9, 2.45, and 1.23 mg/mL and over a temperature range from 5 to 30 °C. Profiles obtained at different protein concentrations exhibited the same shape and were flat at low q values, indicating the absence of aggregation (Figure 6a). The radius of gyration, R_g , values derived from the Guinier approximation at low q value exhibited clear temperature dependence, with higher R_g at higher temperatures (Figure 6a). SAXS curves were simulated for the closed conformation and the open state using the *flexible-meccano* statistical coil description^{22,23} for the linker region and the CRY SOL²⁴ algorithm for prediction of scattering profiles. Combination of these limiting curves in proportions coincident with those determined using NMR spectroscopy (CEST) results in excellent reproduction of the experimental SAXS curves obtained at the limiting temperatures (5 and 30 °C) (Figure 6b,c).

The Open Conformation of 627-NLS Binds Importin $\alpha 1$. The 627-NLS:importin $\alpha 1$ complex was investigated using small angle scattering. SAXS profiles from the complex were analyzed using two procedures. First, *flexible-meccano* was used to sample the conformational space available to the 627 domain, assuming that the linker peptide (residues 676–689) can sample all statistical coil conformations available to this

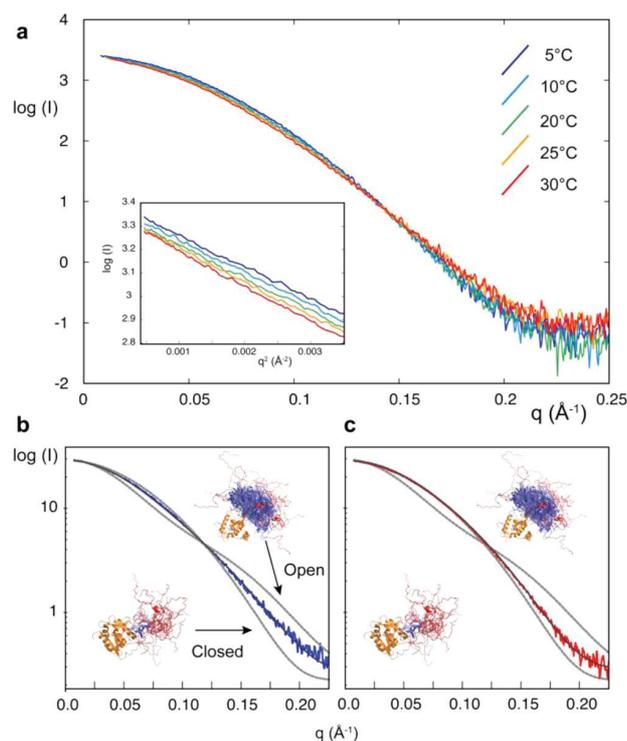


Figure 6. Small-angle X-ray scattering from 627-NLS in solution. (a) Temperature dependence of SAXS curves of the same 627-NLS sample measured at 5, 10, 20, 25, and 30 °C. Inset: Guinier region of the SAXS curve plotted as $\log_e(I)$ vs q^2 , showing that the slope gets steeper in this region as temperature increases. The extracted radii of gyration are 20.4, 20.8, 21.1, 21.3, and 21.7 Å for 5, 10, 20, 25, and 30 °C, respectively. (b,c) Comparison of experimental SAXS data to an admixture of scattering curves calculated using Crysol for the closed form and averaged over an open ensemble calculated using *flexible-meccano* (limiting curves shown in gray). The ratios of open and closed curves were taken from the populations determined from NMR line shape and CEST analyses: (b) 0.2:0.8 at 5 °C and (c) 0.5:0.5 at 30 °C.

peptide sequence. Conformations were built onto the X-ray crystal structure of the NLS domain bound to importin $\alpha 1$, assuming that the NLS domain, whose conformation is conserved in all four complexes with importin, retains this crystallographic position, an assumption supported by mutagenesis and crystallographic studies.¹⁹ SAXS curves were predicted from each of these 25 000 conformers. Note that the reproduction of experimental data by individual curves provides significant improvement over the average of all curves, indicating that the sterically available conformational space is not sampled uniformly. Conformers were then ranked with respect to their ability to reproduce the experimental SAXS curves (Figure 7a), revealing that all high-ranking solutions are positioned near the C-terminal end of importin (Figure 7b). The data reproduction derived from this procedure is already of high quality (Figure 7d); nevertheless, sub-ensembles of conformers may be selected from all possible available positions (Figure 7a), thereby accounting explicitly for the flexibility of 627 relative to NLS-importin. Not surprisingly this approach results in ensembles with a high density of states in the same region of three-dimensional space relative to importin as the single-conformation solutions (Figure 8).

Single-conformer and ensemble solutions that best reproduce the experimental SAXS curve also provide a significantly better

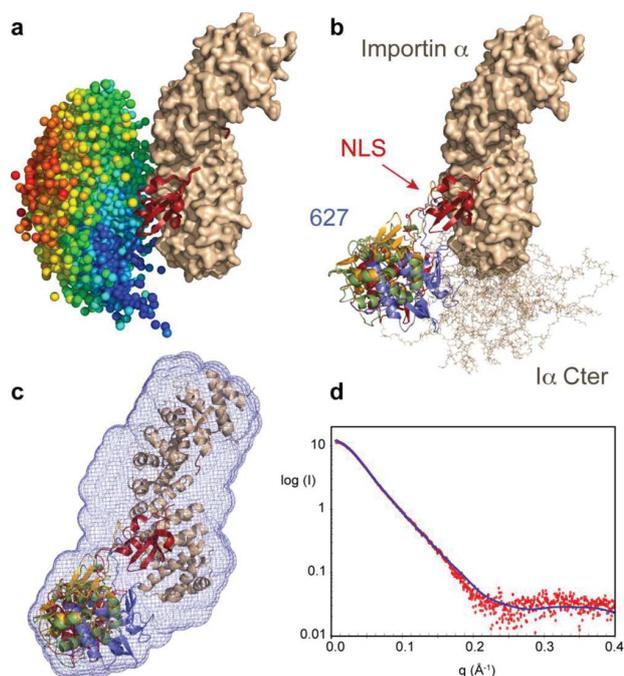


Figure 7. Structural analysis of the 627-NLS:importin $\alpha 1$ complex from small-angle X-ray scattering. (a) Representative depiction of the 25 000 centers-of-mass of different conformers of 627-NLS calculated using the *flexible-meccano* statistical coil algorithm to model the conformational space available to the 627 domain when the NLS domain is bound to importin. The center-of-mass spheres are colored with respect to the ability of the associated conformation to reproduce experimental SAXS data: red, $\chi > 4$; yellow, $4 > \chi > 3$; green, $3 > \chi > 2$; light blue, $2 > \chi > 1$; and dark blue, $\chi < 1$. (b) Representative conformations of 627 that best reproduce the experimental data are shown in cartoon representation. The position of the C-terminal disordered tail of importin $\alpha 1$ is also shown. (c) Comparison of the position of the best-fitting conformations with the *ab initio* SAXS envelope determined using DAMMIF and DAMAVER (blue). (d) Comparison of experimental SAXS data (red) with scattering profile predicted from the best fitting conformation (blue).

reproduction of independently measured small-angle neutron scattering (SANS) data (hydrogenated protein in buffer containing 100% D₂O) than the average over all conformational space (Figure S5). Under these conditions, SANS data report more faithfully on the scattering properties of the proteins alone,^{25,26} whereas SAXS data report on a combination of the solvent shell and the protein, making the two sources of information complementary and thereby further substantiating the model of the complex.

Second, an *ab initio* reconstruction was performed using DAMMIF²⁷ to produce a space-filling envelope of the scattering object. Although some empty density can be seen in this model, possibly due to flexibility of either importin $\alpha 1$ or the 627 domain, the envelope encompasses very well the single conformation solutions derived from broad sampling of conformational space, showing good overall agreement of the two approaches (Figure 7c).

smFRET Analysis of the Dynamic Properties and Importin α Binding Mode of 627-NLS. The conformational equilibria sampled in the free and bound forms of 627-NLS were also investigated by smFRET, using a double-cysteine mutant of 627-NLS (residues 539 and 707, positioned at either

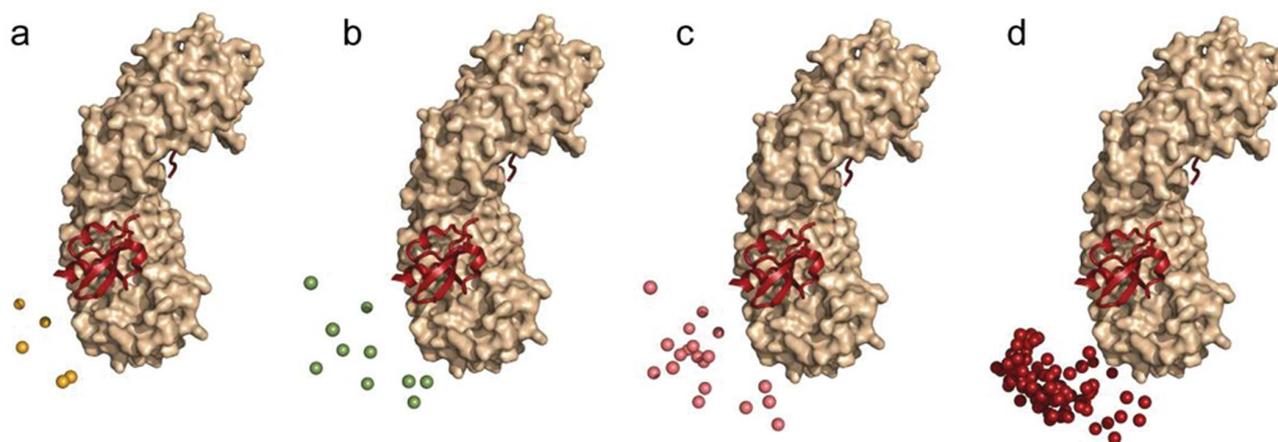


Figure 8. Ensemble selections of the 627-NLS:importin $\alpha 1$ complex. Comparison between centers-of-mass of the 627 domain in ASTEROIDS-selected ensembles of 627-NLS-importin $\alpha 1$ containing (a) 5, (b) 10, (c) 20, and (d) the 100 conformers presenting the lowest χ^2 of individual conformations against the experimental data. The ensembles shown in (a-c) were selected from the entire 25000 orientations represented in Figure 7a.

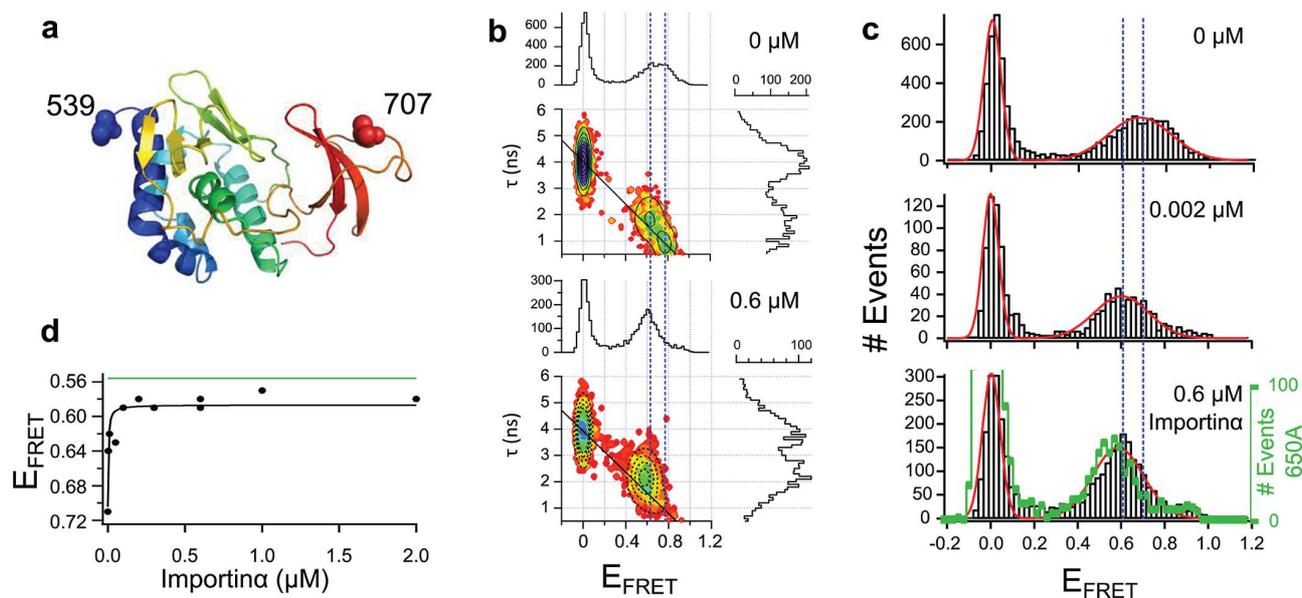


Figure 9. 627-NLS interaction with importin $\alpha 1$ probed by smFRET. (a) Structure of 627-NLS with labeling sites marked as spheres. (b) Donor fluorescence lifetime (τ) versus FRET efficiency histograms of 627-NLS double-cysteine mutant I539C A707C in the presence and absence of 0.6 μM importin $\alpha 1$ with corresponding two-dimensional triple-Gaussian fits (black dashed curves) and the black line describing the relationship between τ and E_{FRET} for a static molecule. The blue dashed lines are centered at the FRET efficiencies of open and closed conformations of the unbound state. (See Figure S7 for the triple Gaussian fits of the corresponding titration.) (c) Typical one-dimensional E_{FRET} histograms of 627-NLS double-cysteine mutant I539C A707C at varying importin $\alpha 1$ concentrations with corresponding Gaussian fits (red curves). The blue dashed lines illustrate E_{FRET} at 0 μM (right) and 0.6 μM importin $\alpha 1$ (left). The green histogram stems from 627-NLS R650A I539C A707C. The histogram at 0 μM importin $\alpha 1$ was accumulated from four independent measurements. (d) E_{FRET} as obtained from a titration of 627-NLS I539C A707C with importin $\alpha 1$. Fitted values (black spheres) were obtained from double Gaussian fits of two-dimensional E_{FRET} versus τ histograms, allowing a tilt between the E_{FRET} and τ axes within one population (Figure S6). E_{FRET} relation with importin $\alpha 1$ concentration was described using a simple binding model, $[AB] = A_{\text{tot}}B/(K_d + B) + \text{offset}$, where A is 627-NLS and B is importin $\alpha 1$, under the assumption that importin $\alpha 1$ was added at large excess at every titration step (black line). The green line illustrates E_{FRET} as obtained from 627-NLS R650A double-cysteine mutant I539C A707C.

extreme of the closed 627-NLS structure) randomly labeled with Alexa488 and Alexa594 (Figure 9a). 627-NLS was subjected to multi-parameter smFRET analysis^{28,29} and yielded a donor (Do, Alexa488)-only population, which lacked active acceptor (Ac, Alexa594) fluorophores, at a FRET efficiency (E_{FRET}) of 0, as well as a FRET population (E_{FRET}) reporting on distance and dynamics between Do and Ac. Quantitative analysis of the width of the FRET population (Table S4)

together with fluorescence lifetime (τ , Figure S6) and observation time analysis (Figure S7) points toward the existence of more than one conformation that is underlying the FRET peak (see Methods for a more detailed description).^{30–32} This is in agreement with the existence of open and closed states in 627-NLS and an exchange rate between these states (50 s^{-1} at 25 $^{\circ}\text{C}$, as determined by NMR) that is significantly slower than the average single-molecule

observation time (up to ~ 5 ms). It is likely therefore that the FRET population in the unbound state is built up from these two states. Following this logic, fitting of the E_{FRET} versus τ histogram with a triple Gaussian function (one Gaussian describing the Do-only population, two describing the FRET population) results in average $E_{\text{FRET}} = 0.62$ and 0.77 (Figure 9b,c). With an approximate dye-to-dye distance of 47 \AA , the molecules giving rise to $E_{\text{FRET}} = 0.77$ likely reflect the closed conformation of 627-NLS as observed in the crystal structure (44 \AA Ca–Ca distance), and behave similarly to static molecules, as can be seen from the corresponding E_{FRET}/τ relationship that follows the “static FRET line” (Figure 9b). Although we do not resolve quantitative modes of 627-NLS dynamics^{28,32,33} from these data, we observe that molecules in the open state, giving rise to $E_{\text{FRET}} = 0.62$, deviate from static behavior,^{31,32} indicating sub-microsecond averaging of different conformations within this state. The two states, open and closed, are occupied with a similar number of molecules at the smFRET temperature of $23 \text{ }^\circ\text{C}$, approximately reflecting the occupancy ratios observed by NMR (45:55 compared to 40:60 from NMR). NMR and smFRET are therefore shown to be highly complementary; while NMR provides a clear separation of time scales of interconversion between conformers in the open-form, and between the open and closed states, smFRET can be used to probe the distance distribution functions present within the two states.

When importin α was added, the FRET population shifted gradually to lower FRET efficiencies (longer distances) in agreement with an increase in the open conformation when compared to the unbound form of the 627-NLS domain (Figure 9c,d and Figure S6). A fit of the corresponding τ versus E_{FRET} histograms with a triple Gaussian function, as was done for the unbound 627-NLS (Figure 9b), suggests that the closed form gradually disappears from the distribution, substantiating the suggestion, obtained from SAXS, that only the open form is capable of binding importin (Figure 9b, Figure S7, and Table S5). It is striking that, even in the bound form, the events coming from the 627-NLS open conformation deviate from the static FRET line, indicating that 627-NLS remains flexible when bound (Figure 9b).

Single-molecule FRET of the R650A mutant confirms that the closed form is barely detectable in the ensemble in the absence of the stabilizing salt bridge, leading to slightly more extended distances compared to the open form measured in the native protein (Figure 9c and Figure S6). Single-molecule FRET on 627-NLS in the presence $5 \mu\text{M}$ importin $\alpha 3$, $\alpha 5$, and $\alpha 7$ gave effectively identical results to those found for importin $\alpha 1$, again revealing an undetectably low population of the closed form of 627-NLS, and very similar FRET efficiencies for the open form (Figure S8). This suggests the 627-NLS binding mode detected for importin $\alpha 1$ is indeed general across the importin α family.

DISCUSSION

Although the three-dimensional structure of influenza virus polymerase has recently been determined to atomic resolution,¹⁰ the molecular basis of import into the nucleus of the constituent PB2 subunit prior to reconstitution of the functional polymerase remains poorly understood. The C-terminal 627-NLS region of influenza PB2 polymerase is crucial for this interaction, as it is terminated by the NLS peptide that interacts with nuclear import receptors. This domain also contains a high density of mutations that are important for

inter-species transmission, suggesting that nuclear import may play an important role in adaptation. Here we have determined the crystal structure of a fully avian H5N1 627-NLS domain that is essentially indistinguishable from human 627-NLS and from the structure of bat influenza 627-NLS in the context of the full-length polymerase. In all available structures^{8–10,17,18} 627-NLS is stabilized in a compact conformation by a tripartite salt bridge, implicating charged amino acids in the interface between the two domains, and a third amino acid in the peptide linking the independently folded domains. On the basis of the crystal structure of the NLS domain in complex with importin, it was previously predicted that this closed conformation could not bind importin α unless 627-NLS or importin α undergo large scale conformational changes that may be further implicated in active trimeric polymerase assembly.¹⁶

Multi-domain proteins comprise 80% of eukaryotic proteomes, where inter-domain dynamics play key roles in a multitude of molecular recognition, transport, and signaling processes.³⁴ These complex dynamic modes cannot be understood from static structures of either the entire protein or individual domains. Indeed intrinsically disordered linkers connecting folded subunits often encode the degrees of conformational flexibility essential to protein function, requiring solution state methods to describe the mechanisms underlying their function.^{35,36} We have therefore used NMR spectroscopy, small angle scattering and smFRET to characterize the behavior of avian 627-NLS in solution and its binding to importin α .

NMR unambiguously detects large-scale conformational changes of 627-NLS in solution, comprising exchange between a closed form and a previously uncharacterized open state that is highly dynamic. The chemical shifts of the open state very closely resemble the independent domains, indicating that the domains evolve more or less independently of each other in solution, sampling multiple conformational states with characteristic exchange time scales faster than $50 \mu\text{s}$. The two domains are separated by a short 12–14 amino acid linker that is partially structured in the closed state, participating in the electrostatic interaction that stabilizes the closed form, but that becomes flexible in the open state, providing the degrees of freedom necessary for the large-scale domain dynamics. Mutation of one of the highly conserved amino acids that constitutes the core of the stabilizing interaction results in a conformational equilibrium comprising only the open-state equilibrium, as demonstrated by both NMR and smFRET.

Exchange between the open and closed forms of 627-NLS occurs on a very slow time scale ($k_{\text{ex}} = 30 \text{ s}^{-1}$ at $15 \text{ }^\circ\text{C}$), exhibiting clear temperature dependence, rising to 80 s^{-1} at $25 \text{ }^\circ\text{C}$. Equally importantly, the population of the open state is strongly temperature-dependent, characterized by a predominantly closed form at $5 \text{ }^\circ\text{C}$ (20%:80% open:closed population) with population of the open form rising to approximately 50% at $30 \text{ }^\circ\text{C}$. NMR exchange and small angle scattering measured over a physiological range concord very closely with this description. A precise investigation of the thermodynamics governing the temperature-dependent equilibrium, derived from simultaneous analysis of CEST as a function of temperature, indicates that although enthalpic differences dominate the activation energy, the free form exhibits both entropic and enthalpic contributions compared to the closed state, suggesting that the open state may exhibit nonspecific nonbonding interactions between the surfaces.

The unambiguous identification of interconversion time-ranges within the open state ($\tau_{\text{ex}} < 50 \mu\text{s}$) and between the open and the closed states ($\tau_{\text{ex}} > 20 \text{ms}$) provides a clear framework within which smFRET can be used to probe the distance distributions between chromophores present in the two domains. This allows us to support the suggestion that the closed state indeed corresponds to the crystal conformation, and identifies significant domain dynamics in the open state, again in agreement with the model derived from NMR and SAXS.

Solution state approaches were also used to probe the interactions with importin $\alpha 1$, demonstrating that the 627 domain is detached from the NLS domain in the complex, sampling conformations in the vicinity of the C-terminus of importin α . The 627 domain clearly does not sample all of the available space provided by the flexible linker, possibly due to interactions with the surface of importin α in this region, or via interactions with the 34 amino acid long flexible C-terminal tail of importin α . This domain clearly remains flexible when the NLS domain is bound to importin as shown from both SAXS and smFRET. We note that the C-terminal tail of importin α presents a highly conserved acidic strand at its N-terminus, and a conserved cluster of aromatic moieties at its C-terminus,³⁷ which may interact with the surface of 627. The smFRET results suggest that the population of the closed state of 627-NLS is reduced to very low levels in complex with importin α , corroborating the model by which the open state is the binding competent form. Experiments carried out on four importin α isoforms ($\alpha 1$, $\alpha 3$, $\alpha 5$, and $\alpha 7$) indicate that this mechanism is general for PB2 binding to all members of the importin α family.

A recent study of residues involved in enhanced polymerase activity in 627-NLS found that mutation of the basic 650 residue, an essential constituent of the stabilizing salt bridge, resulted in efficient nuclear import but reported a lack of polymerase activity in the nucleus.³⁸ This supports our observation that the open form mediates interaction with importin α , and further supports the recent observation that the closed form is essential for function within the context of the reconstituted polymerase-RNA complex, where it has been proposed to play a crucial role in interaction with transcribed RNA in the polymerase exit tunnel.¹⁰

Although NMR exchange experiments comparing the avian and human forms were not able to detect any significant difference in the thermodynamics of the equilibrium, it is possible that the thermodynamic dependence of the solution state equilibrium of 627-NLS is related to temperature adaptation from the warmer environment of the bird intestine (40 °C) to the cooler mammalian respiratory tract (33 °C). Further studies are required to substantiate this possibility.

CONCLUSION

A combination of complementary solution-state approaches reporting on local and long-range protein dynamics is used to characterize the conformational behavior of the C-terminal 627-NLS domain of influenza polymerase PB2. Although in the crystalline form 627-NLS systematically adopts a closed, compact conformation, in solution the protein samples a dynamic equilibrium between the closed state and an open state, in which the two domains dislocate when a highly conserved tripartite salt bridge in the 627-NLS interface is ruptured. The open state is highly dynamic, comprising independent domains joined by the linker that also becomes

flexible. Open and closed states are present in equilibrium in slow exchange, with the open-state population, and the exchange rate, both increasing with temperature. Crucially only the open-form is capable of binding to the nuclear import receptor, and upon binding the 627 domain samples a dynamic conformational equilibrium in the vicinity of the C-terminus of importin α . This study demonstrates the functional importance of large-scale domain dynamics, mediated by the intrinsically disordered linker peptide connecting the two domains, and thereby provides a framework for understanding nuclear import of influenza polymerase. In particular we observe that the interaction of PB2 with importin α likely occurs via conformational selection from the intrinsic pre-existing equilibrium, rather than induced fit. More generally, this study provides a fascinating example of how a combination of solution-state approaches can precisely report on dynamic conformational equilibria, and as such offer the possibility to study the function of proteins exhibiting extensive and functionally important domain dynamics.^{39,40}

METHODS

Protein Expression and Purification. Codon-optimized *pb2* synthetic genes were synthesized encoding amino acids 538–759 (627-NLS domain) from human H3N2 A/Victoria/3/1975 and avian H5N1 A/duck/Shantou/4610/2003 for optimal expression in *E. coli* (Geneart, Regensburg, Germany). Two additional PB2 constructs from both human H3N2 A/Victoria/3/1975 and avian H5N1 A/duck/Shantou/4610/2003 were used: the 627 domain, aa 538–693 and the NLS domain, aa 678–759. The human importin- $\alpha 1$ isoform lacking the autoinhibitory N-terminal importin- β -binding domain, but with native C-terminus (KPNA2; Uniprot P52292, aa 60–529) was also expressed from a synthetic gene as above. All constructs were subcloned into a pET9a-derived vector with an N-terminal hexahistidine tag and TEV protease cleavage sequence (MGHHH-HHHYDIPTTENLYFQG).

The plasmids were transformed into *E. coli* BL21(DE3)RIL cells (Agilent), and the expression of the recombinant proteins was induced with 1 mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) for 18 h at 16 °C. Cells were harvested by centrifugation, suspended in buffer A (50 mM Tris-HCl, 200 mM NaCl), pH 7.5, containing protease inhibitors (Complete EDTA-Free, Roche Diagnostics) and disrupted by sonication. The extract was centrifuged at 20000g for 30 min at 4 °C. All proteins were purified by affinity chromatography on a Ni²⁺ resin column (Qiagen) followed by incubation with TEV protease overnight at 4 °C coupled with a dialysis against buffer A, pH 7. A second affinity chromatography on a Ni²⁺ resin column was performed. The flow-through was then further purified using SEC with a Superdex S75 column (GE Healthcare) equilibrated in buffer A at pH 6.5.

Samples for NMR spectroscopy were produced in M9 minimal medium containing MEM vitamins (Gibco). For producing ¹⁵N-labeled or ¹⁵N,¹³C-labeled 627-NLS, 627, and NLS domains, the medium was supplemented with 1.0 g/L of ¹⁵NH₄Cl and 2.0 g/L of unlabeled glucose or 1.0 g/L of ¹⁵NH₄Cl and 2.0 g/L of ¹³C-glucose, while for producing ¹⁵N,¹³C,²H-labeled 627-NLS, the minimal medium was prepared in D₂O and supplemented with 1.0 g/L of ¹⁵NH₄Cl and 2.0 g/L of deuterated ¹³C-glucose. The purity and monodispersity of the samples were checked by SDS-PAGE, dynamic light scattering, and mass spectrometry.

The avian 627-NLS:human importin $\alpha 1$ complex was obtained by mixing both purified proteins in a ratio 1.5:1. Because SAXS experiments require a very high quality sample in term of monodispersity, the obtained complex was further purified using SEC on Superdex 75 (GE Healthcare) equilibrated with 50 mM Tris-HCl, pH 6.5, 200 mM NaCl, and the subsequent sample was ultracentrifuged at 50000g during 20 min. The quality of the sample

containing the complex and the monodispersity were assessed by SDS-PAGE and dynamic light scattering, respectively.

Mutants of avian 627-NLS were made using the QuikChange site-directed mutagenesis kit (Agilent). A single mutant (R650A) was made for NMR measurements of the open conformation of 627-NLS, a double mutant I539C/A707C and a triple mutant R650A/I539C/A707C were made for smFRET measurements. All mutants were expressed and purified as described above for the wild-type protein. For the cysteine mutations, 10 mM dithiothreitol (DTT) was added after the second Ni-NTA column to keep the proteins in a monomeric state until labeling them for smFRET.

Structure Determination of the Avian 627-NLS Domain.

Hanging drop vapor diffusion trials were carried out at 20 °C. Native crystals of avian 627-NLS were obtained by mixing 1 μ L of protein solution at 12 mg/mL in 20 mM HEPES (pH 7.5), 200 mM NaCl, with an equal volume of a solution containing 0.1 M Bis-Tris (pH 6.5), 30% (w/v) PEG 4000, and 0.08 M magnesium acetate tetrahydrate. Crystals were washed in crystallization solution supplemented with 30% glycerol and frozen in liquid nitrogen.

Data were collected to 1.4 Å resolution on beamline BM14 at the European Synchrotron Radiation Facility (ESRF). All data collection and refinement statistics are given in Table S6. All data were integrated with XDS⁴¹ and analyzed using the CCP4i package.⁴² The structure was determined by molecular replacement using PHASER⁴³ with the human PB2 627-NLS domain (PDB 2VY6) as search model. ARP/wARP⁴⁴ was used for automatic model building. All refinements were performed with REFMAC.⁴⁵ According to MOLPROBITY (<http://molprobity.biochem.duke.edu/>), all structures have excellent geometry.

NMR Measurements. All samples for NMR contained 50 mM Tris-HCl pH 6.5, 200 mM NaCl, and 10% D₂O. The assignment of the avian NLS domain was obtained at 25 °C using a ¹⁵N,¹³C-labeled sample (1.0 mM) using BEST-type triple-resonance experiments⁴⁶ recorded on a Varian spectrometer operating at a ¹H frequency of 600 MHz, equipped with a cryoprobe. The assignment of the avian 627 domain was obtained at 25 °C using a ¹⁵N,¹³C-labeled sample (0.66 mM) using standard triple-resonance experiments recorded on a Varian spectrometer operating at a ¹H frequency of 800 MHz, equipped with cryoprobe.

The assignment of the avian 627-NLS domain was obtained at 10 °C using a ¹⁵N,¹³C,²H-labeled sample (0.5 mM) using BEST-TROSY triple-resonance experiments⁴⁷ recorded on a Bruker spectrometer operating at a ¹H frequency of 950 MHz. The assignment of 627-NLS was carried out at 10 °C to minimize line broadening due to exchange between open and closed conformations. The assignment of the R560A mutant of the avian 627-NLS domain was obtained at 10 °C using a ¹⁵N,¹³C-labeled sample (2.0 mM) using BEST-TROSY triple-resonance experiments recorded on a Bruker spectrometer operating at a ¹H frequency of 950 MHz. For all four proteins the spectra were processed in NMRPipe,⁴⁸ manually peak-picked in Sparky,⁴⁹ and MARS⁵⁰ was used for identification of spin systems followed by manual verification.

The evolution of the spectra with temperature was investigated by performing a series of ¹H-¹⁵N TROSY (627-NLS) or ¹H-¹⁵N HSQC (627 and NLS) spectra at 5, 10, 15, 20, 25, 30, and 35 °C. Intensities of the resonances corresponding to the open and closed conformations of 627-NLS were obtained by a line shape analysis (Lorentzian) at each temperature using the program FuDA (<http://www.biochem.ucl.ac.uk/hansen/fuda/>). The populations of the open and closed forms at each temperature were calculated from the extracted intensities.

CEST experiments²¹ were performed on a sample of 627-NLS (0.8 mM) at 5, 10, 15, and 20 °C on a Bruker spectrometer operating at a ¹H frequency of 700 MHz. The experiments were recorded with a ¹⁵N B₁ saturating field of 22 Hz (5, 10, 15, and 20 °C) and an additional CEST experiment was recorded at 10 °C using a B₁ field of 44 Hz. The B₁ field was applied during a constant period of 0.3 s. The experiments were recorded with 84 two-dimensional planes with the position of the ¹⁵N B₁ field ranging from 100 to 136 ppm in steps of 30 Hz. The CEST data were analyzed using ChemEx assuming a two-

site exchange model. To increase the robustness of the analysis of the CEST experiments, data at 5, 10, 15, and 20 °C were analyzed simultaneously assuming an Eyring relationship,

$$k = \frac{k_B T}{h} \exp\left(-\frac{\Delta H - T\Delta S}{RT}\right) \quad (1)$$

where k is the rate constant, ΔH (in J/mol) and ΔS (in J/mol/K) are the enthalpy and entropy of activation, respectively, k_B is the Boltzmann constant, h is Planck's constant, and R is the gas constant. Multiple resonances corresponding to the closed conformation of 627-NLS were included in the global fit (R650N, G651N, G682N, S688N, R692N, G693N, F694N, and I726N) at each temperature, and uncertainties in the entire fitting procedure were determined using 500 noise-based Monte Carlo simulations. The global fit of the CEST data thus characterizes the thermodynamic properties of the exchange process between the open and closed state of 627-NLS.

Nuclear relaxation rates of the 627-NLS R650A mutant were measured at 10 °C on a Bruker spectrometer operating at a ¹H frequency of 600 MHz. The ¹⁵N longitudinal relaxation rates (R_1) and $\{^1\text{H}\}^{15}\text{N}$ heteronuclear NOEs were measured using the methods of Lakomek et al.,⁵¹ the ¹⁵N transverse relaxation rates (R_2) were measured using the method of Kay et al.⁵² and the ¹⁵N-¹H CSA/DD cross-correlated transverse cross-relaxation rates (η_{xy}) were measured using the pulse scheme of Wang et al.⁵³ All spectra were processed in NMRPipe,⁴⁸ and in-house scripts were used to extract peak heights and calculate the relaxation rates.

Small-Angle X-ray Scattering of Avian 627-NLS. SAXS experiments of avian 627-NLS were obtained at 5, 10, 20, 25, and 30 °C at the BM29 beamline, ESRF, Grenoble⁵⁴ using an energy of 12.5 keV, $\lambda = 0.9918$ Å. All samples were purified as described above. Fifty μ L of protein solution at three different concentrations (4.9, 2.45, and 1.23 mg/mL) and the corresponding buffers were exposed to X-rays and scattering data collected using the robotic sample handling available at the beamline.⁵⁵ Ten individual frames were collected for every exposure, each 1 s in duration using the Pilatus 1M detector (Dectris) placed 2.864 m from the sample. Individual frames were processed automatically and independently using the EDNA processing pipeline incorporating the ATSAS tools from EMBL-Hamburg, yielding individual radially averaged curves of normalized intensity versus scattering angle $q = 4\pi \sin \theta/\lambda$. Time frames were combined, excluding any data points affected by aggregation induced by radiation damage, to give the average scattering curve for each measurement (buffer was measured before and after every sample and averaged for background subtraction). Data were processed using the ATSAS package⁵⁶ and in-house software.

To analyze the experimental SAXS data of 627-NLS at the different temperatures, two different ensembles were created corresponding to the closed and open conformation of 627-NLS. For the "closed" ensemble, the crystal structure of avian 627-NLS (encompassing residues 541–741) was used and *flexible-meccano*^{22,23} was employed to add the flexible NLS peptide to the crystal structure assuming statistical coil distributions for these amino acids (residues 742–759). The "open" ensemble was generated as for the "closed" ensemble, but in addition the inter-domain linker between the 627 and the NLS domain (677–689) was also constructed using *flexible-meccano*, thereby obtaining an ensemble of structures where the 627 and the NLS domain are dislocated and can evolve freely within the limits of the conformational degrees of freedom allowed by the linker peptide. SAXS curves were predicted for each of the two ensembles ("open" and "closed") by averaging predicted SAXS curves for individual structures over the entire ensembles. The two predicted SAXS curves were then combined in a population-weighted average using the populations of the open and closed conformations extracted from the line shape analysis of the TROSY spectra or the CEST experiments at the different temperatures. These population-weighted averages were then compared to the experimental SAXS curves at the different temperatures (Figure 6b,c).

Small-Angle X-ray Scattering of Avian 627-NLS in Complex with Human Importin α 1. SAXS experiments of the complex of avian 627-NLS and human importin α 1 were obtained at 5 °C at the

ID14-3 beamline, ESRF, Grenoble, which operated at a fixed energy (13.32 keV, $\lambda = 0.931 \text{ \AA}$). All samples were purified as described above and checked for monodispersity by dynamic light scattering. Aliquots of 30 μL of protein solution at three different concentrations (1, 0.5, and 0.2 mg/mL) for each sample (and buffer) were exposed to X-rays, and scattering data were collected using the robotic sample handling available at the beamline. Ten individual frames were collected for every exposure, each 30 s in duration using the Pilatus 1M detector (Dectris) placed 1.83 m from the sample. Individual frames were processed automatically and independently using the EDNA processing pipeline using the ATSAS tools from EMBL-Hamburg, yielding individual radially averaged curves of normalized intensity versus scattering angle $q = 4\pi \sin \theta/\lambda$. Time frames were combined, excluding any data points affected by aggregation induced by radiation damage, to give the average scattering curve for each measurement (buffer was measured before and after every sample and averaged for background subtraction). Data were processed and analysis steps were performed using the ATSAS package.⁵⁶ Forty *ab initio* models were calculated using DAMMIF,⁵⁷ and then averaged and aligned using DAMAVER.⁵⁸

In order to analyze the experimental SAXS curve of the avian 627-NLS:human importin $\alpha 1$ complex in more detail, an ensemble of structures was generated on the basis of the crystal structure of the human NLS domain in complex with importin $\alpha 1$.¹⁹ *Flexible-meccano* was used to construct the linker peptide between the NLS and the 627 domain providing an ensemble of structures of the 627-NLS:importin complex where the 627 domain samples a large number of conformations as allowed by statistical coil sampling of the linker peptide (Figure 7a). Compared to the available crystal coordinates, the experimental construct comprised 13 amino acids at the N terminus and 34 amino acids at the C terminus that are predicted to be disordered and were built using *flexible-meccano* (residues 60–72 and 496–529, respectively). For each structure of the generated ensemble, the SAXS curve was predicted and compared to the experimental scattering curve (Figure 7a). In addition, sub-ensembles of structures (of sizes ranging from 5 to 25) that agree with the experimental SAXS curve were selected from the entire 25 000 orientations represented in Figure 7a, using the genetic algorithm ASTEROIDS.^{59–61}

Small-Angle Neutron Scattering of Avian 627-NLS in Complex with Human Importin $\alpha 1$. SANS experiments of the complex of avian 627-NLS and human importin $\alpha 1$ were obtained at 8 °C at the D33 beamline, Institut Laue-Langevin (ILL), Grenoble. The neutron beam had a wavelength $\lambda = 6 \text{ \AA} \pm 10\%$, and was passed through a 2 or 14.4 m rectangular collimator. Sample-to-detector distances were 2 and 14 m. The q range covered in this experiment extended from 0.004 to 0.048 nm^{-1} . The complex, at 2.45 mg/mL in 100% D_2O buffer, and the corresponding buffer were measured in 2 mm thick rectangular cuvettes.

The SANS data were used to validate the SAXS-derived structure/ensemble of the 627-NLS:importin $\alpha 1$ complex. Thus, SANS curves were calculated using CRYSON²⁵ from (1) the best-fitting structure of the complex (Figure 7b) and (2) the ensemble sampling all possible conformations of 627 on the surface of the NLS:importin complex (Figure 7a). The calculated SANS curves for the three cases were subsequently compared to the experimental SANS data (Figure S5).

Single-Molecule FRET of 627-NLS. Double-cysteine mutants of 627-NLS were randomly labeled with Alexa488- and Alexa594-maleimide in a 1:1 ratio, following previously described protocols.²⁹ Briefly, the protein was buffer exchanged into 50 mM Tris pH 7, 200 mM NaCl with 10 mM DTT, and transferred quickly into DTT-free buffer prior to labeling. Free cysteines were labeled with a 3–5-fold molar excess of dye, with unincorporated dye being removed by size exclusion chromatography. The buffer was exchanged into 50 mM Tris pH 6.5, 200 mM NaCl, 2 mM DTT for single-molecule measurements.

The smFRET measurements were performed at 627-NLS concentrations of 50 pM on a custom-built multi-parameter fluorescence spectrometer with a confocal detection geometry, as described previously.⁶² A laser diode (LDH 48S, Picoquant, Berlin, Germany) and a white light laser (SuperK Extreme, NKT Photonics,

Denmark) filtered through a 572/15 excitation filter were pulsed alternately with a repetition frequency of 27 MHz each to excite Alexa488 and Alexa594, respectively. Excitation light was linearly polarized. Fluorescence light was spatially filtered by a 100 μm pinhole, split into parallel and perpendicular polarized light, and then into Alexa488 (Do) and Alexa594 (Ac) fluorescence.

Fluorescence time traces were subjected to a burst recognition algorithm,⁶³ and all bursts with a Do+Ac intensity of more than 50 photons arising from excitation with the green laser were then further analyzed by multi-parameter analysis. FRET efficiencies (E_{FRET}) were calculated from fluorescence Do and Ac signals that were corrected for leakage and direct excitation according to

$$E_{\text{FRET}} = \frac{I_{\text{Ac}}}{I_{\text{Ac}} + \gamma I_{\text{Do}}} \quad (2)$$

with the correction for differences in detection efficiencies and quantum yields $\gamma = 1.14$. γ for 627-NLS was determined using established protocols.^{29,64} An independent validation of the measured γ can be obtained from considering the closed conformation of 627-NLS, which does not exhibit domain dynamics according to NMR and SAXS. For such a static molecule, γ can also be obtained from the following relationship:

$$E_{\text{FRET}} = \frac{I_{\text{Ac}}}{I_{\text{Ac}} + \gamma I_{\text{Do}}} = 1 - \frac{\tau_{\text{DA}}}{\tau_{\text{D}}} \quad (3)$$

with τ_{D} and τ_{DA} being the Do lifetimes in the absence and presence of an active Ac. Indeed, the closed conformation of 627-NLS follows this behavior with the γ used (Figure 9 and Figure S7).

Labeling stoichiometries (S) were calculated according to

$$S = \frac{I_{\text{Do}}}{I_{\text{Do}} + I_{\text{Ac}} + I_{\text{Ac}}^{\text{dir}}} \quad (4)$$

with $I_{\text{Ac}}^{\text{dir}}$ being the intensity of Ac arising from direct excitation with the Ac laser. In most cases (Figures S6 and S7, and all derived analyses), only bursts of $S = 0.2$ – 1.2 were selected for further analysis.²⁸

Fluorescence lifetimes were fit using a maximum likelihood estimator.⁶⁵ E and τ were then plotted in two-dimensional histograms and fit with double ($x = 1$) or triple ($x = 2$) Gaussian functions according to

$$f(E, \tau) = \sum_{n=0}^x A_n \exp \left[-\frac{1}{2(1 - c_n^2)} \times \left(\left(\frac{E - E_n}{\sigma_{E,n}} \right)^2 + \left(\frac{\tau - \tau_n}{\sigma_{\tau,n}} \right)^2 - \frac{2c_n(E - E_n)(\tau - \tau_n)}{\sigma_{E,n}\sigma_{\tau,n}} \right) \right] \quad (5)$$

with the widths of the populations σ , and the centers E_n and τ_n . The parameter c describes a tilt of the Gaussian fit in the τ – E plane and was set to 0 when $x = 2$ was selected, unless noted otherwise.

Single-molecule lifetime fits yielding $\tau < 500$ ps arose from background bursts, as could be shown by more restrictive burst recognition parameters, and were excluded from consideration when triple Gaussian fits were attempted.

Three different approaches were taken up to analyze the presence of slow dynamics within the 627-NLS FRET histograms:

- FRET peaks resulting from a titration of 627-NLS I539C A707C with importin $\alpha 1$ were first fit with two two-dimensional Gaussian functions, allowing a tilt in the τ versus E_{FRET} histogram ($x = 1$ and $c_n \neq 0$ in the fitting equation). The peak widths of the FRET populations (σ_1 in the fitting equation, corresponding to σ_E in Table S4) were then compared with the widths as expected from a shot-noise (sn)-limited FRET peak according to

$$\sigma_{\text{sn}} = \frac{\sqrt{\langle E_{\text{FRET}}^m \rangle (1 - \langle E_{\text{FRET}}^m \rangle) \langle N \rangle^{-1}}}{\langle \sqrt{\langle E_{\text{FRET}}^m \rangle (1 - \langle E_{\text{FRET}}^m \rangle) N_T}^{-1}} \quad (6)$$

where $\langle N \rangle$ is the average number of photons per burst and N_T the minimum number of photons per burst, corresponding to the threshold applied. $\langle E_{\text{FRET}}^m \rangle$ is the average FRET efficiency without γ correction.^{30,66} N_T was used for a conservative approximation of σ_{sn} . σ_1 that were significantly larger than σ_{sn} (as was the case throughout the whole importin $\alpha 1$ titration of 627-NLS, but particularly for 627-NLS alone and with low concentrations of the transport receptor) were indicative of slow dynamics of the dye-to-dye distance or the existence of multiple overlaid populations.^{30,66}

- (b) A tilt between E_{FRET} and τ within the FRET population of 627-NLS I539C A707C was necessary to reasonably fit the two-dimensional histograms with a double Gaussian ($x = 1$) population. However, for folded proteins with a clearly defined dye-to-dye distance, and for fast-fluctuating dye-to-dye distance distributions (much faster than inter-photon times, $\sim 50 \mu\text{s}$ in our case), no tilt would be expected within the same population in the E_{FRET} vs τ histogram.^{30,32,66}
- (c) Analyses considering different single-molecule observation times (binning analysis, Figure S6, shown for 0.5, 1, and 2 ms as well as resulting from burst recognition, i.e., up to ~ 5 ms) that yield similar peak widths and FRET positions are indicative of dynamics that are considerably slower than the time of a burst, or for two/several static molecules contributing to the observed FRET populations. Note that fast dynamics ($< \mu\text{s}$) are not covered by this analysis and can still be present.³²

Approaches (a)–(c) all point toward the existence of very slowly (> 5 ms) interconverting states, or static states contributing to the widths of the FRET histogram. The exchange rate between open and closed conformations in the unbound 627-NLS was estimated to be 50 s^{-1} at $25 \text{ }^\circ\text{C}$ by NMR. Fast interconversion rates within the open population of 627-NLS were found to be smaller than $\sim 10 \mu\text{s}$, whereas the closed population appears devoid of significant inter-domain fluctuations. All dynamics observed by NMR were therefore shorter than the inter-photon time ($\sim 50 \mu\text{s}$), or much longer than the observation time (up to ~ 5 ms), which suggests that a triple Gaussian function ($x = 2$, Figure S7c), without allowing tilt between τ and E_{FRET} within the same population ($c_i = 0$), should describe the two-dimensional τ versus E_{FRET} histograms reasonably well, which is indeed the case. However, in case small amounts of microsecond to millisecond time scale dynamics were present, we also allowed $c_i \neq 0$ in the fit. Under this scenario we determined FRET efficiencies for the open and closed forms, and both methods of fitting three Gaussian populations gave rise to the observation that the closed form was on the static E_{FRET} line, whereas the open form showed fast dynamics, leading to a deviation from the static FRET line (Figure S7d, Table S5).^{31,32} Additionally, both approaches yield a similar occupancy ratio between open and closed conformations to that found by NMR.

While determining distances for FRET populations that are dynamically averaged from a conformational ensemble requires the input of a specific model, the dye-to-dye distance R_{DA} for the closed, static, conformation could be calculated using

$$E_{\text{FRET}} = 1 / (1 + (R_{\text{DA}}/R_0)^6) \quad (7)$$

The Förster distance R_0 was determined using quantum yields and the overlap integral determined from singly labeled 627-NLS to be $58 \pm 1 \text{ \AA}$.²⁸

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b07765.

Tables S1–S3, chemical shift assignments; Tables S4 and S5, FRET peak widths, E_{FRET} , and τ for titration of 627-NLS with importin $\alpha 1$; Table S6, crystallographic data collection and refinement statistic of avian 627-NLS (538–759); Figures S1–S4, showing chemical shift differences between open and closed forms of 627-NLS, NMR spectra of 627, NLS, and 627-NLS, sequence comparison, NMR relaxation data, and all CEST data and fits; Figure S5, reproduction of SANS data; and Figures S6–S8, showing analysis of FRET data of 627-NLS alone and in complex with importins $\alpha 1$, $\alpha 3$, $\alpha 5$, and $\alpha 7$ (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*hart@embl.fr

*martin.blackledge@ibs.fr

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the Agence Nationale de Recherche under ComplexDynamics (SIMI7, M.B.). S.M. acknowledges funding from an EMBO long-term fellowship. N.S. acknowledges funding from a Swiss National Science Foundation Early Postdoc Mobility Fellowship (P2ELP2_148858). E.D. acknowledges funding from the Région Rhône-Alpes. D.B. and S.B. acknowledge funding from the FINOVI foundation. This work used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The authors thank Dr. Frank Gabel (IBS) for insightful discussions and the ESRF and ILL for the provision of beamtime for these experiments and the support of the staff. Parts of the NMR experiments were carried out at the large-scale NMR facility in Frankfurt, Germany, through access provided by the bio-NMR network. The authors thank Dr. Frank Löhr and Dr. Christian Richter for support in this context. E.L. acknowledges funding from the Emmy Noether program of the DFG.

■ REFERENCES

- Ruigrok, R. W. H.; Crépin, T.; Hart, D. J.; Cusack, S. *Curr. Opin. Struct. Biol.* **2010**, *20*, 104–113.
- Arranz, R.; Coloma, R.; Chichón, F. J.; Conesa, J. J.; Carrascosa, J. L.; Valpuesta, J. M.; Ortín, J.; Martín-Benito, J. *Science* **2012**, *338*, 1634–1637.
- Resa-Infante, P.; Jorba, N.; Coloma, R.; Ortín, J. *RNA Biol.* **2011**, *8*, 207–215.
- Moeller, A.; Kirchdoerfer, R. N.; Potter, C. S.; Carragher, B.; Wilson, I. A. *Science* **2012**, *338*, 1631–1634.
- Fodor, E. *Acta Virol.* **2013**, *57*, 113–122.
- Tarendeau, F.; Boudet, J.; Guilligay, D.; Mas, P. J.; Bougault, C. M.; Boulo, S.; Baudin, F.; Ruigrok, R. W. H.; Daigle, N.; Ellenberg, J.; Cusack, S.; Simorre, J.-P.; Hart, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14*, 229–233.
- Deng, T.; Engelhardt, O. G.; Thomas, B.; Akoulitchev, A. V.; Brownlee, G. G.; Fodor, E. *J. Virol.* **2006**, *80*, 11911–11919.
- Kuzuhara, T.; Kise, D.; Yoshida, H.; Horita, T.; Murazaki, Y.; Nishimura, A.; Echigo, N.; Utsunomiya, H.; Tsuge, H. *J. Biol. Chem.* **2009**, *284*, 6855–6860.
- Tarendeau, F.; Crépin, T.; Guilligay, D.; Ruigrok, R. W. H.; Cusack, S.; Hart, D. J. *PLoS Pathog.* **2008**, *4*, e1000136.

- (10) Pflug, A.; Guilligay, D.; Reich, S.; Cusack, S. *Nature* **2014**, *516*, 355–360.
- (11) Mehle, A.; Doudna, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 21312–21316.
- (12) Massin, P.; van der Werf, S.; Naffakh, N. *J. Virol.* **2001**, *75*, 5398–5404.
- (13) Gabriel, G.; Herwig, A.; Klenk, H.-D. *PLoS Pathog.* **2008**, *4*, e11.
- (14) Czudai-Matwich, V.; Otte, A.; Matrosovich, M.; Gabriel, G.; Klenk, H.-D. *J. Virol.* **2014**, *88*, 8735–8742.
- (15) Resa-Infante, P.; Jorba, N.; Zamarreño, N.; Fernández, Y.; Juárez, S.; Ortín, J. *PLoS One* **2008**, *3*, e3904.
- (16) Boivin, S.; Hart, D. J. *J. Biol. Chem.* **2011**, *286*, 10439–10448.
- (17) Yamada, S.; Hatta, M.; Staker, B. L.; Watanabe, S.; Imai, M.; Shinya, K.; Sakai-Tagawa, Y.; Ito, M.; Ozawa, M.; Watanabe, T.; Sakabe, S.; Li, C.; Kim, J. H.; Myler, P. J.; Phan, I.; Raymond, A.; Smith, E.; Stacy, R.; Nidom, C. A.; Lank, S. M.; Wiseman, R. W.; Bimber, B. N.; O'Connor, D. H.; Neumann, G.; Stewart, L. J.; Kawaoka, Y. *PLoS Pathog.* **2010**, *6*, e1001034.
- (18) Smith, E. R.; Begley, D. W.; Anderson, V.; Raymond, A. C.; Haffner, T. E.; Robinson, J. L.; Edwards, T. E.; Duncan, N.; Gerdt, C. J.; Mixon, M. B.; Nollert, P.; Staker, B. L.; Stewart, L. J. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2011**, *67*, 1015–1021.
- (19) Pumroy, R. A.; Ke, S.; Hart, D. J.; Zachariae, U.; Cingolani, G. *Structure* **2015**, *23*, 374–384.
- (20) Gabriel, G.; Fodor, E. *Curr. Top. Microbiol. Immunol.* **2014**, *385*, 35–60.
- (21) Vallurupalli, P.; Bouvignies, G.; Kay, L. E. *J. Am. Chem. Soc.* **2012**, *134*, 8148–8161.
- (22) Bernadó, P.; Blanchard, L.; Timmins, P.; Marion, D.; Ruigrok, R. W. H.; Blackledge, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 17002–17007.
- (23) Ozenne, V.; Bauer, F.; Salmon, L.; Huang, J.-R.; Jensen, M. R.; Segard, S.; Bernadó, P.; Charavay, C.; Blackledge, M. *Bioinformatics* **2012**, *28*, 1463–1470.
- (24) Svergun, D.; Barberato, C.; Koch, M. J. *J. Appl. Crystallogr.* **1995**, *28*, 768–773.
- (25) Svergun, D. I.; Richard, S.; Koch, M. H. J.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 2267–2272.
- (26) Kim, H. S.; Gabel, F. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2015**, *71*, 57–66.
- (27) Konarev, P. V.; Petoukhov, M. V.; Volkov, V. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2006**, *39*, 277–286.
- (28) Sisamakris, E.; Valeri, A.; Kalinin, S.; Rothwell, P. J.; Seidel, C. A. M. *Methods Enzymol.* **2010**, *475*, 455–514.
- (29) Milles, S.; Tyagi, S.; Banterle, N.; Koehler, C.; VanDelinder, V.; Plass, T.; Neal, A. P.; Lemke, E. A. *J. Am. Chem. Soc.* **2012**, *134*, 5187–5195.
- (30) Merchant, K. A.; Best, R. B.; Louis, J. M.; Gopich, I. V.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 1528–1533.
- (31) Kalinin, S.; Valeri, A.; Antonik, M.; Felekyan, S.; Seidel, C. A. M. *J. Phys. Chem. B* **2010**, *114*, 7983–7995.
- (32) Gopich, I. V.; Szabo, A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 7747–7752.
- (33) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späh, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 17800–17806.
- (34) Vogel, C.; Bashton, M.; Kerrison, N. D.; Chothia, C.; Teichmann, S. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 208–216.
- (35) Ma, B.; Tsai, C.-J.; Haliloglu, T.; Nussinov, R. *Structure* **2011**, *19*, 907–917.
- (36) Huang, J.-R.; Warner, L. R.; Sanchez, C.; Gabel, F.; Madl, T.; Mackereth, C. D.; Sattler, M.; Blackledge, M. *J. Am. Chem. Soc.* **2014**, *136*, 7068–7076.
- (37) Chang, C.-W.; Couñago, R. L. M.; Williams, S. J.; Bodén, M.; Kobe, B. *Plant Cell* **2012**, *24*, 5074–5088.
- (38) Kirui, J.; Bucci, M. D.; Poole, D. S.; Mehle, A. *J. Virol.* **2014**, *88*, 5977–5986.
- (39) Bernadó, P.; Blackledge, M. *Nature* **2010**, *468*, 1046–1048.
- (40) Mackereth, C. D.; Sattler, M. *Curr. Opin. Struct. Biol.* **2012**, *22*, 287–296.
- (41) Kabsch, W. *J. Appl. Crystallogr.* **1993**, *26*, 795–800.
- (42) Collaborative Computational Project, Number 4. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1994**, *50*, 760–763.
- (43) Read, R. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *57*, 1373–1382.
- (44) Perrakis, A.; Morris, R.; Lamzin, V. S. *Nat. Struct. Biol.* **1999**, *6*, 458–463.
- (45) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 240–255.
- (46) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187*, 163–169.
- (47) Solyom, Z.; Schwarten, M.; Geist, L.; Konrat, R.; Willbold, D.; Brutscher, B. *J. Biomol. NMR* **2013**, *55*, 311–321.
- (48) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
- (49) Goddard, T. D.; Kneller, D. G. *SPARKY*; University of California, San Francisco, 2003.
- (50) Jung, Y.-S.; Zweckstetter, M. *J. Biomol. NMR* **2004**, *30*, 11–23.
- (51) Lakomek, N.-A.; Ying, J.; Bax, A. *J. Biomol. NMR* **2012**, *53*, 209–221.
- (52) Kay, L. E.; Torchia, D. A.; Bax, A. *Biochemistry* **1989**, *28*, 8972–8979.
- (53) Wang, C.; Rance, M.; Palmer, A. G. *J. Am. Chem. Soc.* **2003**, *125*, 8968–8969.
- (54) Pernot, P.; Theveneau, P.; Giraud, T.; Fernandes, R. N.; Nurizzo, D.; Spruce, D.; Surr, J.; McSweeney, S.; Round, A.; Felisaz, F.; Foedinger, L.; Gobbo, A.; Huet, J.; Villard, C.; Cipriani, F. *J. Phys. Conf. Ser.* **2010**, *247*, 012009.
- (55) Round, A. R.; Franke, D.; Moritz, S.; Huchler, R.; Fritsche, M.; Malthan, D.; Klaering, R.; Svergun, D. I.; Roessle, M. *J. Appl. Crystallogr.* **2008**, *41*, 913–917.
- (56) Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A. G.; Gajda, M.; Gorba, C.; Mertens, H. D. T.; Konarev, P. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2012**, *45*, 342.
- (57) Franke, D.; Svergun, D. I. *J. Appl. Crystallogr.* **2009**, *42*, 342–350.
- (58) Volkov, V. V.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36*, 860–864.
- (59) Nodet, G.; Salmon, L.; Ozenne, V.; Meier, S.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2009**, *131*, 17908–17918.
- (60) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407–8418.
- (61) Jensen, M. R.; Zweckstetter, M.; Huang, J.; Blackledge, M. *Chem. Rev.* **2014**, *114*, 6632–6660.
- (62) Milles, S.; Lemke, E. A. *Biophys. J.* **2011**, *101*, 1710–1719.
- (63) Eggeling, C.; Berger, S.; Brand, L.; Fries, J. R.; Schaffer, J.; Volkmer, A.; Seidel, C. A. *J. Biotechnol.* **2001**, *86*, 163–180.
- (64) Ferreon, A. C. M.; Gambin, Y.; Lemke, E. A.; Deniz, A. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 5645–5650.
- (65) Schaffer, J.; Volkmer, A.; Eggeling, C.; Subramaniam, V.; Striker, G.; Seidel, C. A. M. *J. Phys. Chem. A* **1999**, *103*, 331–336.
- (66) Gopich, I.; Szabo, A. *J. Chem. Phys.* **2005**, *122*, 014707.
- (67) Marsh, J. A.; Singh, V. K.; Jia, Z.; Forman-Kay, J. D. *Protein Sci.* **2006**, *15*, 2795–2804.
- (68) Maltsev, A. S.; Ying, J.; Bax, A. *J. Biomol. NMR* **2012**, *54*, 181–191.

Supporting Information

Large Scale Conformational Dynamics Control H5N1 Influenza Polymerase PB2 Binding to Importin α

Authors:

Elise Delaforge¹, Sigrid Milles¹, Guillaume Bouvignies¹, Denis Bouvier^{1,2}, Stephane Boivin²,
Nicola Salvi¹, Damien Maurin¹, Anne Martel⁴, Adam Round³, Edward A. Lemke,⁵ Malene
Ringkjøbing Jensen¹, Darren Hart^{2, *}, Martin Blackledge^{1, *}

Communicating author:

martin.blackledge@ibs.fr, hart@embl.fr

Table S1: Chemical shifts (ppm) of avian 627-NLS obtained at 10°C in a buffer composed of 50 mM Tris, 200mM NaCl, pH 6.5. Chemical shifts are reported for both open and closed conformations (where applicable).

Res	aa	N	H ^N	C'	C α	C β	N	H ^N	C'	C α	C β
Closed form (or single conformation)						Open form					
537	G	-	-	176.9	42.8	-	-	-	-	-	-
538	E	121.0	8.73	176.9	55.7	29.7	-	-	-	-	-
539	I	123.1	8.55	176.1	60.8	37.6	-	-	-	-	-
540	N	123.3	8.68	175.2	52.6	38.6	-	-	175.2	52.6	38.6
541	G	109.8	8.44	172.8	45.0	-	109.7	8.43	172.8	45.0	-
542	P	-	-	177.9	64.9	31.6	-	-	178.0	64.8	31.6
543	E	117.3	8.97	178.3	60.2	27.4	117.2	9.03	178.3	60.4	27.3
544	S	114.8	7.92	176.4	60.7	62.1	-	-	176.4	-	-
545	V	123.3	8.03	180.3	66.0	30.5	123.0	8.02	180.5	66.0	30.5
546	L	123.9	8.16	178.0	59.3	41.3	123.9	8.18	178.0	59.2	41.3
547	V	118.4	8.21	178.2	67.0	31.2	118.8	8.28	178.1	67.0	31.1
548	N	116.7	9.06	179.2	55.8	37.6	116.7	8.98	179.2	55.9	37.6
549	T	117.9	8.30	175.6	67.4	68.2	117.7	8.27	-	67.4	68.2
550	Y	-	-	-	-	-	-	-	-	-	-
551	Q	-	-	178.5	-	-	-	-	-	-	-
552	W	120.6	8.23	179.1	-	-	-	-	-	-	-
553	I	-	-	-	-	-	-	-	-	-	-
554	I	-	-	180.3	-	-	-	-	180.2	-	-
555	R	119.9	8.66	178.1	58.4	29.4	119.9	8.61	178.2	58.6	29.3
556	N	115.7	7.39	174.7	52.6	37.7	115.2	7.38	174.7	52.6	37.7
557	W	123.4	7.42	175.4	59.6	30.3	123.9	7.44	175.4	59.6	30.3
558	E	114.1	8.58	179.3	59.2	28.0	114.1	8.60	179.3	59.2	28.0
559	T	110.5	7.40	-	-	-	110.5	7.35	177.0	64.1	68.2
560	V	-	-	177.0	-	-	-	-	-	-	-
561	K	119.0	8.58	178.8	59.3	29.7	-	-	-	-	-
562	I	-	-	178.7	-	-	-	-	-	-	-
563	Q	121.9	7.87	-	-	-	-	-	-	-	-
564	W	118.4	9.30	177.3	-	-	-	-	-	-	-
565	S	111.0	7.57	175.5	60.9	63.4	-	-	-	-	-
566	Q	118.0	8.38	176.5	57.0	29.7	-	-	176.5	-	-
567	D	118.4	8.85	172.5	51.9	39.7	118.1	8.84	172.6	51.9	39.9
568	P	-	-	177.0	64.2	30.5	-	-	176.9	-	30.3
569	T	109.6	8.06	176.5	66.0	67.7	109.8	8.02	176.5	65.9	67.7
570	M	119.3	6.96	178.7	56.9	30.9	119.3	6.93	178.8	56.9	30.9
571	L	121.0	6.91	178.3	-	40.3	121.0	6.98	178.3	56.9	40.3
572	Y	115.5	8.36	177.0	55.9	36.1	-	-	176.9	55.8	36.1
573	N	112.1	7.29	-	52.6	38.6	111.5	7.20	175.9	52.6	38.7
574	K	121.3	7.67	179.5	57.9	30.2	-	-	179.5	57.8	-
575	M	125.2	8.95	178.7	57.4	30.3	125.3	8.94	178.7	57.4	30.3
576	E	122.2	9.84	177.8	60.0	26.2	122.1	9.78	177.8	60.0	26.4
577	F	114.8	7.02	175.2	56.7	38.3	-	-	-	-	-
578	E	-	-	-	-	-	-	-	-	-	-
579	P	-	-	179.7	66.1	30.3	-	-	-	-	-
580	F	120.4	8.52	177.8	60.7	37.9	-	-	177.6	-	-
581	Q	116.2	8.01	178.5	58.7	28.5	115.7	7.93	178.5	58.6	28.5
582	S	112.8	7.58	175.5	60.9	62.6	112.8	7.54	175.5	60.9	62.6
583	L	120.3	7.98	176.7	54.5	40.3	-	-	-	-	-
584	V	-	-	-	-	-	-	-	-	-	-
585	P	-	-	178.7	63.2	30.9	-	-	-	-	-
586	K	127.2	8.67	178.5	60.1	31.7	-	-	-	-	-
587	A	117.8	8.80	178.4	53.3	17.5	-	-	178.4	53.2	17.6
588	A	118.9	7.52	177.1	51.3	19.0	118.7	7.49	177.2	51.3	18.9
589	R	116.7	7.11	179.1	60.5	29.6	-	-	179.0	60.5	29.8
590	G	108.6	8.70	178.2	46.6	-	108.4	8.69	177.9	46.6	-
591	Q	128.2	8.82	180.9	58.4	27.4	127.7	8.65	180.1	58.5	27.7
592	Y	120.2	9.04	177.9	56.8	35.7	119.6	8.96	178.0	56.9	35.6
593	S	113.0	7.74	-	59.7	63.3	-	-	-	-	-
594	G	-	-	-	-	-	-	-	-	-	-
595	F	-	-	-	-	-	-	-	-	-	-
596	V	-	-	178.4	62.3	27.5	-	-	-	-	-
597	R	118.6	9.36	180.5	54.7	32.2	-	-	-	-	-
598	T	119.3	7.91	175.6	66.7	67.7	-	-	-	-	-
599	L	126.6	7.74	-	-	-	-	-	-	-	-
600	F	-	-	177.4	55.5	35.6	-	-	177.7	56.0	-
601	Q	119.9	7.81	177.7	58.0	26.4	120.5	8.07	178.1	58.5	26.2
602	Q	119.9	8.24	176.9	59.4	26.9	119.8	8.14	178.4	58.8	27.0

603	M	119.1	7.76	177.7	58.2	33.2	-	-	-	-	-
604	R	116.3	7.94	179.1	58.9	28.9	-	-	-	-	-
605	D	117.5	8.20	176.9	55.5	41.0	117.7	8.00	-	-	-
606	V	123.4	8.01	174.6	61.3	29.0	-	-	-	-	-
607	L	127.8	8.09	178.6	56.1	40.6	-	-	-	-	-
608	G	110.1	9.02	176.2	45.4	-	-	-	-	-	-
609	T	108.6	8.16	173.6	60.1	68.8	110.8	8.08	-	-	-
610	F	125.6	8.17	175.8	59.8	38.8	125.1	8.29	-	-	-
611	D	130.7	6.85	173.7	52.1	42.3	128.2	6.40	-	-	-
612	T	-	-	174.5	67.0	69.6	-	-	-	-	-
613	V	119.0	7.73	176.5	66.2	29.9	-	-	-	-	-
614	Q	119.4	7.75	178.8	57.4	27.9	-	-	-	-	-
615	I	-	-	-	-	-	-	-	-	-	-
616	I	-	-	-	-	-	-	-	-	-	-
617	K	-	-	-	-	-	-	-	-	-	-
618	L	-	-	-	52.3	38.0	-	-	-	-	-
619	L	123.1	8.60	-	60.5	33.5	-	-	-	-	-
620	P	-	-	-	-	-	-	-	-	-	-
621	F	-	-	-	-	-	-	-	-	-	-
622	A	-	-	-	-	-	-	-	-	-	-
623	A	-	-	177.0	-	-	-	-	177.0	-	-
624	A	118.2	6.69	174.4	49.8	17.6	118.4	6.68	174.4	49.8	17.6
625	P	-	-	-	-	-	-	-	-	-	-
626	P	-	-	176.9	62.7	31.9	-	-	-	-	-
627	E	120.6	7.32	175.5	53.6	32.2	-	-	175.4	53.6	32.0
628	Q	123.7	8.91	175.6	55.5	28.4	122.3	8.68	176.1	55.7	27.7
629	S	116.7	8.04	173.5	57.7	64.7	119.4	8.56	174.8	57.9	64.3
630	R	119.1	8.78	177.0	54.9	29.4	122.6	8.86	176.3	55.5	28.9
631	M	119.7	7.74	176.1	55.5	31.9	119.2	7.90	175.0	54.7	32.2
632	Q	120.2	8.16	175.0	53.0	26.8	121.1	8.25	174.4	53.2	28.0
633	F	119.1	5.58	-	57.1	39.3	-	-	-	-	-
634	S	117.3	9.05	175.3	58.5	63.2	-	-	175.3	58.9	-
635	S	116.8	7.96	170.6	56.7	65.5	118.5	7.95	170.8	58.2	65.1
636	L	123.6	8.79	175.2	53.2	42.5	-	-	175.2	53.2	42.5
637	T	121.0	8.07	173.0	63.3	69.8	119.9	7.96	173.0	62.9	70.1
638	V	-	-	174.3	58.8	34.2	-	-	-	-	-
639	N	123.9	9.39	175.1	52.1	38.1	-	-	-	-	-
640	V	-	-	175.0	60.5	33.6	-	-	175.0	60.5	33.3
641	R	123.4	8.90	176.8	56.7	27.2	124.1	8.96	176.8	56.7	27.0
642	G	106.6	8.63	174.4	45.1	-	106.4	8.60	174.4	45.1	-
643	S	116.2	7.75	173.2	56.8	64.4	116.2	7.70	173.2	56.8	64.4
644	G	111.9	8.69	174.3	45.8	-	112.3	8.72	174.3	45.8	-
645	M	127.5	9.30	174.7	55.0	33.7	127.6	9.36	174.7	55.0	34.2
646	R	118.6	8.43	176.1	55.3	31.2	118.1	8.42	176.4	55.5	31.7
647	I	125.4	9.03	176.1	61.2	38.0	125.3	9.18	175.2	61.2	38.5
648	L	128.5	7.39	177.2	54.5	39.9	128.5	8.52	177.2	54.0	40.1
649	V	124.4	7.88	176.2	66.6	30.6	124.4	7.78	176.8	66.8	30.7
650	R	123.3	9.07	176.0	54.3	30.0	118.3	8.68	175.8	54.6	29.6
651	G	112.7	8.59	173.8	44.2	-	108.9	7.74	173.4	44.0	-
652	N	117.3	8.45	174.9	52.2	41.1	118.0	8.55	174.7	52.2	40.8
653	S	114.3	9.12	172.8	56.4	64.8	115.4	9.06	172.6	56.0	65.0
654	P	-	-	179.3	64.8	31.2	-	-	179.3	64.9	31.2
655	V	113.7	7.73	172.9	65.1	31.7	113.6	7.66	173.1	65.2	31.7
656	F	119.3	7.06	172.4	57.1	41.1	117.7	6.88	172.7	57.1	41.1
657	N	126.8	8.18	172.8	51.7	40.5	127.7	8.22	172.8	51.7	40.2
658	Y	127.3	8.78	173.5	56.2	39.4	127.0	8.76	173.5	56.4	39.4
659	N	128.6	7.47	174.3	51.0	38.7	128.9	7.51	174.4	51.0	38.8
660	K	124.5	8.41	177.3	58.1	31.7	124.5	8.44	177.5	58.1	31.7
661	A	120.1	8.08	179.8	53.9	17.8	-	-	179.8	-	-
662	T	107.2	7.34	175.3	60.9	69.2	107.1	7.36	-	60.8	69.3
663	K	117.3	8.06	175.1	56.7	27.7	117.2	8.06	175.0	57.1	27.8
664	R	114.8	7.14	176.0	54.3	31.0	-	-	-	-	-
665	L	127.5	8.75	174.3	52.8	43.4	127.5	8.79	174.3	52.8	43.4
666	T	122.2	8.97	173.5	60.4	69.7	121.7	8.95	173.4	60.3	69.7
667	I	127.6	9.06	175.9	56.2	37.4	127.7	9.08	175.8	56.2	37.3
668	L	126.5	8.79	177.1	54.9	38.8	126.3	8.58	177.2	54.9	38.6
669	G	102.7	8.44	173.9	44.9	-	102.6	8.43	173.9	44.9	-
670	K	123.0	8.24	175.6	53.5	32.3	122.8	8.22	175.6	53.5	32.3
671	D	124.1	8.74	175.8	55.4	41.1	-	-	175.9	55.3	41.1
672	A	129.1	9.28	176.0	50.6	18.7	129.0	9.19	175.9	50.7	18.8
673	G	107.0	7.85	171.1	43.3	-	106.8	7.89	170.4	43.2	-
674	A	119.8	8.09	177.5	50.2	20.0	119.5	8.10	177.5	49.9	20.5
675	L	122.1	8.45	176.1	54.0	42.3	122.0	8.44	176.8	54.1	41.7
676	T	117.9	8.46	173.8	60.7	69.6	-	-	-	-	-
677	E	124.1	8.76	175.8	56.5	29.1	-	-	175.7	55.3	30.2
678	D	126.2	8.93	-	52.3	42.9	124.0	8.71	175.2	52.2	40.4

679	P	-	-	177.1	64.3	31.1	-	-	-	-	-	-
680	D	118.1	8.99	175.5	53.7	39.8	-	-	-	-	-	-
681	E	121.6	7.73	177.6	57.1	28.9	-	-	-	177.3	-	-
682	G	114.8	9.35	175.7	46.1	-	110.8	8.72	-	-	45.2	-
683	T	113.8	8.57	175.7	61.6	69.7	-	-	-	-	-	-
684	A	126.1	8.69	178.6	52.7	18.2	-	-	-	178.6	52.7	18.2
685	G	108.3	8.35	174.8	45.3	-	108.8	8.52	-	173.1	45.3	-
686	V	118.9	7.29	176.2	60.6	32.2	119.4	8.01	-	175.0	-	-
687	E	124.2	8.92	178.2	54.0	29.6	123.5	8.66	-	176.5	-	-
688	S	122.2	9.48	172.8	57.4	64.3	116.0	8.39	-	-	-	-
689	A	125.7	8.85	173.8	49.5	17.6	-	-	-	179.5	-	-
690	V	115.2	7.45	173.7	62.8	34.2	118.2	8.01	-	177.9	64.0	31.2
691	L	116.4	7.73	176.9	51.4	45.7	-	-	-	177.1	-	-
692	R	125.9	10.19	177.7	56.8	29.3	121.2	7.81	-	178.2	58.1	28.9
693	G	118.6	10.42	173.5	44.7	-	-	-	-	173.4	44.6	-
694	F	116.1	7.37	173.4	55.9	43.4	119.5	8.20	-	173.9	57.1	41.1
695	L	117.5	8.75	176.7	52.4	43.5	122.2	9.17	-	176.3	53.1	43.4
696	I	121.7	8.19	175.9	61.4	36.5	124.3	8.91	-	176.9	61.4	36.0
697	L	126.3	9.41	177.1	54.8	41.9	129.4	9.46	-	-	55.0	41.4
698	G	105.1	7.34	172.2	44.4	-	105.4	7.39	-	172.2	44.4	-
699	K	120.5	8.28	174.2	56.0	32.5	120.6	8.28	-	174.2	-	-
700	E	118.2	7.46	175.7	56.1	29.9	-	-	-	175.7	56.1	29.9
701	D	125.0	8.68	178.5	51.9	42.7	125.0	8.63	-	178.5	51.9	42.7
702	K	127.6	9.41	178.1	58.6	31.0	127.8	9.38	-	178.3	58.5	31.0
703	R	121.1	8.46	178.0	58.2	28.6	121.2	8.49	-	178.0	58.2	28.6
704	Y	114.4	7.83	176.6	58.0	36.5	114.6	7.79	-	176.4	57.9	36.4
705	G	107.9	7.38	171.8	45.1	-	-	-	-	-	-	-
706	P	-	-	176.5	62.3	31.5	-	-	-	176.5	62.3	31.4
707	A	124.0	8.58	177.6	51.8	18.5	124.0	8.63	-	177.6	51.8	18.3
708	L	125.2	8.99	177.1	53.2	42.8	125.3	8.98	-	177.1	53.2	42.8
709	S	115.2	8.68	176.6	56.7	64.4	114.9	8.61	-	176.6	56.7	64.5
710	I	120.9	8.72	178.2	63.6	36.6	121.6	8.79	-	178.3	63.6	36.6
711	N	117.3	8.24	176.8	54.7	37.5	117.3	8.29	-	177.2	55.0	37.5
712	E	119.3	7.74	177.4	56.4	29.9	118.9	7.77	-	178.1	56.8	30.0
713	L	119.0	7.55	179.1	56.2	39.7	119.7	7.62	-	179.1	56.4	39.9
714	S	112.8	7.75	174.9	60.5	62.2	113.2	7.77	-	175.1	60.6	62.1
715	N	117.5	7.83	175.9	53.1	38.2	117.4	7.65	-	175.8	53.0	38.4
716	L	121.2	7.34	176.6	55.2	41.8	121.7	7.39	-	176.7	55.3	41.9
717	A	126.7	9.10	178.1	50.3	18.9	127.1	9.01	-	178.1	50.4	19.0
718	K	119.5	8.53	178.0	58.8	31.7	120.3	8.54	-	178.0	58.5	31.3
719	G	113.0	7.98	173.0	45.1	-	113.3	8.48	-	173.0	45.0	-
720	E	121.7	8.02	175.2	56.1	30.3	121.5	7.88	-	175.1	56.4	30.4
721	K	121.0	8.08	175.9	54.1	34.4	121.6	8.17	-	175.9	54.2	34.2
722	A	120.2	8.46	175.2	51.1	22.8	120.5	8.57	-	175.2	51.2	22.7
723	N	119.7	9.16	175.3	52.4	39.9	119.7	9.11	-	175.2	52.4	39.9
724	V	121.2	9.74	173.8	59.1	35.2	121.4	9.65	-	172.8	59.1	34.6
725	L	129.2	8.35	175.2	53.7	40.4	129.2	8.19	-	175.2	53.7	40.4
726	I	128.0	8.40	175.3	59.0	39.7	123.4	8.62	-	-	-	-
727	G	110.6	7.97	173.5	44.2	-	-	-	-	174.2	45.1	-
728	Q	120.7	8.86	177.0	56.6	26.8	121.7	8.86	-	176.8	56.0	26.9
729	G	111.0	8.58	173.7	45.4	-	110.0	8.42	-	173.6	45.4	-
730	D	122.1	7.91	173.8	53.5	41.1	122.1	7.89	-	173.8	53.6	41.1
731	V	124.3	8.43	176.1	60.6	32.2	123.2	8.31	-	176.1	60.6	32.2
732	V	119.3	8.71	173.2	57.6	33.8	120.4	8.63	-	173.0	-	34.2
733	L	-	-	176.5	53.9	42.4	-	-	-	176.5	53.9	42.4
734	V	121.0	9.37	174.0	57.7	34.7	119.7	9.24	-	174.3	57.8	35.5
735	M	117.1	8.64	175.8	53.1	37.3	115.7	8.57	-	175.5	53.4	37.2
736	K	128.5	8.64	176.6	56.0	30.9	127.7	8.97	-	176.7	57.0	31.1
737	R	127.8	7.90	176.7	57.3	29.6	126.7	7.94	-	176.3	56.9	30.1
738	K	121.2	8.73	176.1	57.1	32.2	122.3	8.58	-	176.1	56.6	32.4
739	R	119.2	7.67	174.7	54.7	31.1	121.7	8.22	-	175.8	55.4	30.2
740	D	120.8	8.43	177.0	53.4	41.5	-	-	-	-	-	-
741	S	118.9	8.50	173.9	58.7	63.2	117.5	8.44	-	174.8	-	-
742	S	118.8	8.65	174.5	57.6	63.4	118.9	8.55	-	174.6	58.4	63.1
743	I	120.4	8.17	176.4	60.5	37.7	122.5	8.04	-	176.5	60.9	37.4
744	L	125.4	8.28	177.9	54.8	41.1	126.2	8.34	-	177.9	54.8	41.0
745	T	116.7	8.45	174.7	61.4	69.3	114.9	8.16	-	174.7	61.4	69.3
746	D	120.1	8.57	176.9	54.3	40.4	123.1	8.46	-	176.9	54.3	40.4
747	S	116.7	8.39	175.4	58.7	62.8	-	-	-	175.3	58.6	62.8
748	Q	122.3	8.47	176.9	56.0	28.4	122.3	8.46	-	176.9	56.0	28.4
749	T	115.3	8.15	174.9	62.2	69.0	-	-	-	-	-	-
750	A	126.4	8.31	178.4	52.7	18.1	-	-	-	-	-	-
751	T	114.1	8.11	174.9	62.1	69.1	-	-	-	-	-	-
752	K	124.3	8.25	176.7	56.2	31.8	-	-	-	-	-	-
753	R	122.8	8.31	176.6	55.9	29.7	-	-	-	-	-	-
754	I	123.2	8.24	176.3	60.8	37.6	-	-	-	-	-	-

755	R	125.6	8.43	176.2	55.7	29.7	-	-	-	-	-
756	M	122.7	8.43	175.9	54.8	32.1	-	-	-	-	-
757	A	126.0	8.34	177.5	51.9	18.3	-	-	-	-	-
758	I	120.5	8.19	175.4	60.7	37.8	-	-	-	-	-
759	N	127.8	8.05	179.6	54.4	39.8	-	-	-	-	-

Table S2: Chemical shifts (ppm) of avian 627 obtained at 25°C in a buffer composed of 50 mM Tris, 200mM NaCl, pH 6.5.

Res	aa	N	H ^N	C'	C α	C β
537	G	-	-	-	-	-
538	E	-	-	-	-	-
539	I	-	-	175.8	61.3	34.3
540	N	122.9	8.59	175.0	53.2	39.4
541	G	109.2	8.35	172.6	45.0	-
542	P	-	-	177.8	65.3	32.6
543	E	116.9	9.13	178.2	61.0	28.3
544	S	114.4	7.88	176.3	61.4	62.5
545	V	122.6	7.97	180.2	66.5	31.7
546	L	123.4	8.15	177.8	59.8	42.5
547	V	118.3	8.23	177.8	67.5	32.3
548	N	116.5	8.96	179.0	56.5	38.3
549	T	117.4	8.26	175.4	67.9	68.7
550	Y	118.9	7.66	178.3	63.9	37.7
551	Q	117.7	9.13	178.3	60.0	29.4
552	W	120.3	8.24	178.9	63.5	28.3
553	I	-	-	177.0	66.6	-
554	I	116.0	8.05	179.9	65.6	38.3
555	R	119.5	8.63	177.9	59.0	30.4
556	N	115.1	7.37	174.9	53.2	38.6
557	W	123.6	7.45	175.4	60.2	31.1
558	E	114.0	8.56	178.8	60.0	29.0
559	T	110.6	7.34	176.7	64.7	68.6
560	V	125.3	8.36	177.0	66.6	31.9
561	K	118.6	8.50	178.5	59.9	30.9
562	I	118.8	6.66	178.5	63.8	37.9
563	Q	121.4	7.91	178.1	59.5	29.2
564	W	118.3	9.28	177.0	62.5	28.5
565	S	110.5	7.56	175.4	61.4	63.8
566	Q	117.8	8.32	176.3	57.8	30.8
567	D	117.5	8.78	172.5	52.6	40.5
568	P	-	-	176.8	64.6	31.3
569	T	109.7	7.94	176.3	66.4	67.9
570	M	118.8	6.97	178.5	57.5	32.0
571	L	120.4	7.03	178.1	57.4	41.7
572	Y	115.3	8.25	176.6	56.5	36.9
573	N	111.5	7.20	175.6	53.1	39.3
574	K	120.9	7.61	179.2	58.2	31.4
575	M	125.1	8.88	178.5	58.1	31.4
576	E	121.8	9.76	177.6	60.5	-
577	F	114.4	7.02	175.2	56.8	39.1
578	E	-	-	-	-	-
579	P	-	-	179.5	66.6	-
580	F	119.8	8.37	177.4	61.3	38.9
581	Q	115.4	7.93	178.3	59.2	29.4
582	S	112.6	7.54	175.4	61.4	63.1
583	L	-	-	-	-	-
584	V	-	-	-	-	-
585	P	-	-	178.5	63.6	32.1
586	K	126.8	8.62	178.3	60.7	32.8
587	A	117.5	8.74	178.2	54.0	18.6
588	A	118.2	7.46	177.1	52.0	20.2
589	R	116.7	7.16	178.8	61.1	30.8
590	G	108.0	8.68	177.7	46.9	-
591	Q	127.2	8.53	179.8	59.2	28.8
592	Y	119.3	8.93	177.8	57.5	36.6
593	S	112.5	7.75	177.7	60.5	64.0
594	G	105.5	8.68	176.6	47.8	-
595	F	122.7	8.25	177.3	61.1	39.5
596	V	119.9	8.60	177.1	66.9	31.5
597	R	117.6	9.28	180.9	61.2	29.7
598	T	118.8	7.89	175.6	67.2	68.1
599	L	126.2	7.80	180.5	58.7	41.8
600	F	117.0	8.95	177.4	56.6	36.7
601	Q	120.2	8.06	178.0	59.2	27.4
602	Q	119.5	8.15	178.4	59.4	28.2
603	M	118.0	7.97	177.6	59.1	34.1
604	R	117.0	7.94	178.2	59.1	30.6
605	D	117.3	8.03	176.9	56.2	42.0
606	V	121.2	8.17	176.1	62.7	32.0

607	L	125.7	8.49	178.5	57.0	41.9
608	G	109.5	8.62	174.9	45.5	-
609	T	110.7	8.04	174.1	61.6	70.0
610	F	124.4	8.25	174.4	59.7	40.1
611	D	127.3	6.43	173.9	53.0	42.6
612	T	122.7	8.46	-	67.3	69.7
613	V	-	-	176.6	66.9	-
614	Q	117.9	7.56	178.1	59.1	30.4
615	I	117.8	7.72	174.1	65.4	37.4
616	I	118.2	7.67	175.6	67.0	38.2
617	K	113.0	7.86	178.2	59.4	33.6
618	L	111.4	7.83	178.1	54.9	42.1
619	L	122.5	8.51	174.7	60.4	40.3
620	P	-	-	-	-	-
621	F	-	-	-	-	-
622	A	-	-	176.1	52.7	19.0
623	A	119.2	6.58	177.2	53.8	20.6
624	A	118.2	6.73	174.3	50.8	18.7
625	P	-	-	-	-	-
626	P	-	-	177.1	63.2	-
627	E	120.3	7.31	175.4	54.3	32.9
628	Q	121.9	8.62	175.9	56.4	28.7
629	S	118.9	8.48	174.7	58.3	64.8
630	R	122.4	8.81	176.0	56.1	30.1
631	M	118.9	7.88	174.8	55.3	33.4
632	Q	120.5	8.15	174.2	54.1	29.1
633	F	119.3	5.27	-	58.0	40.8
634	S	-	-	-	-	-
635	S	-	-	170.7	57.3	65.7
636	L	123.2	8.79	175.1	53.9	43.6
637	T	119.6	7.95	172.9	63.4	70.4
638	V	114.3	8.38	174.3	59.4	35.3
639	N	123.8	9.35	174.9	52.9	38.9
640	V	-	-	175.0	61.2	34.3
641	R	123.6	8.91	176.6	57.3	28.3
642	G	106.0	8.52	174.3	45.3	-
643	S	115.8	7.71	173.2	57.4	64.9
644	G	112.1	8.67	174.1	46.4	-
645	M	126.9	9.31	174.6	55.5	35.4
646	R	117.9	8.40	176.3	56.1	33.0
647	I	124.8	9.14	175.2	61.7	39.7
648	L	128.4	8.54	177.0	54.6	41.4
649	V	124.1	7.78	176.6	67.5	31.8
650	R	117.8	8.65	175.6	55.2	30.7
651	G	108.5	7.68	173.2	44.0	-
652	N	117.7	8.50	174.5	52.8	41.5
653	S	115.2	9.04	172.4	56.5	65.6
654	P	-	-	179.0	-	-
655	V	113.1	7.64	172.7	65.5	32.8
656	F	117.4	6.89	172.7	57.7	42.0
657	N	127.5	8.23	172.8	52.4	41.0
658	Y	126.7	8.69	173.4	57.0	40.0
659	N	128.5	7.54	174.3	51.8	39.4
660	K	124.2	8.40	177.2	58.7	32.8
661	A	119.9	8.07	179.5	54.6	19.0
662	T	106.3	7.30	175.1	61.2	69.6
663	K	116.7	8.05	174.7	57.7	29.2
664	R	114.6	7.13	175.8	55.3	32.1
665	L	127.0	8.74	174.3	53.5	44.6
666	T	121.2	8.90	173.3	60.8	70.1
667	I	127.2	9.04	175.6	56.9	38.7
668	L	125.9	8.58	176.9	55.6	39.8
669	G	102.3	8.41	173.9	45.4	-
670	K	122.3	8.22	175.5	54.3	33.4
671	D	123.8	8.65	175.8	55.9	41.8
672	A	128.8	9.14	175.8	51.6	20.0
673	G	106.3	7.90	171.0	43.8	-
674	A	119.2	8.03	177.4	50.8	21.6
675	L	121.7	8.42	176.6	54.8	42.8
676	T	115.9	8.32	174.1	61.9	69.4
677	E	122.5	8.26	175.5	56.0	31.3
678	D	123.4	8.59	174.9	52.9	41.2
679	P	-	-	175.7	63.5	32.3
680	D	120.0	8.50	176.5	54.7	41.3
681	E	122.2	8.34	177.1	57.0	30.5
682	G	110.4	8.64	174.6	45.3	-

683	T	113.6	8.07	174.6	61.9	69.6
684	A	126.5	8.37	178.2	53.1	19.4
685	G	108.4	8.38	174.3	45.1	-
686	V	119.1	7.95	176.5	62.4	32.9
687	E	124.4	8.60	176.6	57.0	30.2
688	S	116.8	8.29	174.4	58.6	63.6
689	A	126.2	8.33	177.7	53.0	19.4
690	V	118.9	7.98	176.1	62.5	32.8
691	L	125.9	8.22	177.1	55.3	42.2
692	R	122.6	8.30	175.7	56.1	31.1
693	G	116.2	8.00	179.0	45.6	-

Table S3: Chemical shifts (ppm) of avian NLS obtained at 25°C in a buffer composed of 50 mM Tris, 200mM NaCl, pH 6.5.

Res	aa	N	H ^N	C'	C α	C β
677	G	117.9	8.48	-	-	-
678	D	122.8	8.18	-	-	-
679	P	-	-	176.8	63.6	32.2
680	D	119.9	8.41	176.5	54.5	41.3
681	E	121.6	8.29	177.3	57.2	30.3
682	G	109.6	8.49	174.7	45.7	-
683	T	113.6	8.03	174.6	62.1	69.8
684	A	126.3	8.35	178.2	53.0	19.2
685	G	107.8	8.32	174.5	45.5	-
686	V	119.4	8.01	176.5	63.0	32.9
687	E	123.3	8.62	176.9	57.7	30.0
688	S	115.8	8.33	175.6	59.2	63.6
689	A	126.3	8.32	179.1	54.1	18.9
690	V	117.2	7.91	177.1	64.4	32.5
691	L	120.3	7.96	176.8	54.7	41.8
692	R	121.1	7.80	177.9	58.6	30.0
693	G	108.6	8.69	173.3	45.2	-
694	F	118.9	8.16	173.9	57.4	42.1
695	L	121.9	9.13	176.2	53.5	44.7
696	I	124.0	8.88	176.7	62.0	37.2
697	L	129.0	9.42	177.5	55.5	42.6
698	G	105.2	7.42	172.2	45.0	-
699	K	120.2	8.23	174.0	56.1	33.3
700	E	117.9	7.39	175.5	56.7	31.1
701	D	124.7	8.59	178.4	52.4	43.6
702	K	127.4	9.30	178.0	59.0	32.1
703	R	120.8	8.51	177.8	58.8	29.8
704	Y	114.3	7.77	176.3	58.5	37.4
705	G	107.6	7.39	-	45.5	-
706	P	-	-	176.4	62.8	32.4
707	A	123.6	8.53	177.4	52.4	19.3
708	L	124.9	8.96	176.9	53.8	43.8
709	S	114.6	8.56	176.5	57.3	65.3
710	I	121.4	8.75	178.0	64.2	37.9
711	N	116.8	8.21	177.0	55.4	38.2
712	E	118.5	7.76	177.8	57.2	31.1
713	L	119.2	7.57	178.8	56.9	41.1
714	S	112.8	7.76	174.9	61.2	62.9
715	N	117.1	7.66	175.6	53.4	39.2
716	L	121.2	7.37	176.5	55.7	43.2
717	A	127.0	8.98	177.9	50.9	20.0
718	K	119.8	8.47	177.7	59.0	32.5
719	G	112.8	8.46	173.0	45.6	-
720	E	121.0	7.87	175.1	56.7	31.4
721	K	121.0	8.10	175.8	54.7	35.5
722	A	120.1	8.54	175.2	51.7	23.8
723	N	119.6	9.07	175.0	52.8	40.6
724	V	121.1	9.64	172.8	59.6	35.8
725	L	128.9	8.20	175.7	54.4	42.7
726	I	123.2	8.60	176.1	60.5	39.9
727	G	110.0	8.33	174.0	45.5	-
728	Q	121.1	8.78	176.7	56.7	27.9
729	G	109.8	8.41	173.6	45.9	-
730	D	121.8	7.85	173.6	54.0	42.0
731	V	122.9	8.20	176.1	61.1	33.8
732	V	120.0	8.64	172.9	58.6	35.2
733	L	124.3	8.26	176.4	54.3	43.5
734	V	119.5	9.21	174.2	58.2	36.5
735	M	115.7	8.59	175.5	53.9	38.3
736	K	127.5	8.88	176.4	57.4	32.4
737	R	126.7	7.97	176.0	57.4	31.4
738	K	121.8	8.51	176.1	56.9	33.6
739	R	121.1	8.16	175.6	55.9	31.3
740	D	122.7	8.45	176.3	54.3	41.5
741	S	116.7	8.33	174.7	58.6	63.8
742	S	118.3	8.48	174.4	58.9	63.9
743	I	121.9	8.00	176.3	61.4	38.6
744	L	125.7	8.27	177.6	55.3	42.3
745	T	114.3	8.07	174.5	61.9	69.8
746	D	122.7	8.36	176.7	54.7	41.3

747	S	116.1	8.31	175.2	59.1	63.7
748	Q	121.8	8.42	176.6	56.6	29.4
749	T	114.5	8.08	174.7	62.6	69.7
750	A	126.0	8.25	178.2	53.2	19.2
751	T	113.2	8.03	174.7	62.5	69.7
752	K	123.6	8.18	176.5	56.7	33.0
753	R	122.2	8.25	176.3	56.4	30.9
754	I	122.3	8.14	176.2	61.4	38.8
755	R	124.9	8.36	176.0	56.2	30.9
756	M	122.1	8.34	175.6	55.4	33.2
757	A	125.5	8.26	177.3	52.4	19.4
758	I	119.8	8.10	175.2	61.2	39.0
759	N	127.4	7.98	179.4	54.8	40.6

Table S4: FRET peak widths for titration of 627-NLS with importin $\alpha 1$.

Importin $\alpha 1$ (μM)	$\langle E_{FRET}^m \rangle$	σ_E	σ_{sn}
0	0.73	0.132	0.062
0.002	0.67	0.123	0.066
0.005	0.67	0.135	0.067
0.01	0.65	0.139	0.068
0.05	0.66	0.144	0.067
0.1	0.62	0.105	0.069
0.2	0.62	0.096	0.069
0.3	0.62	0.099	0.069
0.6	0.61	0.095	0.069
1	0.61	0.108	0.069
2	0.61	0.097	0.069

The width expected for a FRET peak delimited by shot noise is

$$\sigma_{sn} = \sqrt{\langle E_{FRET}^m \rangle (1 - \langle E_{FRET}^m \rangle) \langle N^{-1} \rangle} < \sqrt{\langle E_{FRET}^m \rangle (1 - \langle E_{FRET}^m \rangle) N_T^{-1}},$$

With $\langle N^{-1} \rangle$ being the mean inverse photon count per burst, and N_T the number of photons determined by the threshold over Do and Ac photons as used for burst selection. For a conservative approximation, we used N_T to determine an upper value for $\sigma_{sn}^{1,2}$.

σ_E is the width of the Gaussian FRET distribution obtained from $\langle E_{FRET}^m \rangle$, the FRET efficiencies that were not corrected for γ using a double Gaussian fit and allowing a tilt in the fit within the population ($c_n \neq 0$).

Table S5: E_{FRET} and τ during 627-NLS titration with importin $\alpha 1$ as obtained from triple Gaussian fits (**Supplementary Fig. S7c**).

Importin $\alpha 1$ (μM)	E_1	τ_1 (ns)	E_2	τ_2 (ns)
0	0.62	1.92	0.77	0.99
0.002	0.58	2.12	0.72	1.12
0.005	0.58	2.04	0.75	1.10
0.01	0.57	2.22	0.75	1.05
0.05	0.58	2.15	0.77	0.89
0.1	0.57	2.25	0.72	1.37
0.2	0.56	2.22	0.71	1.35
0.3	0.58	2.19	0.74	1.21
0.6	0.57	2.18	0.71	1.16
1	0.53	2.39	0.65	1.62
2	0.58	2.16	0.75	0.84
average	0.57	2.33	0.73	1.15
SD	0.02	0.54	0.03	0.23

All E_{FRET} versus τ histograms (Supplementary Fig. S7c) were fit with triple Gaussian functions, not allowing a tilt within the population ($c_n = 0$ in the fitting equation). Starting values for the fit were the same at all importin $\alpha 1$ concentrations. E_0 and τ_0 describing the Do-only populations were centered around 0 and 4 ns, respectively. In the static case, the FRET efficiency can be calculated with the following equation

$$E_{\text{FRET}} + \frac{\tau_{\text{DA}}}{\tau_{\text{D}}} = 1$$

with the Do lifetimes τ_{DA} and τ_{D} in the presence (τ_1 or τ_2 in the table) and absence ($\tau_0 \approx 4$ ns) of an active acceptor, respectively. For fast conformational averaging ($<$ inter photon time) however, the following relationship holds:³

$$E_{\text{FRET}} + \frac{\tau_{\text{DA}}}{\tau_{\text{D}}} > 1$$

Since we do not constrain the position of the FRET peaks, E_1 and E_2 scatter between the different titration points. Nevertheless, the closed conformation of 627-NLS fulfills this relationship static relationship in all cases ($E_{\text{FRET}} + \frac{\tau_{\text{DA}}}{\tau_{\text{D}}} = 1.01 \pm 0.03$ on average). For the open

conformation, we obtain $E_{FRET} + \frac{\tau_{DA}}{\tau_D} = 1.12 \pm 0.01$ on average for all different importin $\alpha 1$ concentrations, in agreement with fast averaging dynamics.

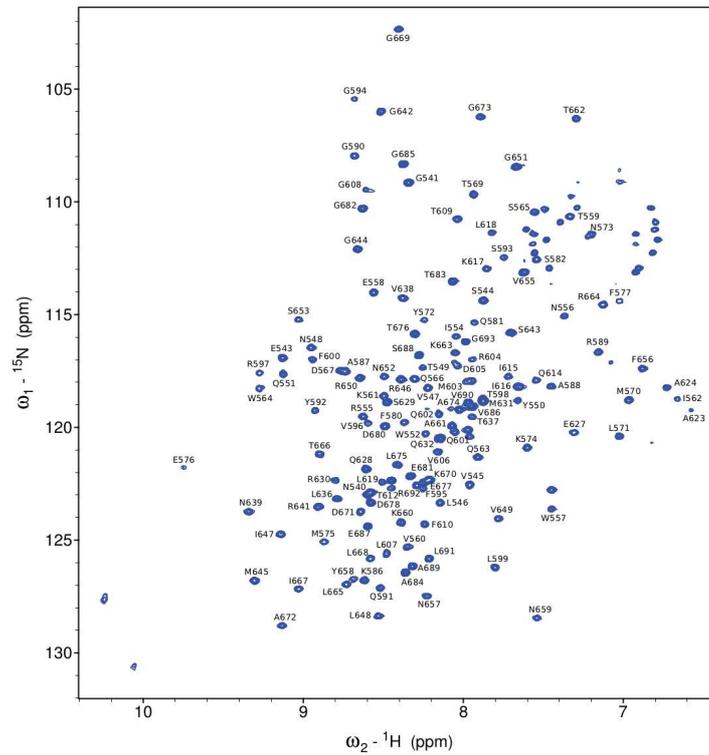
Note that the FRET efficiencies in the unbound form are slightly larger than in all forms where importin $\alpha 1$ was present. While we cannot easily determine whether this comes from slightly different photophysical properties in the absence and presence of the transport receptor, or whether it has a structural meaning, it does not affect our conclusions. The distance of the closed conformation was determined from the unbound state, for which also all correction parameters were measured.

Table S6. Crystallographic data collection and refinement statistic of avian 627-NLS (538-759)

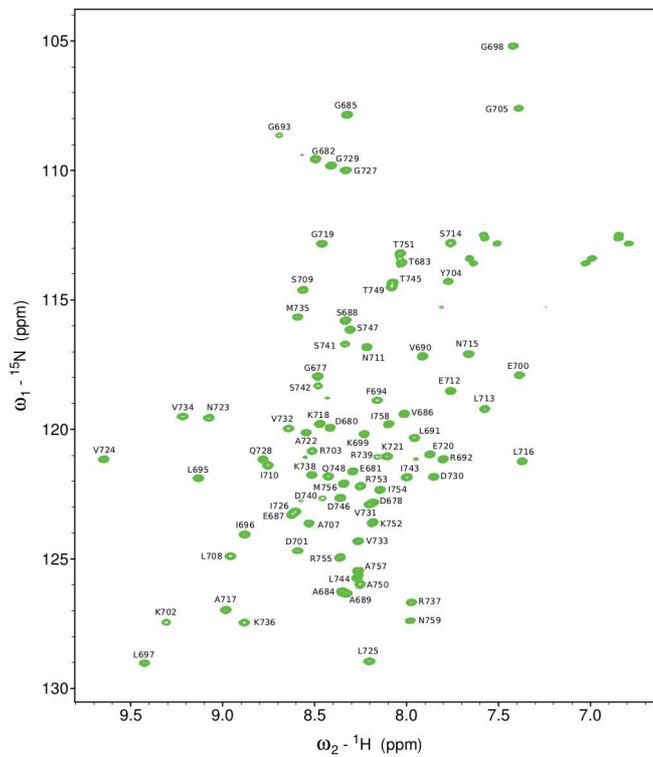
	Av 627-NLS domain (native)
Data collection	
Beamline (ESRF)	ID14-4
Wavelength (Å)	0.939
Space group	<i>P</i> 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	41.2, 58.4, 83.5
α , β , γ (°)	90.0, 102.7, 90.0
Resolution (Å)	15-1.43 (1.54-1.48)
R _{merge}	10.9 (50.8)
<i>I</i> / σ	17.56 (3.60)
Completeness (%)	97.9 (99.2)
Redundancy	3.98 (3.86)
Refinement	
Resolution (Å)	1.5
Total No. reflections/free	57170
R _{work}	0.171
R _{free}	0.200
No. atoms	3558
<i>Protein</i>	3272
<i>Non-protein</i>	35
<i>Metal</i>	1
<i>Water</i>	251
No. residues modeled with multiple conformations	chain A: 6 chain B: 4
Average B-factors (Å ²)	14.6
R.m.s. deviations	
<i>Bond lengths</i> (Å)	0.010
<i>Bond angles</i> (°)	1.307
Ramachandran Plot	
<i>Most Favoured</i> (%)	99.5
<i>Favoured</i> (%)	0.5
Values in parentheses are for highest-resolution shell. Two molecules in the asymmetric unit.	

Figure S1: NMR spectra of 627, NLS and 627-NLS

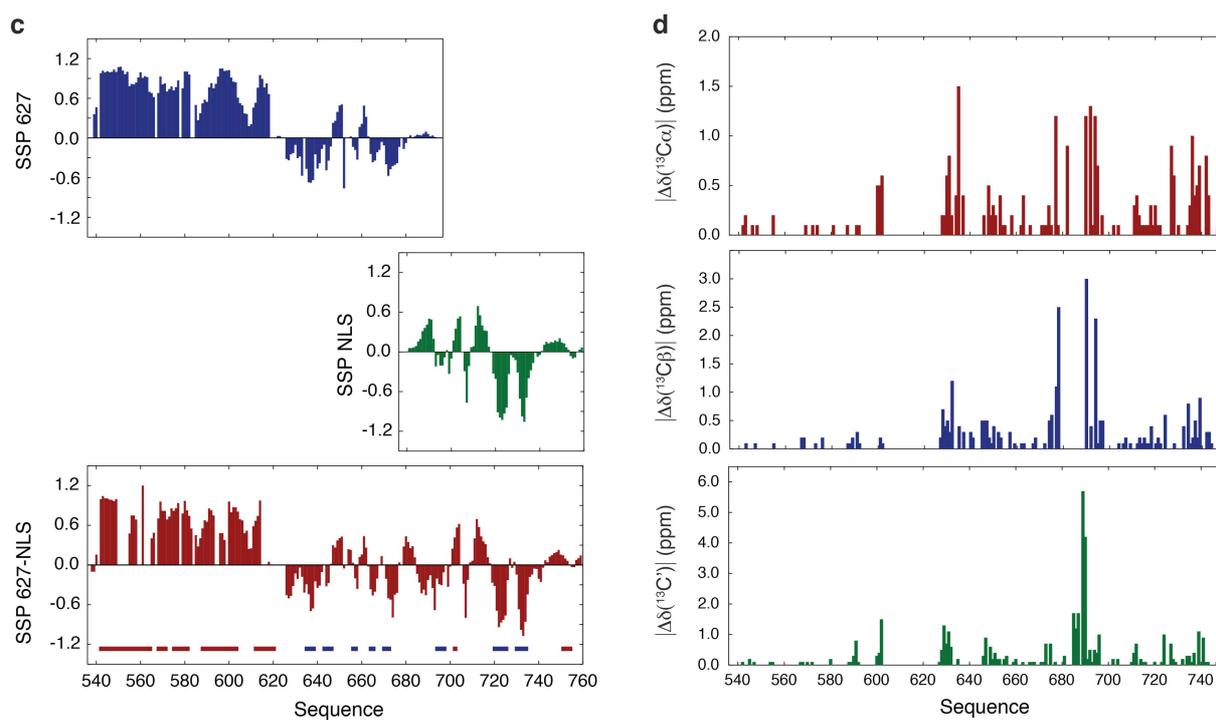
S1a) ^1H - ^{15}N HSQC spectrum of avian 627 obtained at 25°C showing the assignment of the resonances.



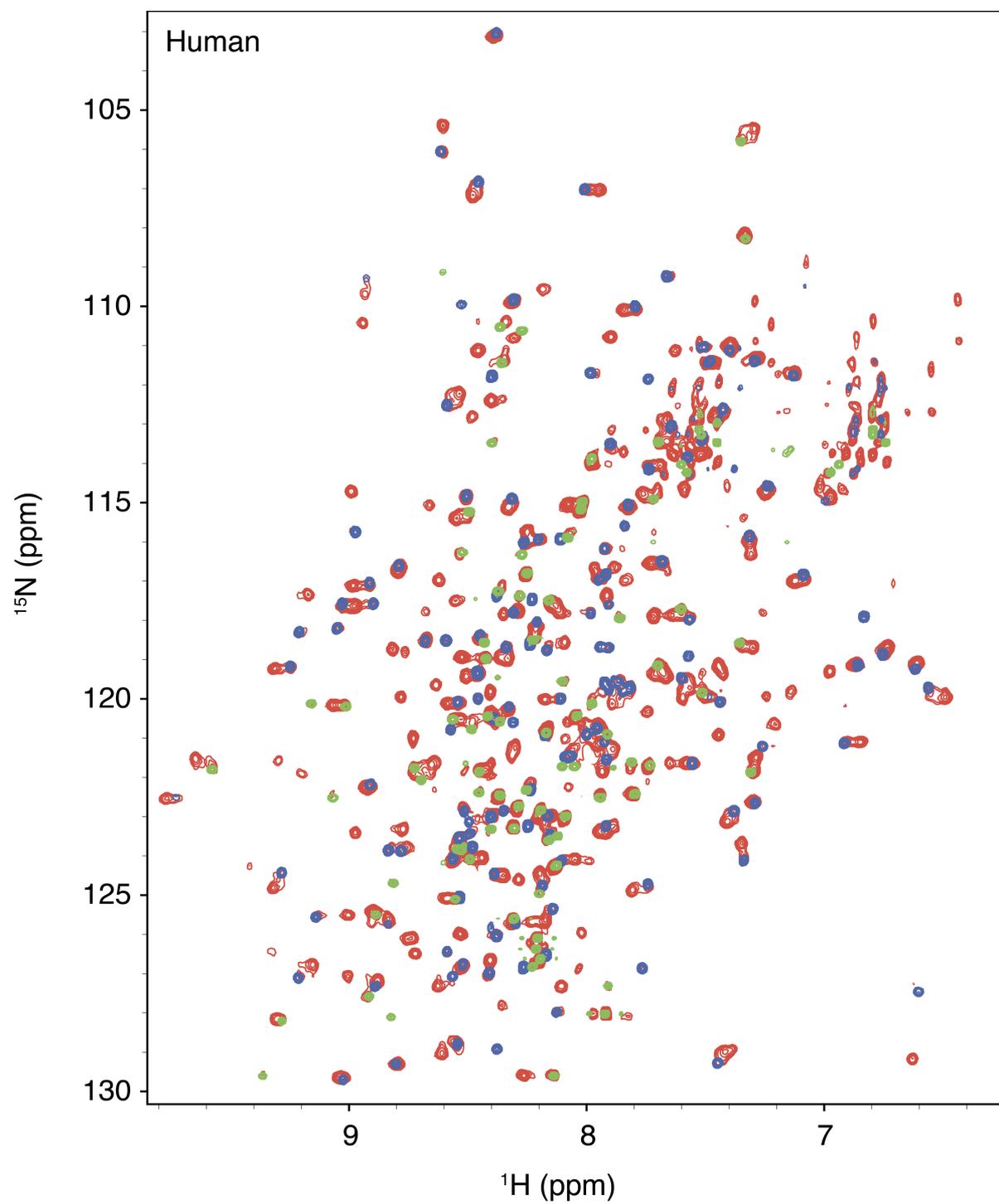
S1b) ^1H - ^{15}N HSQC spectrum of avian NLS obtained at 25°C showing the assignment of the resonances.



S1c, S1d) Analysis of experimental chemical shifts. **(c)** Secondary structure propensities (SSP) of 627 (blue, 25°C), NLS (green, 25°C) and 627-NLS (red, 10°C) calculated from experimental $^{13}\text{C}\alpha$ and $^{13}\text{C}\beta$ chemical shifts.³ The chemical shifts of 627 were obtained using a deuterated sample, while the chemical shifts of 627-NLS were obtained using a perdeuterated sample. Corrections were made for isotope effects on experimental chemical shifts of perdeuterated 627-NLS using the corrections recently published by Maltsev et al.⁴ In the case of 627-NLS the SSP score is reported for the set of resonances corresponding to the closed form of 627-NLS or, if applicable, the resonances where the open and closed forms are indistinguishable. Horizontal bars in the bottom panel indicate the position of helices (red bars) and sheets (blue bars) as observed in the crystal structure of avian 627-NLS. **(d)** ^{13}C chemical shift differences between the resonances corresponding to open and closed conformations as function of the sequence of 627-NLS: $^{13}\text{C}\alpha$ (red), $^{13}\text{C}\beta$ (blue) and $^{13}\text{C}'$ (green).



S1e) ^1H - ^{15}N TROSY spectrum of human 627-NLS (red) obtained at 25°C superimposed with the spectra of the isolated 627 (blue) and NLS (green) domains. Resonances corresponding to open and closed conformations of the 627-NLS domain are also visible for the human form.



S1f) ^{15}N chemical shift differences between the resonances corresponding to open and closed conformations as function of the sequence of 627-NLS. Figure shows the same points as figure 3 in the main manuscript, showing observed chemical shift differences between open and closed, and additionally the chemical shift differences (grey bars) for residues where HSQC correlation peaks of the open form is not present, due to excessive broadening, but whose chemical shift difference could be determined from CEST. It can be seen that the presence of two peaks is almost universal throughout the protein (including both 627 and NLS domains as well as the linker).

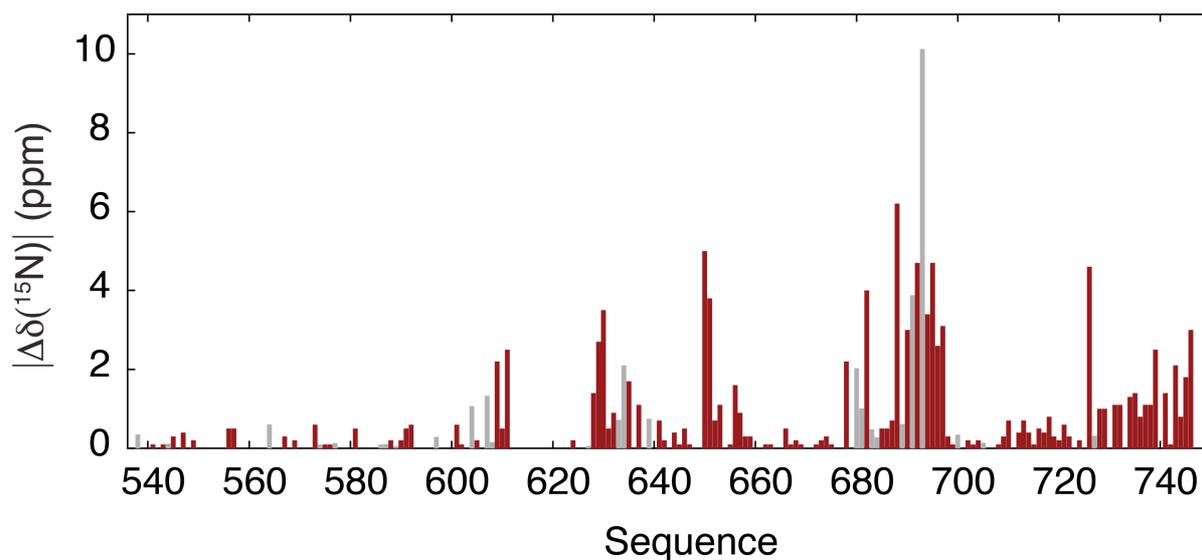


Figure S2: Sequence alignment of avian, human and bat 627-NLS domains from influenza A. The residue numbering corresponds to the numbering of avian 627-NLS.

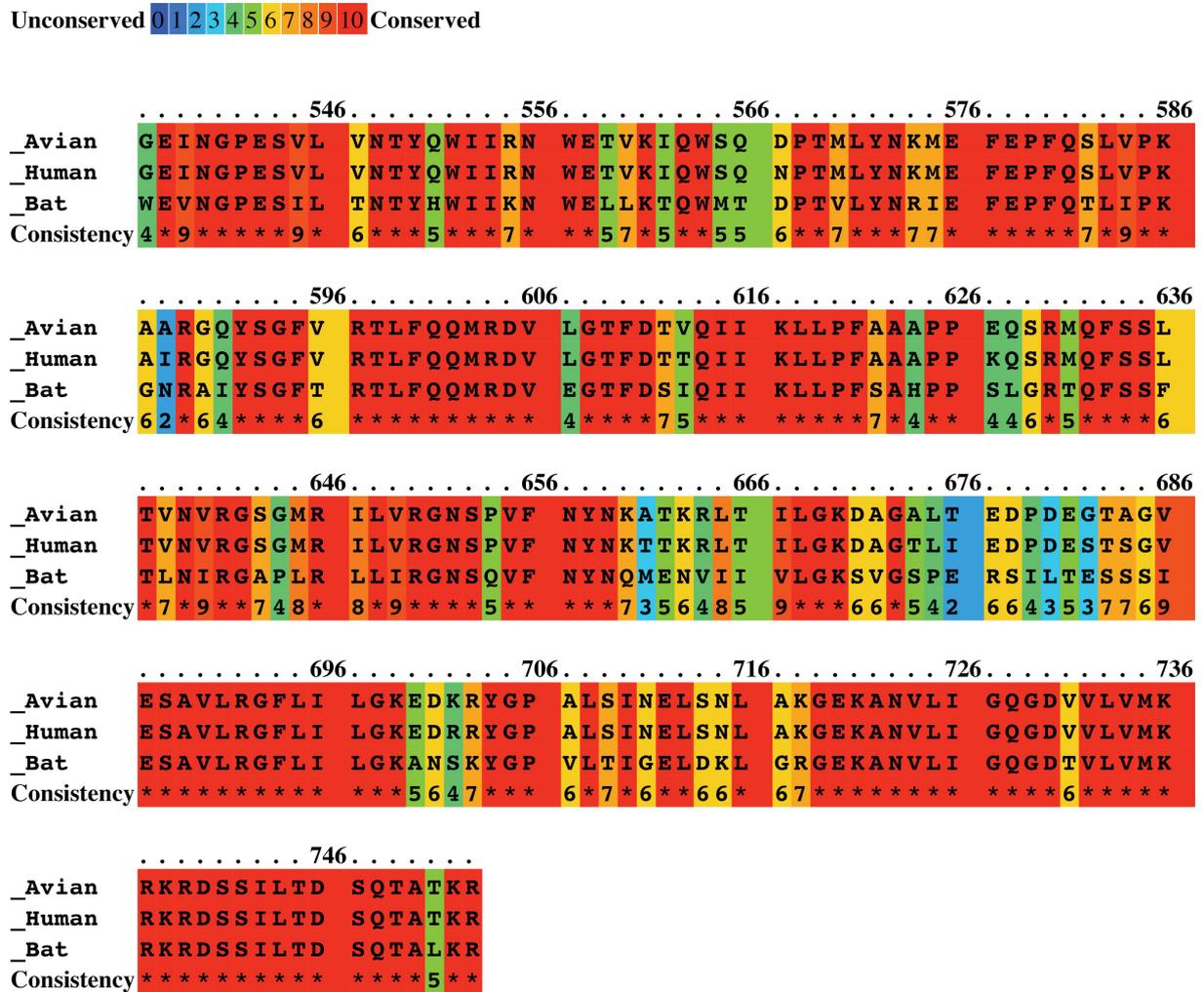


Figure S3: ^{15}N relaxation rates of avian 627-NLS R650A obtained at a ^1H frequency of 600 MHz and 10°C . Top : Transverse backbone ^{15}N relaxation. Middle: Transverse cross-correlated ^{15}N dipole-CSA relaxation. Bottom: Heteronuclear ^1H - ^{15}N nuclear Overhauser effect (ratio of saturated and reference spectra shown as a function of sequence).

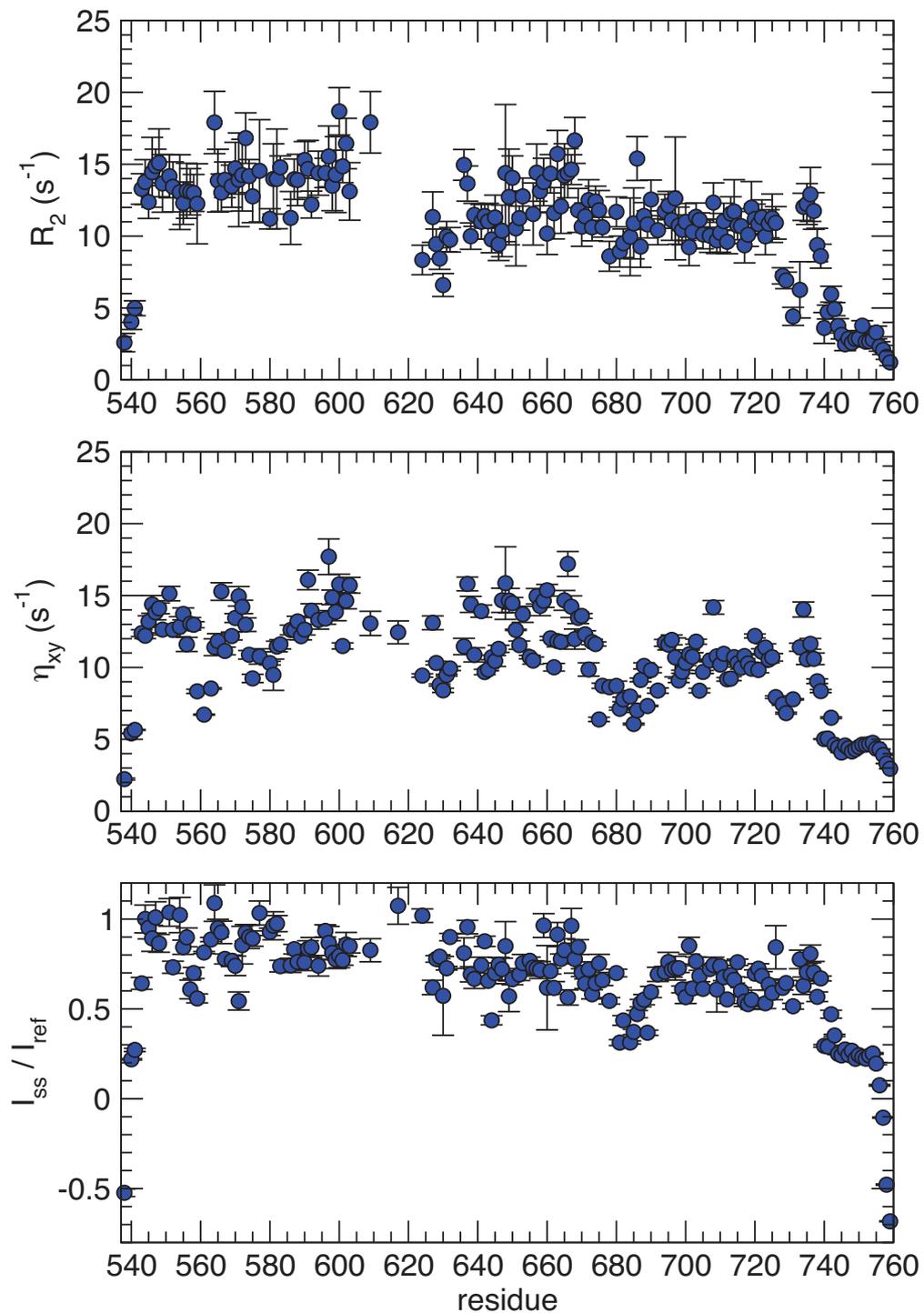
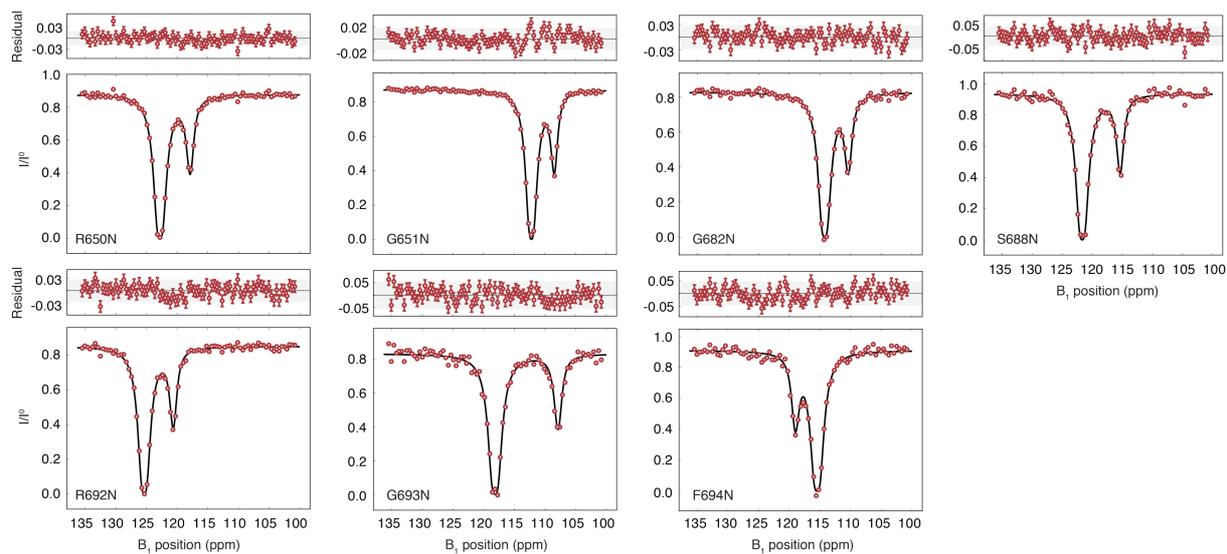


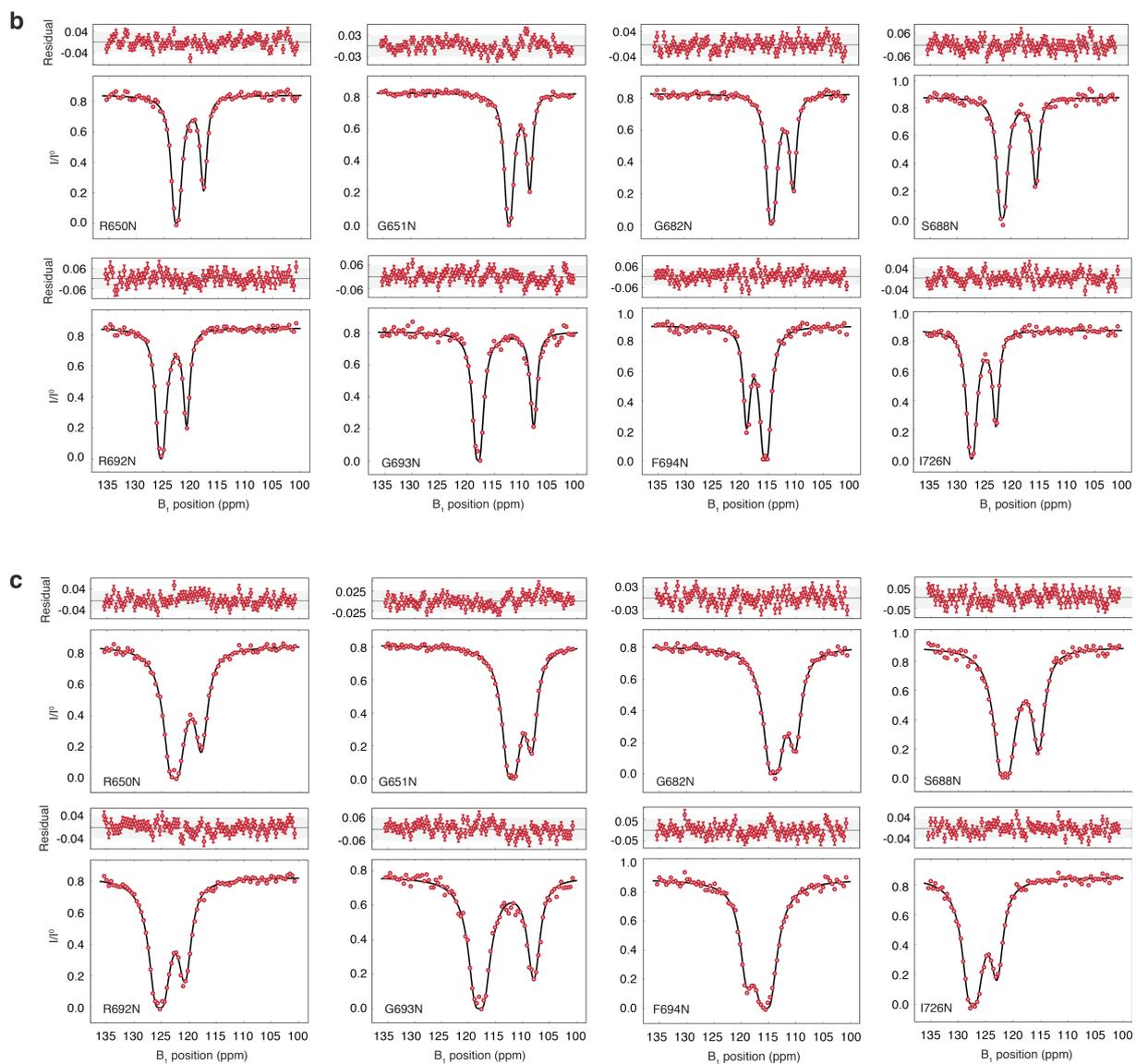
Figure S4: CEST experiments applied to 627-NLS

a) Experimental CEST profiles (red) obtained of selected residues in 627-NLS at 5°C and a B_1 saturating field of 22 Hz. The black lines correspond to a fit to a two-site exchange model by analyzing simultaneously data for selected residues at 5°C, 10°C, 15°C and 20°C assuming an Eyring relationship (see online methods).

S4a

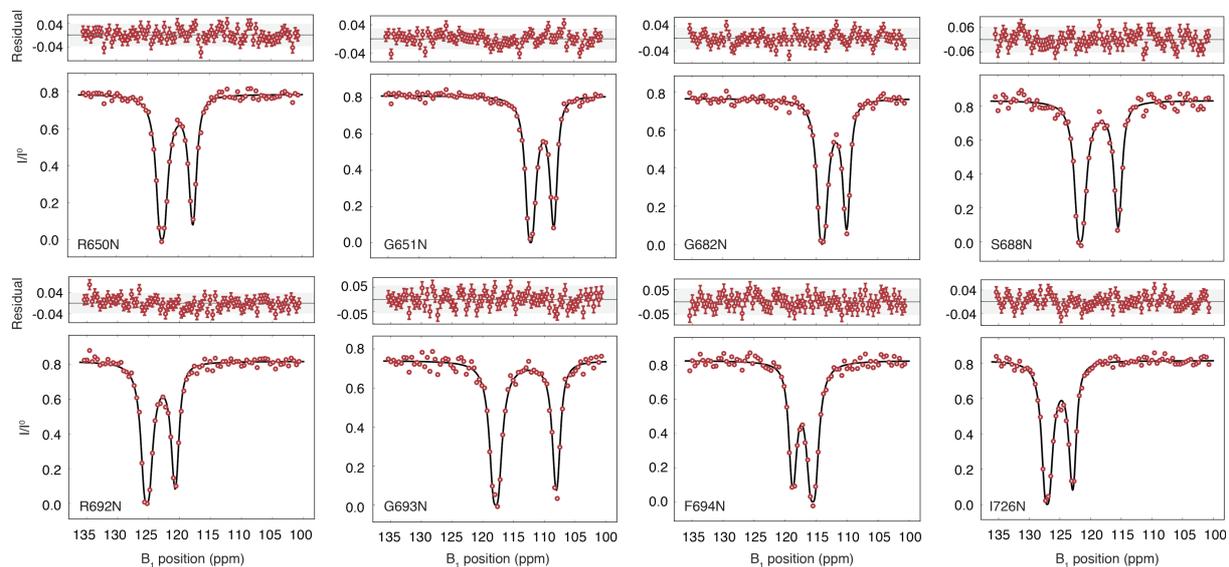


S4b, S4c) Experimental CEST profiles (red) obtained of selected residues in 627-NLS at 10°C and a B_1 saturating field of 22 Hz (B) and 44 Hz (C). The black lines correspond to a fit to a two-site exchange model by analyzing simultaneously data for selected residues at 5°C, 10°C, 15°C and 20°C assuming an Eyring relationship (see online methods).



S4d, S4e) Experimental CEST profiles (red) obtained of selected residues in 627-NLS at 15°C (D) and 20°C (E) and a B_1 saturating field of 22 Hz. The black lines correspond to a fit to a two-site exchange model by analyzing simultaneously data for selected residues at 5°C, 10°C, 15°C and 20°C assuming an Eyring relationship (see online methods).

S4d)



S4e)

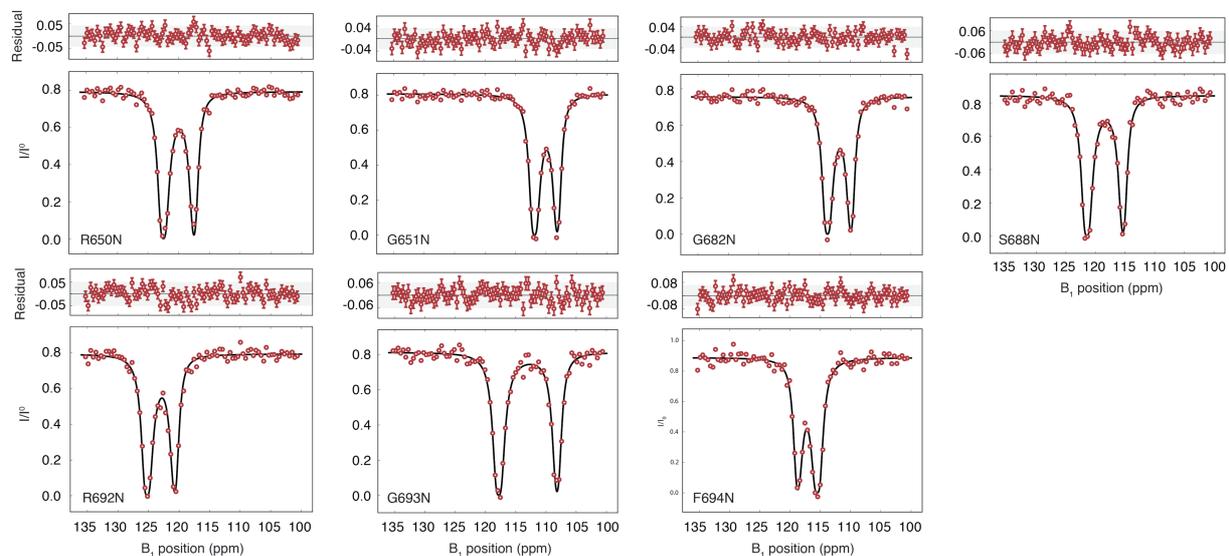


Figure S5:

Reproduction of SANS data. Blue: experimental data points measured on protonated proteins in 100% D₂O. Green: Average SANS profile calculated over 25000 conformers sampling the space available to the 627 domain when the NLS domain is bound to importin α 1 in the conformation found in the crystal structure (pdb code 4UAF). Red: comparison of SANS curves calculated from ensembles selected on the basis of SAXS data (see methods) using the ASTEROIDS algorithm.

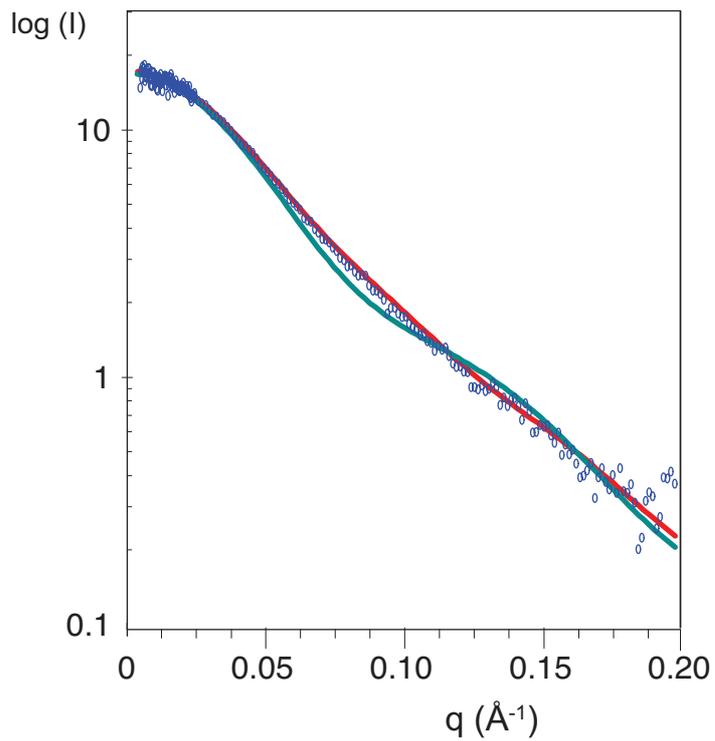


Figure S6: Single molecule FRET of 627-NLS:Importin α interaction

(a) Donor fluorescence lifetime (τ) versus FRET efficiency (E_{FRET}) histograms of 627-NLS I539C A707C double cysteine mutant at various importin α 1 concentrations with corresponding two-dimensional double-Gaussian fits (black dashed curves). Tilts in the fit between the E_{FRET} and τ axis within one population was allowed ($c_n \neq 0$). (b) τ versus E_{FRET} histogram of 627-NLS R650A I539C A707C double cysteine mutant. Fit as described in (a).

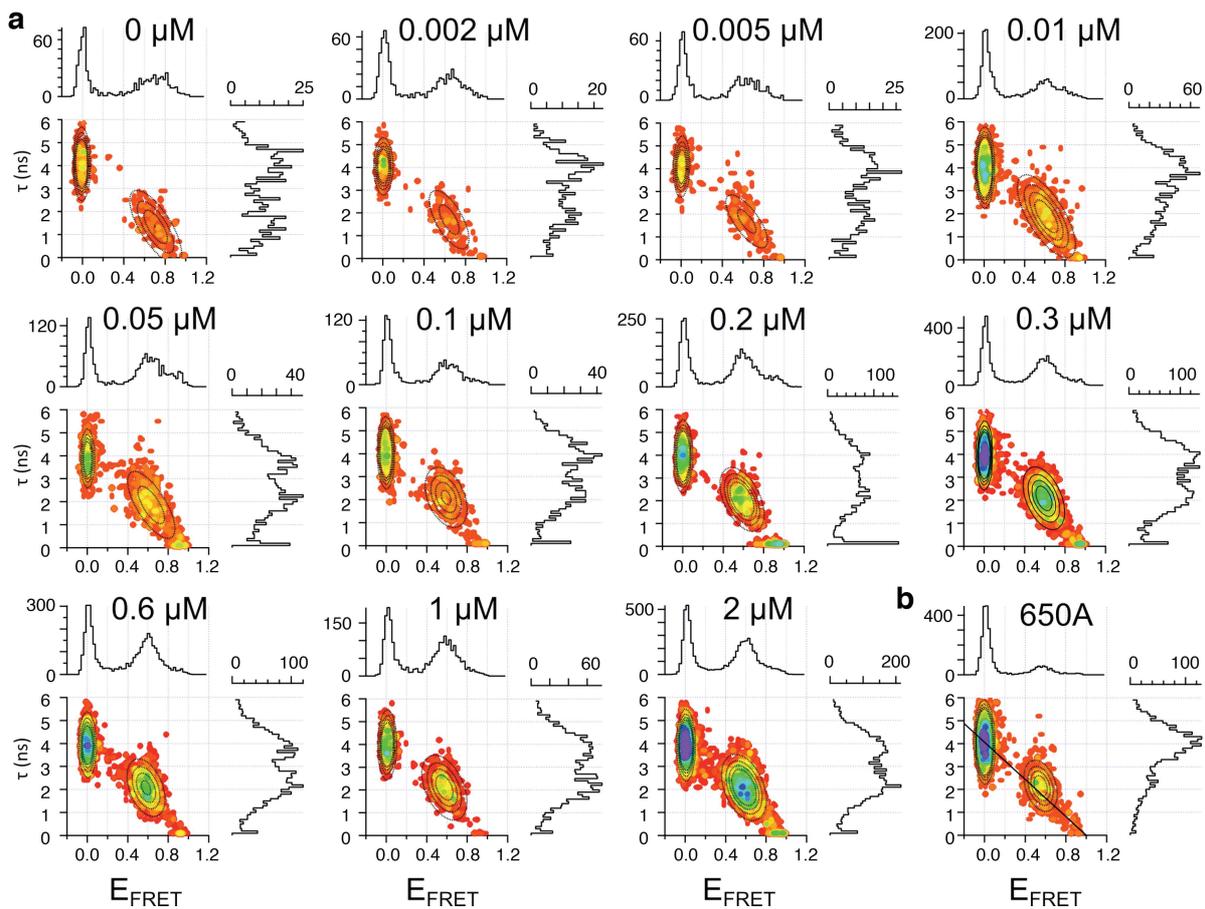
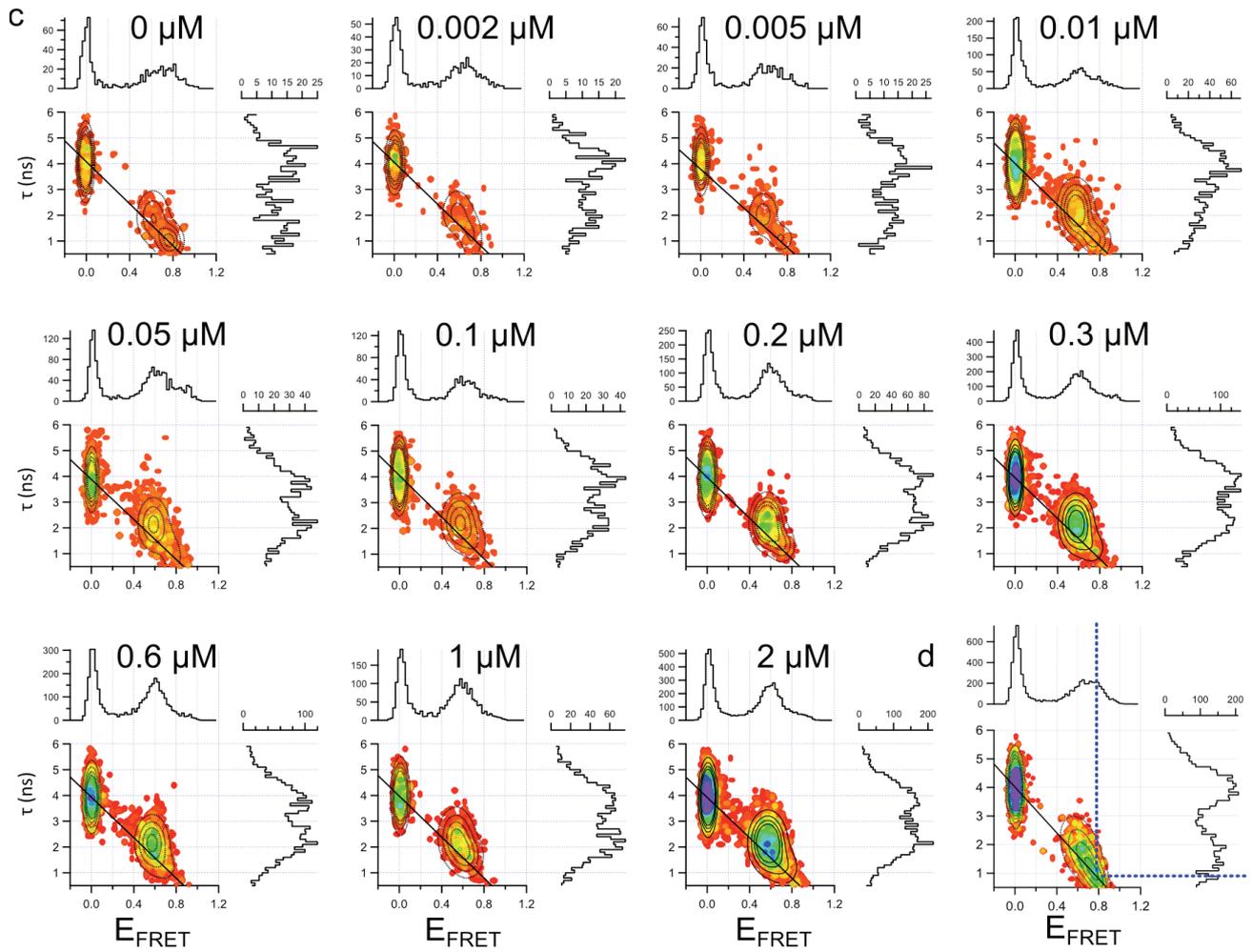
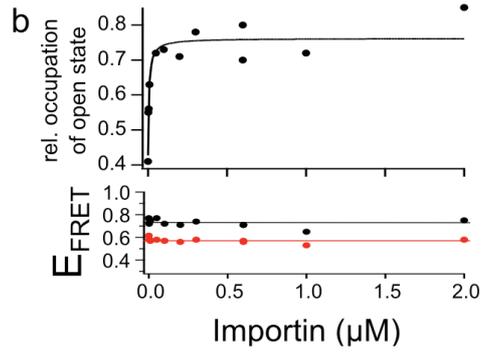
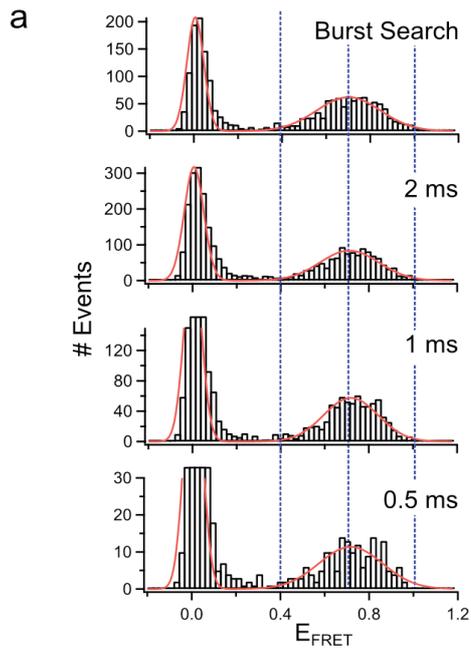


Figure S7: Identification and analysis of two (open and closed) 627-NLS conformations by smFRET.

(a) Binning analysis of 627-NLS FRET histograms. Single molecules were identified using a burst search algorithm⁵, or division of the entire time trace into time bins of different lengths (0.5 ms, 1 ms, or 2 ms), with subsequent thresholding over the Do+Ac signal after Do excitation. The threshold was set to 50 photons in all cases. FRET histograms calculated from all events above threshold and corresponding Gaussian fits suggest that neither the position ($E_{\text{FRET}} = 0.71 \pm 0.01$), nor the width ($\sigma_E = 0.14 \pm 0.01$) of the FRET peak change significantly with observation time. At least two populations, whose interconversion appears static with respect to the diffusion time of 627-NLS through the confocal volume (~ 5 ms), should therefore exist within the observed FRET peak. (b) Titration of 627-NLS I539C A707C double cysteine mutant analysed using two-dimensional triple Gaussian fits in the Do-lifetime (τ) versus E_{FRET} histograms in (c). E_{FRET} values E_1 (low FRET, ‘open’) and E_2 (high FRET, ‘closed’, see also **Table 5**) as well as corresponding amplitudes A_1 and A_2 were extracted from the fits, and the relative occupancy of the open conformation was calculated according

to $\frac{A_1}{A_1 + A_2}$, which could be fit with a simple binding model under the assumption that

importin $\alpha 1$ was always in excess (see also **Fig. 9**). The transition from ‘bound’ to ‘unbound’ follows a remarkably similar trend as when average E_{FRET} are used (compare to **Fig. 9b**) Apart from small deviations, E_1 (lower graph, black) and E_2 (lower graph, red) remain rather constant across wide ranges of importin- $\alpha 1$ concentrations when the fit is not constrained in E nor τ . (c) τ versus E_{FRET} histograms of 627-NLS I539C A707C double cysteine mutant at various importin- $\alpha 1$ concentrations with corresponding two-dimensional triple-Gaussian fits (black dashed curves). Tilt in the Gaussian fit for the τ and E_{FRET} axis were not allowed within a single Gaussian population. The black line corresponds to the behavior expected from a static molecule (‘static FRET line’) (**Table 5**). (d) τ versus E_{FRET} histogram of 627-NLS I539C A707C double cysteine mutant with two-dimensional triple Gaussian fits allowing tilt between τ and E_{FRET} axis in the fit within a single population. E_1 (low FRET, ‘open’) and E_2 (high FRET, ‘closed’) converge to the same values as obtained from (c) and **Fig. 9d**. The closed conformation is again centred on the static FRET line (blue dashed lines), while the open conformation shows fast dynamics.



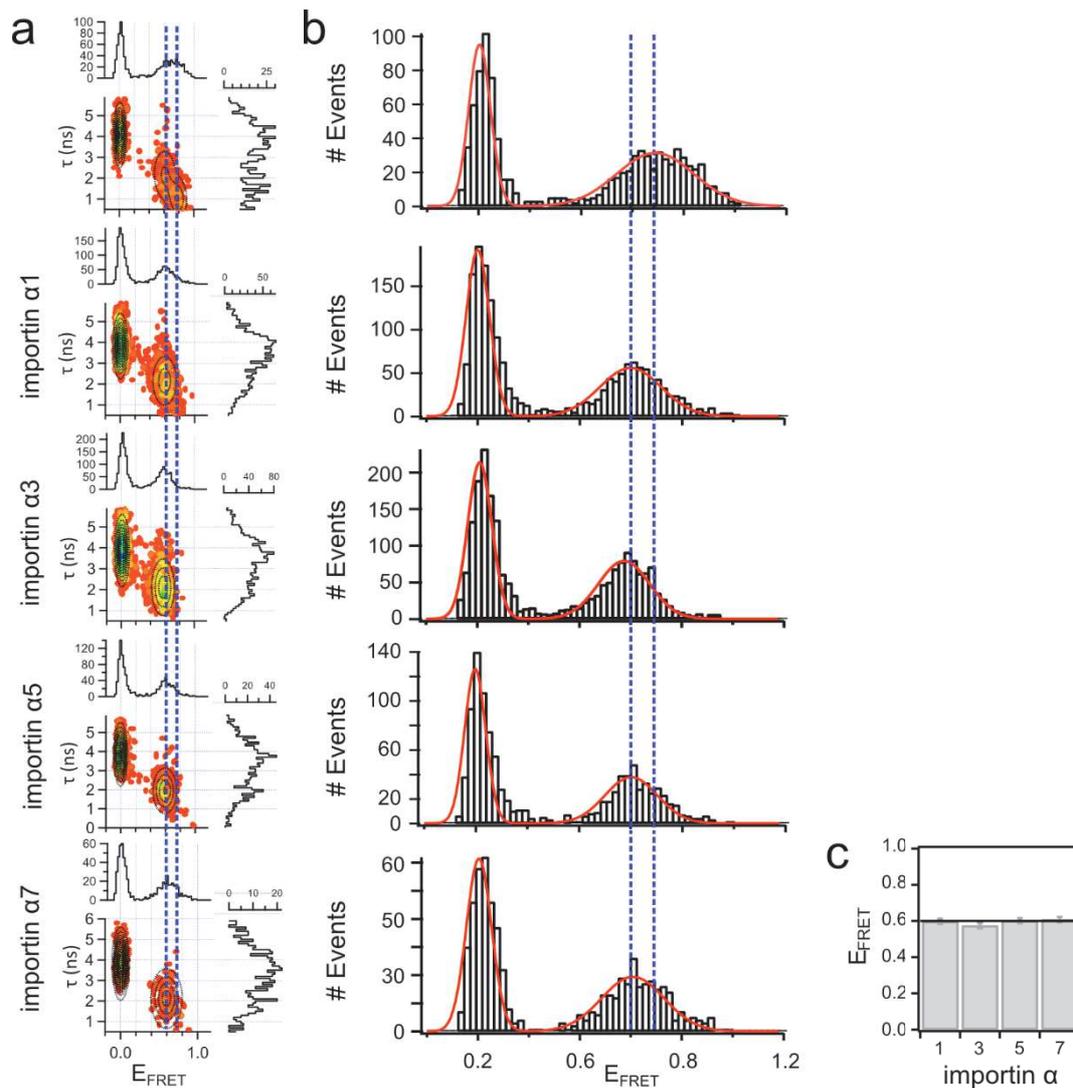


Figure S8.

Interaction of 627-NLS with different importin α proteins. (a) Donor fluorescence lifetime (τ) versus FRET efficiency histograms of 627-NLS 539C 707C in the presence and absence of 5 μM importin $\alpha 1$, $\alpha 3$, $\alpha 5$ and $\alpha 7$ with corresponding two-dimensional double (presence of importin) or triple-Gaussian fits (black dashed curves). Blue dashed lines indicate the centers of the open and closed conformations. Note that the presence of 5 μM importin α does not allow fitting an additional population of the closed conformation. (b) Double Gaussian fits of the corresponding 1D histograms. Blue dashed lines indicate the center of the FRET population. (c) Display of the FRET efficiencies in the presence of the different importin α forms as obtained from (b). The black line indicates the average obtained from all measurements in the presence of importin. Error bars report on the standard deviation obtained from all FRET efficiencies in the presence of different importin forms. All data in this figure have been analysed with 1 ms binning instead of burst recognition, which delivers more robust results in the case of elevated fluorescent background at high importin concentrations.

1. Gopich, I. & Szabo, A. Theory of photon statistics in single-molecule Förster resonance energy transfer. *J Chem Phys* 122, 14707 (2005).
2. Merchant, K. A., Best, R. B., Louis, J. M., Gopich, I. V. & Eaton, W. A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1528–1533 (2007).
3. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci* 15, 2795–2804 (2006).
4. Maltsev, A. S., Ying, J. & Bax, A. Deuterium isotope shifts for backbone ^1H , ^{15}N and ^{13}C nuclei in intrinsically disordered protein α -synuclein. *J. Biomol. NMR* 54, 181–191 (2012).
5. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* 86, 163–180 (2001).

3 Résultats complémentaires

3.1 Relaxation et temps de corrélation

Une des premières preuves de la différence importante de dynamique entre la forme ouverte et la forme fermée du 627-NLS a été obtenue grâce à des données de relaxation du 627-NLS aviaire à 600 MHz à 10°C. Sous certaines conditions, le rapport de R_2/R_1 peut nous informer sur le temps de corrélation τ_c de la molécule observée [Kay et al., 1989].

Les données obtenues sont présentées dans la figure 6.1. On peut voir que la première partie du domaine 627 (résidus 538 à 600) présente un rapport R_2/R_1 et donc un temps de corrélation τ_c différent du reste de la protéine, et que les deux formes attribuées (ouverte et fermée) également. Pour les deux formes, les résidus 538 à 600 présentent un temps de corrélation supérieur au reste de la protéine. Cette première partie correspond au coeur d'hélices compactes en N-terminal du 627-NLS. C'est un domaine rigide, il est donc cohérent que son τ_c soit supérieur à celui du reste de la protéine, plus flexible. D'autre part, la forme ouverte présente un temps de corrélation inférieur à la forme fermée, et ce systématiquement sur toute la longueur de la protéine. Ces observations sont cohérentes avec l'hypothèse d'une forme ouverte dont les deux domaines se comportent indépendamment et ont donc un temps de corrélation plus court, et une forme fermée, plus compacte, dans laquelle les deux domaines sont solidaires l'un de l'autre et qui a donc un temps de corrélation plus long. Néanmoins, il est évident que le rapport R_2/R_1 peut être fortement influencé par des mouvements internes ou de l'échange.

3.2 Equilibre conformationnel du 627-NLS humain

Le 627-NLS humain présente également un équilibre lent en solution. Comme pour le 627-NLS aviaire, une des populations a les mêmes déplacements chimiques que les domaines séparés, la population ouverte, et la seconde a des déplacements chimiques différents et correspond probablement à la forme cristallographique, fermée (Figure 6.2A ; Publication, figure S1E).

Comme pour le 627-NLS aviaire, une dépendance en température des populations a été observée, avec une prédominance de la population fermée à basse température, et une augmentation progressive de la population de conformation ouverte au fur et à mesure de l'augmentation de la température, pour atteindre environ 60 % à 30°C selon l'analyse des populations par les formes de raie (Figure 6.2B). Cependant, les erreurs sur l'estimation de ces populations sont trop importantes pour pouvoir

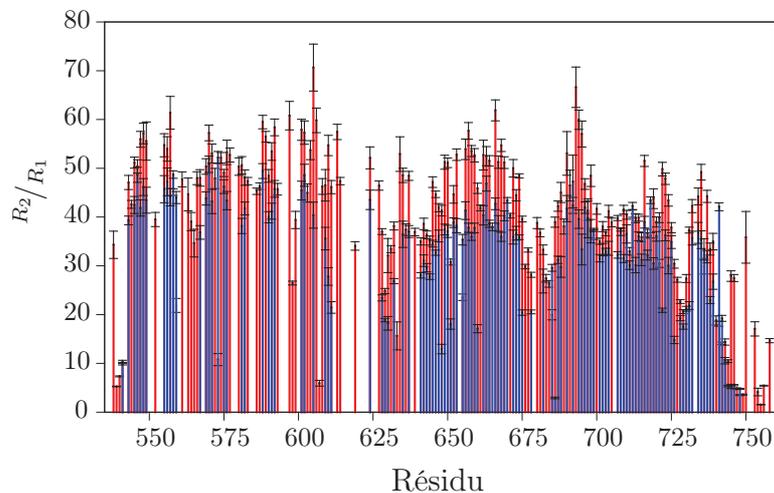


FIGURE 6.1 – Temps de corrélation des formes ouvertes et fermées.
Rapport R_2/R_1 en fonction de la séquence du 627-NLS aviaire à 10°C. La forme ouverte (bleu) a un temps de corrélation plus court que la forme fermée (rouge).

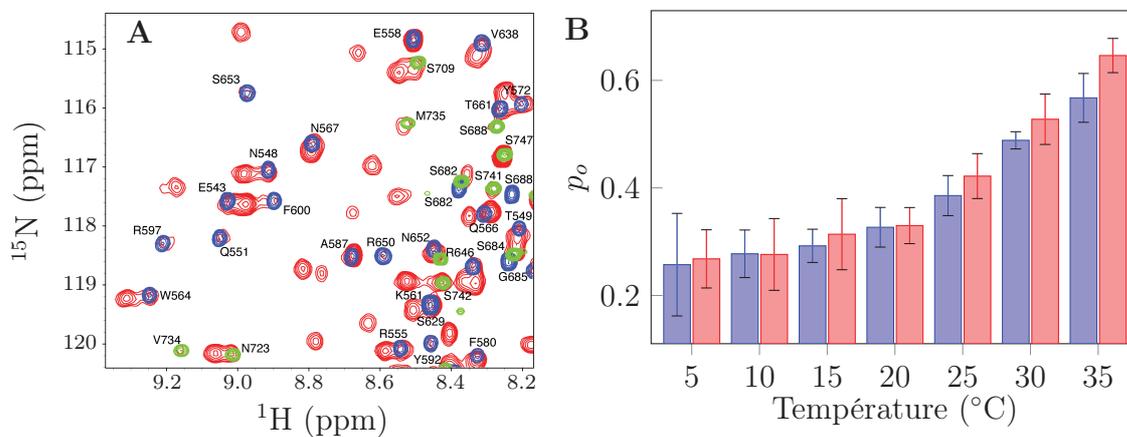


FIGURE 6.2 – Le 627-NLS humain est en équilibre entre les formes ouverte et fermée.
A. Superposition des spectres du 627-NLS (rouge), 627 (bleu) et NLS (vert) humain à 25°C. L'attribution du 627 a été effectuée par Denis Bouvier, celle du NLS publiée [Tarendeau et al., 2007]. **B.** Comparaison entre les populations ouvertes du 627-NLS aviaire (bleu) et humain (rouge) en fonction de la température.

conclure quant à une réelle différence comparées aux populations du 627-NLS aviaire (Figure 6.2B).

L'analyse CEST de la cinétique du 627-NLS humain a également été entreprise, cependant, l'instabilité de cette protéine ne permet pas l'enregistrement d'expériences de qualité au delà de 15°C. Les populations et paramètres cinétiques extraits sont similaires à ceux obtenus pour le 627-NLS aviaire, et ne permettent pas de conclure quant à une différence significative entre les deux souches (Figure 6.3).

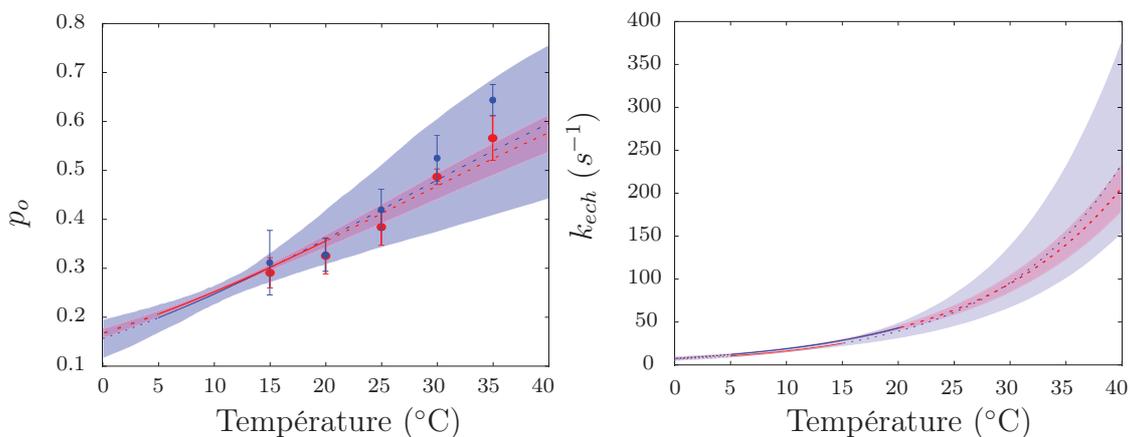


FIGURE 6.3 – Populations et constantes d'échange obtenues par l'ajustement des expériences de CEST sur le 627-NLS humain (en bleu) et comparaison avec l'aviaire (en rouge). La ligne continue représente les données expérimentales, la pointillée son extrapolation de 0 à 40°C et la zone colorée représente l'erreur. Les points correspondent aux valeurs obtenues par l'analyse de la forme de raie.

3.3 La forme ouverte

La forme fermée est stabilisée par un pont salin tripartite entre les résidus R650, E687 et D730. En mutant le résidu 650 en alanine, il a été possible de casser ce pont salin et d'obtenir un mutant ouvert (table des déplacements chimiques en annexe B), présentant uniquement sur le spectre RMN les signaux correspondant à la forme ouverte. Une expérience de SAXS a permis d'obtenir la courbe de diffusion des rayons X de ce mutant. Il a donc ensuite été possible de comparer ces données avec la prédiction de la courbe de SAXS correspondant à l'ensemble structural représentant la forme ouverte (Figure 6.4). L'accord entre ces deux courbes est correct, bien que la courbe expérimentale ne se superpose pas à la courbe prédite. Ceci suggère la présence d'une population de conformation fermée ou que l'échantillonnage de l'espace conformationnel disponible par le 627-NLS dans sa forme ouverte n'est pas uniforme. D'autre part, nous avons enregistré une expérience de CEST sur ce mutant, et malgré l'absence de signaux correspondant à la forme ouverte, quelques

résonances présentent un échange, bien que beaucoup plus faible que celui observé pour le 627-NLS non muté ($k_{ech} = 132 \text{ s}^{-1}$, population fermée de 1,7 % à 10°C) (Figure 6.5). La présence de ce faible échange résiduel n'explique donc pas totalement la différence observée entre la courbe expérimentale du 627-NLS R650A et la courbe prédite à partir de l'ensemble structural décrivant la forme ouverte. Enfin, le FRET du mutant R650A démontre que la conformation ouverte observée est encore plus ouverte (distance entre les fluorophores plus grande) que la conformation ouverte du 627-NLS aviaire (Publication, Figure S6B). Ces résultats suggèrent que le mutant R650A ne correspond pas exactement à la forme ouverte de la protéine sauvage, expliquant les différences observées entre la courbe de SAXS expérimentale et prédite.

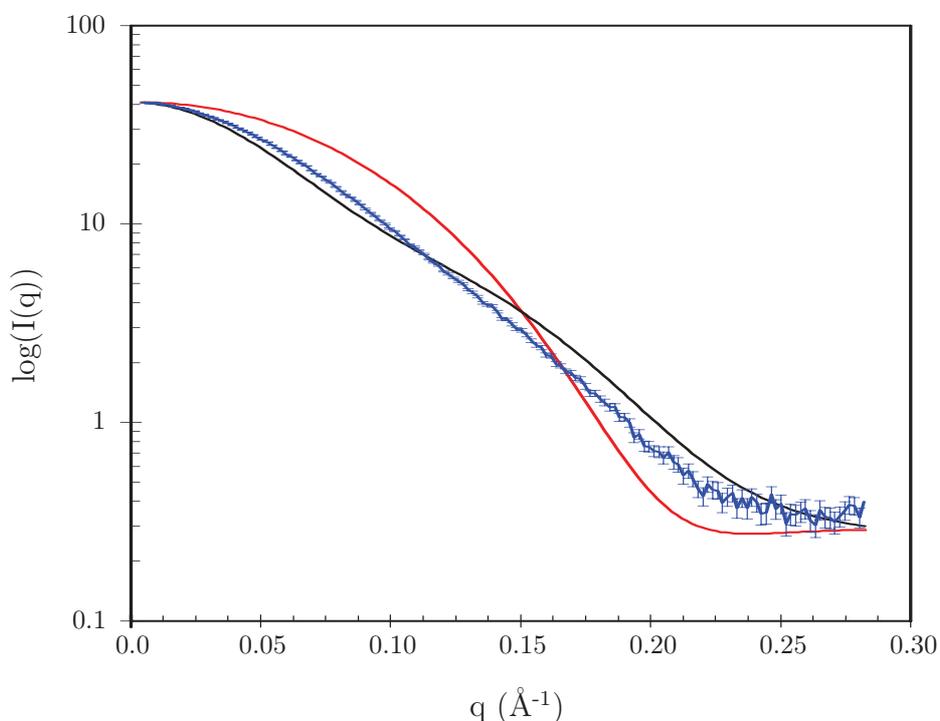


FIGURE 6.4 – Comparaison de la courbe SAXS expérimentale du 627-NLS aviaire R650A (en bleu) avec la courbe SAXS prédite de l'ensemble ouvert (en noir) et de l'ensemble fermé (en rouge).

3.3.1 Augmentation paramagnétique de la relaxation

Dans le but d'obtenir des contraintes longues distances pour confirmer la description structurale de la forme ouverte du 627-NLS, j'ai effectué des expériences d'augmentation paramagnétique de la relaxation sur le mutant R650A. Pour ce faire, j'ai couplé une sonde TEMPO grâce à une liaison thioéther sur une cystéine introduite dans la protéine, à la position 587 ou 675. J'ai ensuite enregistré un spectre

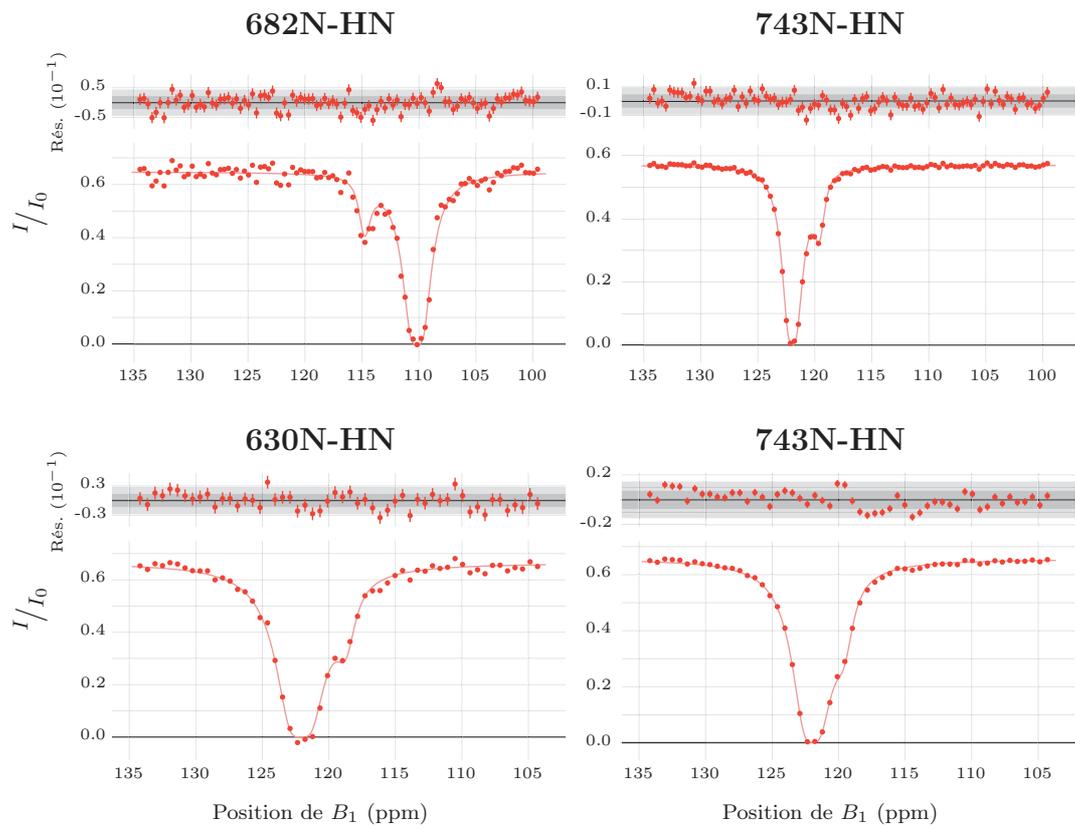


FIGURE 6.5 – Exemples de profils CEST obtenus pour le 627-NLS aviaire R650A, avec un champ B_1 de 22 et 44 Hz en haut et en bas respectivement.

TROSY ^1H - ^{15}N des protéines marquées, à 10°C . Un spectre de la protéine non marquée a été utilisé comme référence, et le rapport de l'intensité des pics de la protéine marquée sur celle des pics de la protéine non marquée a été calculé (Figure 6.6A). C'est pourquoi certains rapports obtenus sont supérieurs à 1, la présence de la sonde pouvant influencer la dynamique de la protéine.

Pour permettre l'interprétation des données, j'ai choisi de reporter l'amplitude des PREs obtenus sur une structure extraite de l'ensemble structural de la forme ouverte du 627-NLS aviaire (Figure 6.6B). Pour chacun des mutants, on observe un effet important de la sonde dans le domaine 627 en fonction de la distance à la sonde, ce qui est attendu. Le domaine NLS est affecté par des PREs de basse amplitude, sauf pour sa première hélice. Dans le cas d'un marquage à la position 675, ceci n'est pas surprenant puisque l'acide aminé 675 est situé sur le domaine 627 mais du côté du NLS. Cette hélice se situe donc dans le rayon d'effet attendu de la sonde. Par contre, pour un marquage à la position 587, ce n'est pas le cas. Enfin, le peptide NLS, en C-terminal du domaine NLS, est dans les deux cas touché par l'augmentation de relaxation, bien qu'il soit éloigné de la position de la sonde. Cette dernière différence peut être expliquée par la flexibilité du peptide NLS, balayant un large éventail de conformations en échange rapide, l'amenant ponctuellement à proximité de la sonde.

De façon plus générale, nous avons démontré grâce au SAXS et au CEST qu'un échange résiduel était présent pour le mutant R650A. Bien que les signaux de la conformation minoritaire ne soient pas présents sur le spectre, il est possible que certains résidus (notamment ici du domaine NLS) soient affectés par la sonde uniquement dans la conformation minoritaire et, la protéine continuant à échanger durant la séquence d'impulsions RMN, que cet effet soit mesurable sur les signaux de la conformation majoritaire. Ce phénomène explique la présence de PREs d'amplitude moindre dans le domaine NLS. Cependant, aucune différence significative n'a été observée entre les PREs enregistrés à 10 et 25°C , suggérant que les différences observées entre les PREs enregistrés et attendus d'après l'ensemble conformationnel de la conformation ouverte ne proviennent pas de la présence de conformation fermée dont la population serait différente à ces deux températures. On remarque que les parties flexibles sont particulièrement touchées, suggérant la présence de sonde libre dans l'échantillon, malgré une purification par chromatographie d'exclusion de taille après marquage. Les PREs obtenus sont donc plutôt en accord avec la description que nous faisons de la structure ouverte, mais la mesure de PREs sur un mutant ne présentant plus aucun échange, et éventuellement sur d'autres mutants choisis stratégiquement, ainsi qu'une purification de la protéine marquée permettant d'éliminer toute la sonde libre, pourraient permettre une confirmation solide de cette conclusion.

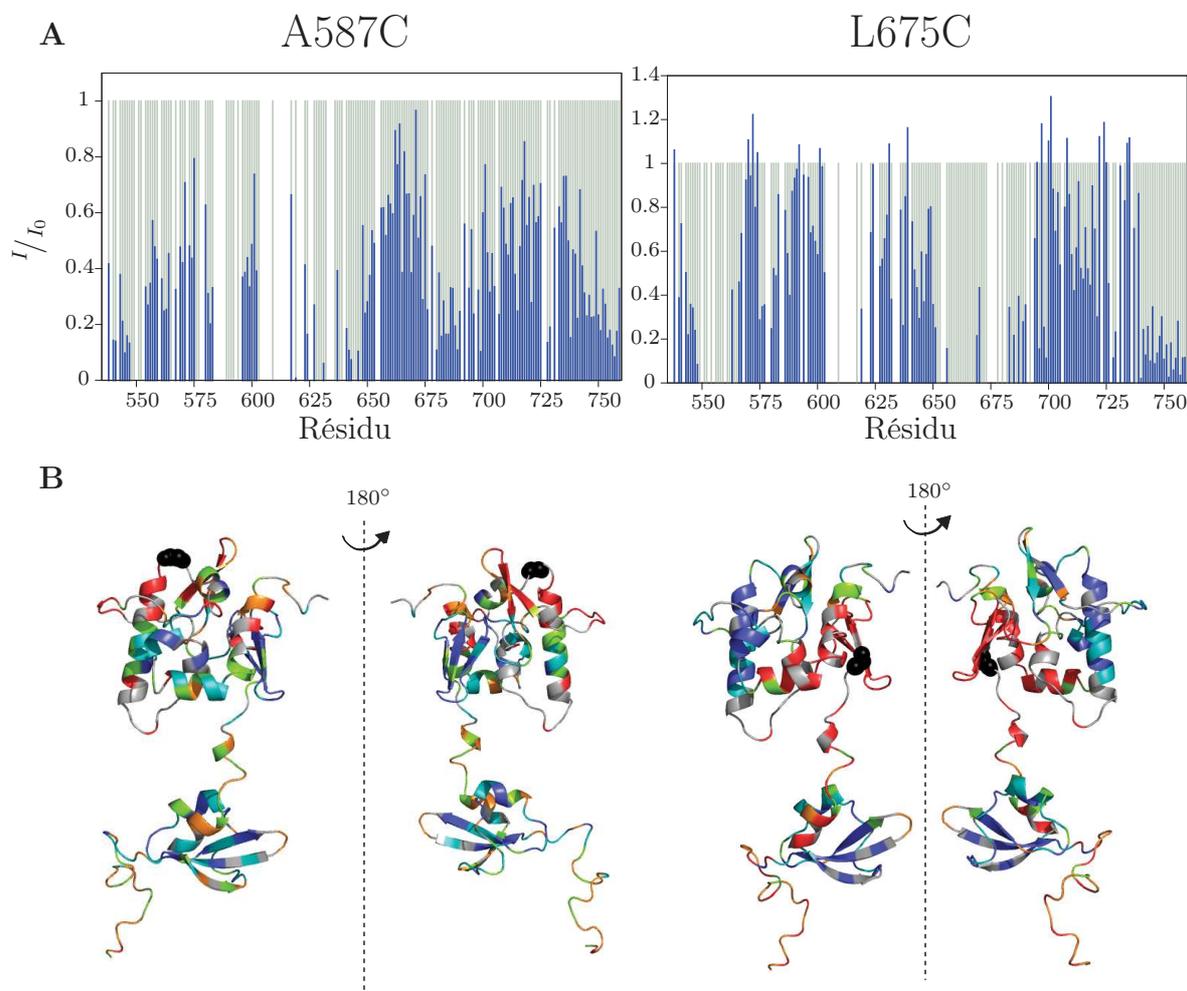


FIGURE 6.6 – Relaxation paramagnétique du 627-NLS aviaire R650A A587C (à gauche) ou L675C (à droite) marqués par du TEMPO à 10°C.

A. En bleu, PRE obtenus en faisant le rapport de l'intensité du pic marqué et du pic non marqué en fonction de la séquence. En gris, le rapport I_0/I_0 . **B.** Représentation sur une structure de l'ensemble ouvert du 627-NLS aviaire des PREs obtenus selon leur amplitude : $< 0,1$, $< 0,25$ et $< 0,3$, $< 0,4$ et $< 0,5$, $< 0,6$ et $< 0,8$, $> 0,6$ et $> 0,8$ pour A587C et L675C respectivement. En gris, les résidus sans données. La boule noire représente le résidu auquel est attachée la sonde.

3.4 Interaction avec l'importine α par RMN

J'ai réalisé une première expérience visant à étudier l'interaction du 627-NLS aviaire avec l'importine $\alpha 1$ par RMN en utilisant le 627-NLS aviaire marqué et l'importine $\alpha 1$ non marquée. Cependant, après formation du complexe, aucun signal ne restait visible dans le spectre TROSY enregistré. Pour pouvoir observer l'interaction du point de vue de l'importine α , nous avons entrepris l'attribution de l'importine $\alpha 1$, qui est trop grande pour que tous ses signaux soient visibles dans le spectre (50 kDa), mais comporte une partie flexible de 34 acides aminés en C-terminal (496-529) ainsi qu'une dizaine d'acides aminés flexibles en N-terminal (60-72). J'ai pu attribuer la majorité de ces parties flexibles (Figure 6.7, table des déplacements chimiques en Annexe B). Le calcul des déplacements chimiques secondaires et des propensions de structures secondaires n'a pas montré de tendance particulière pour les parties attribuées, proches du *random coil* (Figure 6.8).

J'ai également effectué une titration par 0, 25, 50, 75, 100 et 150 % de 627-NLS aviaire non marqué (Figure 6.9A). Deux phénomènes sont observables. Premièrement, certains pics disparaissent progressivement dès les premiers ajouts de 627-NLS jusqu'à disparaître complètement lorsqu'on atteint la saturation de l'importine α . Pour mesurer cet effet, j'ai calculé le rapport de l'intensité des pics avant et après saturation (Figure 6.9B). Deuxièmement, certains pics bougent au cours de la titration. Pour quantifier ces changements, j'ai calculé les perturbations de déplacement chimique (CSP) selon la formule 3.12 avec $c = 0,15$ (Figure 6.9C).

Tout d'abord, il est à noter que les signaux de la partie N-terminale gagnent en intensité à saturation. Ceci pourrait être expliqué par un gain de flexibilité de cette partie lors de l'interaction, donnant des signaux plus fins et donc plus intenses. Le N-terminal de la construction utilisée correspond au *linker* entre le "corps" de l'importine $\alpha 1$ et le peptide d'interaction avec l'importine β (IBB, ici supprimé). Une augmentation de flexibilité de cette partie lors de l'interaction avec le cargo, ici le 627-NLS, pourrait permettre de favoriser l'interaction de l'importine α avec l'importine β pour l'import au noyau.

Les signaux de la partie C-terminale perdent en moyenne les deux tiers de leur intensité lors de l'interaction, avec une diminution progressive du rapport I/I_0 autour des acides aminés 516 et 517. De plus, quatre pics disparaissent complètement : ceux correspondant aux acides aminés T516, F517, T526 et F527. Cette observation, corrélée avec l'ensemble décrivant le complexe construit par ailleurs (Publication, Figure 7B), suggère une interaction transitoire entre le domaine 627 et la partie C-terminale de l'importine α dans le contexte du complexe. De plus, la disparition des signaux de deux fois deux acides aminés consécutifs, deux motifs "TF", suggère qu'un site d'interaction pour ce motif pourrait être présent sur le domaine 627.

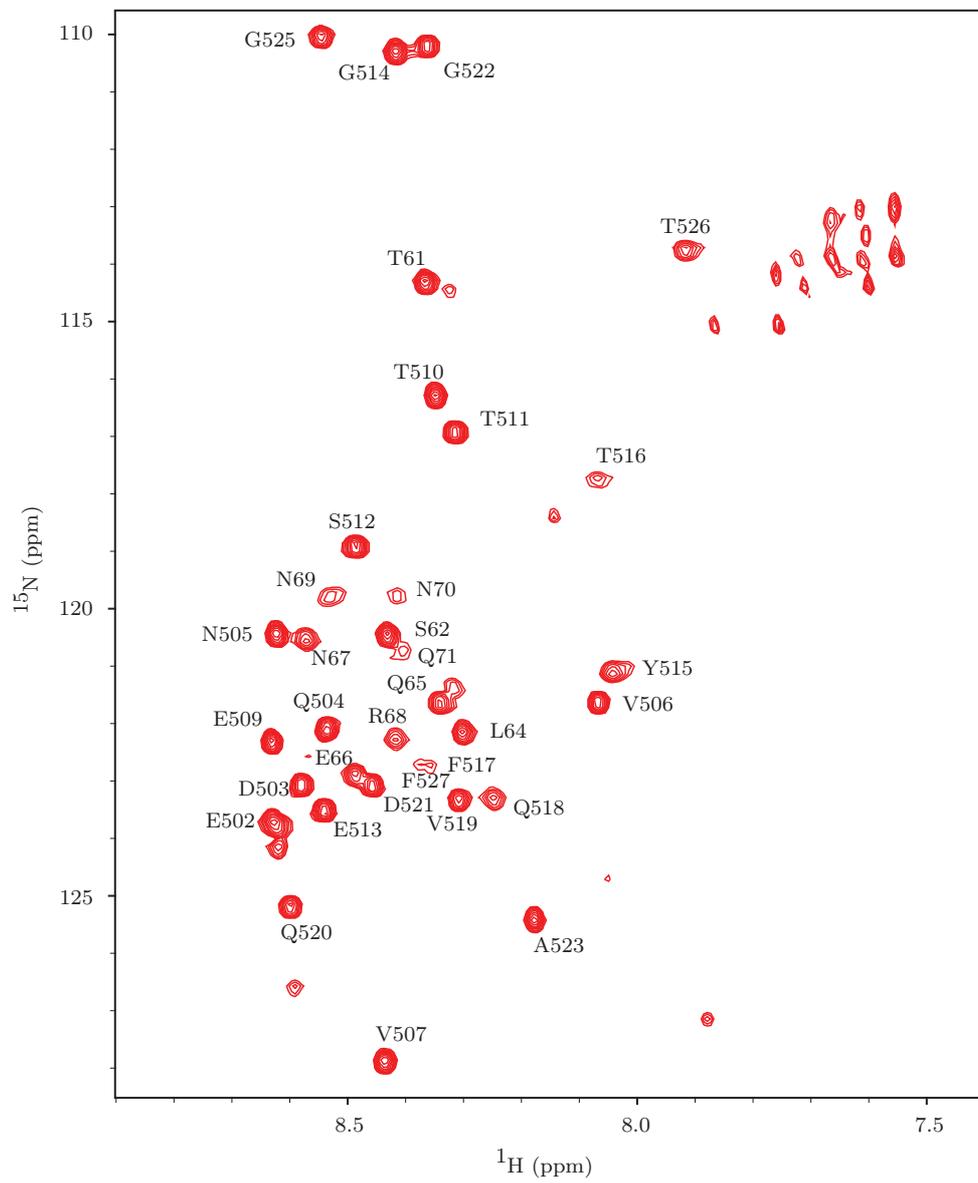


FIGURE 6.7 – Spectre TROSY ^1H - ^{15}N à 10°C attribué de l'importine $\alpha 1$.

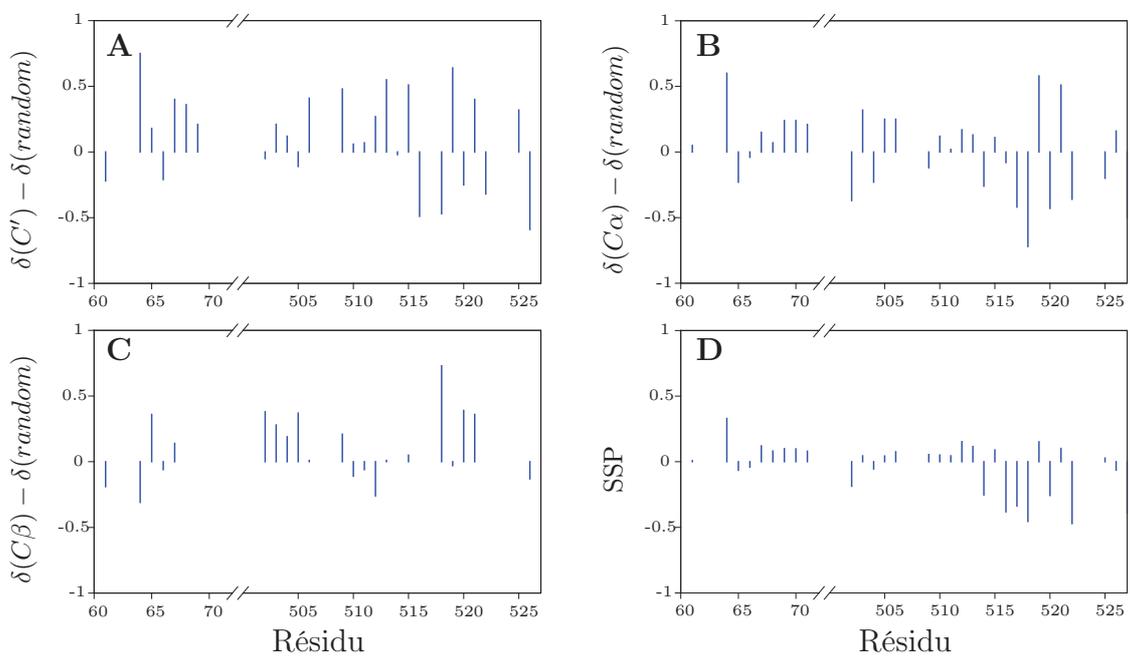


FIGURE 6.8 – Déplacements chimiques secondaires et SSP de l'importine $\alpha 1$. Déplacements chimiques secondaires de l'importine $\alpha 1$ des carbonyles (**A**), carbones α (**B**) et carbones β (**C**). **D**. Propensions de structures secondaires calculées par l'algorithme SSP à partir des déplacements chimiques C_α , C_β et C' .

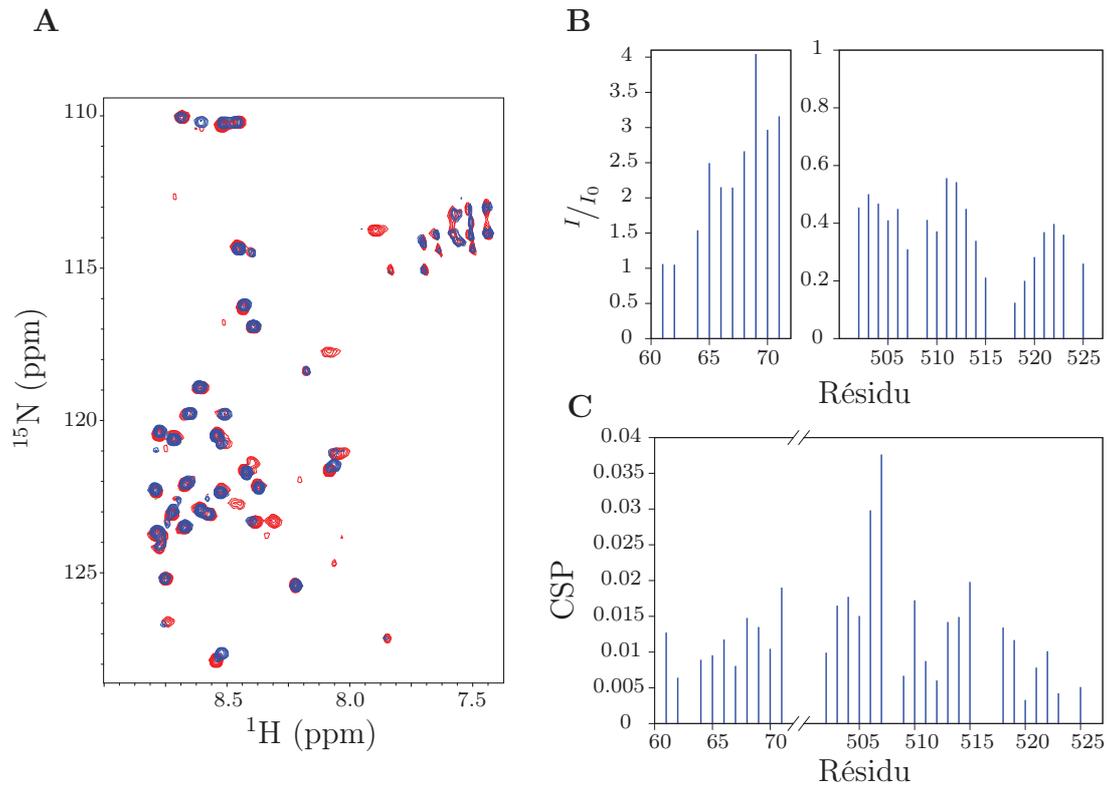


FIGURE 6.9 – Interaction du 627-NLS aviaire avec l'importine $\alpha 1$. **A.** Superposition du spectre de l'importine $\alpha 1$ libre (rouge) et titrée par 150 % de 627-NLS (bleu). **B,C.** Rapport des intensités (**B**) et perturbation des déplacements chimiques (**C**) observés lorsque l'importine $\alpha 1$ est saturée par 150 % de 627-NLS.

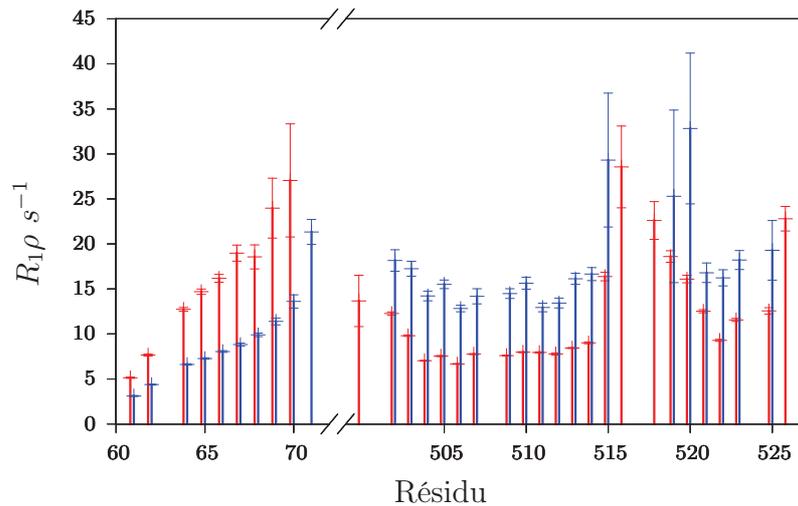


FIGURE 6.10 – Relaxation $R_{1\rho}$ de l'importine $\alpha 1$ libre (rouge) ou liée au 627-NLS (bleu) à 10°C à 600 MHz.

La perturbation des déplacements chimiques (CSP) est faible, et donc délicate à interpréter. Les CSP observés en N-terminal sont très faibles, entre 0,05 et 0,015. Cette partie n'interagissant pas avec le 627-NLS d'après l'analyse des rapports d'intensités, ceci suggère que seuls les CSP supérieurs à ces valeurs sont à prendre en compte. En C-terminal, seuls deux acides aminés présentent une valeur de CSP supérieur à 0,2 : V506 et V507. Ces deux résidus sont potentiellement affectés par la proximité du domaine 627. Cependant, les acides aminés voisins ne changent pas de déplacements chimiques.

L'hypothèse d'une interaction des résidus 516, 517, 526 et 527 avec le domaine 627 est confortée par les données de relaxation enregistrées sur les formes libre et liée de l'importine α . En effet, les résidus 516, 517, 526 et 527 présentent dans ce cas une relaxation plus rapide que le reste des résidus (Figure 6.10). Deux hypothèses principales permettraient d'expliquer ce phénomène. D'une part, il est possible que les deux motifs TF soient engagés dans une interaction hydrophobe intramoléculaire. D'autre part, une interaction transitoire entre ces résidus et le corps de l'importine α pourrait avoir lieu, même en l'absence du cargo 627-NLS. La relaxation de ces mêmes résidus augmente lors de l'interaction de l'importine α avec le 627-NLS, suggérant soit une interaction directe avec le domaine 627, soit une interaction intramoléculaire plus stable. Par ailleurs, la relaxation de la partie N-terminale de l'importine α diminue lors de l'interaction entre l'importine α et le 627-NLS, confirmant l'augmentation de flexibilité proposée. Les rôles de l'interaction intramoléculaire des motifs TF de l'importine α et/ou le 627-NLS ainsi que de l'augmentation de flexibilité du N-terminal de l'importine α lors de l'interaction restent peu clairs, mais ces observations sont très intéressantes dans la perspective de mieux comprendre les mécanismes régissant la spécificité des différentes importines α pour leurs différents cargos.

3.5 Interaction avec l'importine α par smFRET

Trois autres couples de mutations cystéines ont été étudiés en plus de celui présenté dans la publication : I539C/A717C, L675C/A707C et L675C/A717C (Figure 6.11). Les données obtenues sur ces mutants sont moins claires que celles obtenues sur le couple I539C/A707C. En effet, il est plus difficile de distinguer deux populations pour les mesures effectuées en absence d'importine α . Cependant, l'éloignement des deux fluorophores, et donc l'ouverture du 627-NLS lors de l'interaction avec l'importine α , reste très clairement visible pour tous les mutants. D'autre part, la flexibilité du 627-NLS dans sa forme liée au complexe est également visible de par la déviation de la population observée de la ligne de FRET statique.