



HAL
open science

RNA binding and assembly of human influenza A virus polymerases

Christopher Swale

► **To cite this version:**

Christopher Swale. RNA binding and assembly of human influenza A virus polymerases. Structural Biology [q-bio.BM]. Université Grenoble Alpes, 2015. English. NNT : 2015GREAV053 . tel-01687520

HAL Id: tel-01687520

<https://theses.hal.science/tel-01687520>

Submitted on 18 Jan 2018

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 GRENOBLE ALPES

Spécialité : **Biologie Structurale et Nanobiologie**

Arrêté ministériel : 7 août 2006

Présentée par

Christopher SWALE

Thèse dirigée par le **Professeur Rob RUIGROK** et
codirigée par le **Docteur Thibaut CREPIN**

préparée au sein de l'**Unit of Virus Host Cell Interaction (UVHCI)**, **Negative RNA Viruses group**

Ecole Doctorale Chimie Science du Vivant (EDCSV)

RNA BINDING AND ASSEMBLY OF HUMAN INFLUENZA A VIRUS POLYMERASES.

Thèse soutenue publiquement le **13 Novembre 2015**,
devant le jury composé de :

Professeur Johanes GEISELMANN

PR, Université Joseph Fourier, Président du Jury

Docteur Alan Hay

DR, Watson Crick Institute, Rapporteur

Docteur Bernard DELMAS

DR1, Unité de Virologie et Immunologie Moléculaires, Rapporteur

Docteur Christoph BIENIOSSEK

Responsable de groupe, Roche Pharma R&D, Examineur

Professeur Rob RUIGROK

PR, Université Joseph Fourier, Directeur de Thèse

Docteur Thibaut CREPIN

CR1, Unit of Virus Host Cell Interactions, co-Directeur de Thèse



Acknowledgements

TABLE OF CONTENT

CHAPTER 1: INTRODUCTION	1
RESUME DU CHAPITRE 1 EN FRANÇAIS	2
1.1 GENERAL INFORMATION ON THE INFLUENZA VIRUS	5
1.1.1 INFLUENZA VIRUSES	5
1.1.2 ZOOONOSIS OF INFLUENZA A VIRUSES	6
1.1.3 HISTORICAL BACKGROUND	7
1.1.4 SEASONAL INFLUENZA	8
1.1.5 PANDEMIC INFLUENZA.....	11
1.1.6 CURRENT TREATMENT STRATEGIES	13
1.2 STRUCTURAL FEATURES AND INFECTIOUS CYCLE OF THE INFLUENZA A VIRUS	17
1.2.1 STRUCTURE OF THE VIRAL PARTICLE	17
1.2.2 SEGMENTED GENOME ORGANISATION	18
1.2.3 VIRUS LIFE CYCLE	31
1.3 EVOLUTION OF THE INFLUENZA VIRUS	36
1.3.1 ANTIGENIC DRIFT	36
1.3.2 ANTIGENIC SHIFT.....	36
1.4 THE INFLUENZA POLYMERASE STRUCTURE, FUNCTION AND ASSEMBLY PATHWAY	38
1.4.1 FOUNDING WORK ON THE POLYMERASE SUBUNITS.....	38
1.4.2 LOW RESOLUTION EM MAPS AND INDIVIDUAL DOMAIN STRUCTURES OF THE POLYMERASE	39
1.4.3 FULL ATOMIC MODELS FOR THE RNA DEPENDANT RNA POLYMERASE.....	44
1.4.4 CURRENT MODELS FOR TRANSCRIPTION AND REPLICATION	49
1.4.5 ASSEMBLY PATHWAY OF THE POLYMERASE	52
1.5 THESIS OBJECTIVES	57
CHAPTER 2: EXPERIMENTAL PROCEDURE.....	59
RESUME DU CHAPITRE 2 EN FRANÇAIS	60
2.1 CLONING OF THE RdRp POLYPROTEIN GENES AND BACULOVIRUS GENERATION	61
2.2 HETEROTRIMERIC AND HETERODIMERIC POLYMERASE COMPLEXES PURIFICATION	63
2.3 CLONING OF RANBP5.....	64
2.4 PURIFICATION OF RANBP5.....	64
2.5 THERMOFLUOR	65
2.5.1 PRINCIPLE.....	65
2.5.2 EXPERIMENTAL PROCEDURE.....	65
2.6 AFFINITY BINDING TECHNIQUES	66
2.6.1 FLUORESCENCE ANISOTROPY	66
2.6.2 FILTER BINDING ASSAY	68
2.6.3 EXPERIMENTAL METHODS.....	68
2.7 SIZE EXCLUSION CHROMATOGRAPHY COUPLED WITH MALLS	69
2.7.1 PRINCIPLE OF MALLS.....	69
2.7.2 EXPERIMENTAL PROCEDURE FOR MALLS	69

2.8 SMALL ANGLE XRAY SCATTERING (SAXS)	71
2.8.1 SAXS PRINCIPLE	71
2.8.2 DATA AQUISION	71
2.8.3 DATA ANALYSIS	72
2.8.4 EXPERIMENTAL PROCEDURES FOR SAXS	74
2.9 X-RAY CRYSTALLOGRAPHY	75
2.9.1 PRINCIPLE.....	75
2.9.2 CRYSTALLOGENESIS	75
2.9.3 ACQUISITION AND INTEGRATION	77
2.9.4 PHASE DETERMINATION.....	78
2.9.5 EXPERIMENTAL PROCEDURES FOR X-RAY CRYSTALLOGRAPHY:	81
2.10 TRANSMISSION ELECTRON MICROSCOPY USING NEGATIVE STAINING	81
2.10.1 PRINCIPLE.....	81
2.10.2 DATA COLLECTION	82
2.11 AUTOMATED MICRO BIO-REACTOR (AMBR)	83
2.11.1 BRIEF OVERVIEW.....	83
2.11.2 AUTOMATED INSECT CULTURES	84
2.11.3 SMALL-SCALE PURIFICATION	84
CHAPTER 3: RNA BINDING AND ASSEMBLY OF HUMAN INFLUENZA A VIRUS POLYMERASES	85
RESUME DU CHAPITRE 3 EN FRANÇAIS	86
NOTE OF INTRODUCTION TO CHAPTER 3	87
CHAPTER 3 MANUSCRIPT.....	88
CHAPTER 4: STRUCTURAL CHARATERISATION OF RANBP5	89
RESUME DU CHAPITRE 4 EN FRANÇAIS	90
NOTE OF INTRODUCTION TO CHAPTER 4	91
4.1 OBJECTIVES	93
4.2 EXPRESSION AND PURIFICATION OF RANBP5	93
4.3 SAXS CHARACTERISATION	94
4.4 INITIAL CHARACTERISATION AND CRYSTALISATION	96
4.5 CRYSTAL OPTIMISATION	99
4.6 DATA COLLECTION.....	101
4.7 INCREASING RESOLUTION THROUGH <i>IN-SITU</i> PROTEOLYSIS	102
4.8 EXPERIMENTAL PHASING OF RANBP5.....	106
4.8.1 SAD PHASING OF RANBP5.....	107
4.8.2 SIR AND MIR PHASING	109
4.9 CONCLUSION AND PERSPECTIVES	110
CHAPTER 5: MICRO-SCALE CONSTRUCT SCREENING USING AMBR	111
RESUME DU CHAPITRE 5 EN FRANÇAIS	112
NOTE OF INTRODUCTION TO CHAPTER 5	113

5.1 INTRODUCTION	115
5.2 CONSTRUCT PANEL OVERVIEW	116
5.3 AMBR PURIFICATION RESULTS	118
5.4 CONCLUSION	121
GENERAL CONCLUSION.....	123
LIST OF FIGURES.....	127
LIST OF ABBREVIATIONS	129
BIBLIOGRAPHY	132

CHAPTER 1: INTRODUCTION

Résumé du chapitre 1 en Français

La Grippe est une maladie d'origine virale causée par un virus possédant un génome ARN segmenté et de polarité négative. Ce virus, dont la découverte et la caractérisation datent du début du siècle appartient à l'ordre des *Mononegavirales* et à la famille des Orthomyxoviridae. Au sein de cette famille virale, 6 genres de virus sont classifiés, dont trois (*influenza A, B et C*) sont des virus grippaux. Ces genres se distinguent à la fois par des différences génomiques et antigéniques. Tous les virus de la grippe présentent des protéines de surface. L'hémagglutinine (HA) et la neuraminidase (NA) sont caractéristiques des types A et B. Le type C présente une seule protéine de surface, l'hémagglutinine-estérase-fusion (HEF).

Les genres et sous-types de virus de la grippe se différencient aussi par leur zoonose, qui comprend de nombreuses espèces mammifères. Le virus influenza de type A est capable d'infecter un nombre important d'espèces mammifères (porcs, chevaux, humains et chauves-souris), mais présente son réservoir naturel chez les oiseaux. A l'heure actuelle, 105 espèces aviaires sont les hôtes de plusieurs souches de virus influenza de type A présentant 16 sous-types HA (H1-H16) et 9 sous-types NA (N1–N9) différents. Plusieurs souches de grippe A (H17N9, H18N10) ont aussi été découvertes ces dernières années chez des chauves-souris vivant dans les jungles d'Amérique du sud. Influenza B ne présente qu'un seul type de HA/NA et infecte principalement l'homme selon un mode épidémique saisonnier. Influenza C est uniquement détectée chez l'homme et présente un danger sanitaire moindre par rapport aux autres sous types grippaux.

Certaines espèces influenza A infectent l'homme selon deux modes épidémiques distincts : de façon épidémique (saisonnière) et de façon pandémique.

Le mode épidémique saisonnier affecte les deux hémisphères du globe. Ce phénomène est permis par des glissements antigéniques, qui sont le produit de mutations ponctuelles du virus de la grippe. Ces mutations permettent aux virus d'échapper à l'immunisation progressive des populations infectées. Les épidémies grippales sévissent généralement pendant les saisons automno-hivernale dans les pays tempérés. Plusieurs facteurs semblent être la cause de ce phénomène. Bien que moins pathogènes que les manifestations pandémiques, les épidémies saisonnières infectent tous les ans un nombre considérable de personnes et engendrent des dommages sanitaires et économiques importants. Il est estimé qu'entre 3 et 5 millions de personnes sont hospitalisées suite à des infections grippales et entre 250 000 à 500 000 personnes en décèdent dans le monde.

La seconde manifestation épidémiologique de la grippe est parfois pandémique. Lors de l'infection simultanée d'une cellule hôte par deux ou plusieurs virus grippaux,

des réassortiments nouveaux peuvent se produire générant de nouvelles souches. Ces virus pandémiques peuvent présenter une infectivité et/ou une pathogénicité grandement accrue. Au cours de l'histoire moderne, plusieurs pandémies ont sévi et parfois causé très rapidement la mort de millions de personnes. L'exemple le plus marquant est celui de la grippe dite « espagnole » de 1918 qui causa entre 50 et 100 millions de morts. Les deux formes épidémiques de grippe sont liées historiquement et génétiquement. Les souches pandémiques sont souvent issues de souches épidémiques adaptées.

La grippe est donc un problème majeur de santé public dans le monde dont le traitement à la fois symptomatique et préventif reste une priorité pour de nombreux organismes de santé public gouvernementaux et internationaux. A l'heure actuelle, la vaccination systémique des populations à risque (jeunes enfants, personnes âgées, personnes immunodéprimées ou souffrant de maladies respiratoires) reste la meilleure méthode de prophylaxie. Les programmes de vaccination mis en place depuis soixante ans ont permis de contenir la grippe épidémique saisonnière. Néanmoins, ce traitement présente des limitations, la plus importante étant durée nécessaire à la production du vaccin qui peut durer jusqu'à 6 mois.

Il y a donc un engouement scientifique à développer des inhibiteurs antiviraux pouvant traiter ou réduire l'infection grippale. A l'heure actuelle, seul un antiviral, l'osaltamivir ou Tamiflu (Roche) présente une efficacité relative envers certaines souches et pour certaines personnes.

Le virus de la grippe présente un matériel génétique divisé en 8 segments dont la réplication et la transcription sont indépendantes. Ces 8 segments codent pour un total de 10 protéines indispensables dont les sous-unités de polymérase (PA, PB1 et PB2), NP, HA, NA, M1, M2, NS1 et NS2. D'autres protéines accessoires et dont la présence n'est détectée que chez certaines souches sont aussi exprimées par ces 8 segments. Toutes ces protéines interviennent à différents cycles pendant la réplication et sont toutes des cibles intéressantes pour développer des inhibiteurs antiviraux.

Au cœur du processus réplcatif du virus de la grippe se trouve l'ARN-polymérase ARN-dépendante constituée de trois protéines majeures : protéine acide (PA) protéine basique 1 (PB1) et protéine basique 2 (PB2). Cet hétérotrimère, une fois associé à la nucléoprotéine (NP) et à l'ARN virale constitue un complexe nommé ribonucléoprotéine (RNP). Les RNP sont dotées de deux activités distinctes, la réplication du génome ARN et la transcription de celui-ci. Ces deux étapes se déroulent exclusivement dans le noyau de la cellule infectée. Lors de la réplication, les RNP génèrent tout d'abord un ARN complémentaire qui sert ensuite de matrice réplcative pour ré-amplifier des ARNv. Lors de la transcription, les RNP génèrent des ARNm à partir de l'ARNv qui permettent ensuite l'expression de protéines

virales. L'ARN polymérase est dépourvue d'activité de synthèse de coiffe, celle-ci effectue donc un vol de coiffe issue d'ARNm cellulaires naissant, en utilisant trois activités différentes réparties sur les trois sous-unités de l'ARN polymérase. La coiffe de l'ARNm est capturée par un domaine de fixation de coiffe sur PB2, la coiffe suivie de 12 nucléotides est ensuite clivée par un domaine endonucléase localisé sur PA. Il s'en suit une étape de transcription réalisée par PB1.

L'étude structurale de la polymérase s'est longtemps limitée à la caractérisation des domaines solubles de PA et PB1 qui a permis de proposer un modèle pour expliquer le vol de coiffe. Il a par contre fallu attendre l'année 2014 pour obtenir les premières structures cristallographiques de polymérase grippale (B et A de chauve-souris) permettant une compréhension accrue de son architecture globale et de son fonctionnement. L'assemblage de la polymérase reste encore mal compris. Le modèle actuel postule sur l'assemblage nucléaire de la polymérase qui impose l'adressage des différentes sous-unités vers le noyau. PB2 serait importée via l'interaction avec l'importine- α . Différentes structures ont déjà permis d'étudier cette interaction. L'import de PA et PB1 est moins connu, certaines études suggèrent un import commun via l'interaction avec une importine- β connue sous le nom de RanBP5.

Pour le but de comprendre le fonctionnement de PA-PB1 comme précurseur de la polymérase et des facteurs contribuant à son assemblage avec PB2, mon projet de thèse se focalise sur la caractérisation structurale et fonctionnelle de PA-PB1 et de RanBP5. Ces deux complexes protéiques ont été étudiés en tant que partenaires séparés et sous forme complexée afin enrichir les modèles d'assemblage actuels, mais aussi d'avancer dans la recherche de cibles potentielles antivirales.

1.1 GENERAL INFORMATION ON THE INFLUENZA VIRUS

1.1.1 INFLUENZA VIRUSES

Influenza viruses are segmented single stranded, negative sense RNA viruses belonging to the *Orthomyxoviridae* family (from the Greek: *orthos*, straight; *myxa*, mucus). According to the Baltimore classification established in 1971 (Baltimore, 1971), this family of viruses belongs to the group V of RNA viruses. Both classification and definition as a negative strand RNA genome were established from the work of David Baltimore who demonstrated that the genome from this group of viruses were complementary to the messenger RNAs (mRNAs) also defined as positive strand RNA. Within the group V classification is the order of *Mononegavirales*, which comprises non-segmented RNA genome virus families, in which we can find the *Bornaviridae*, *Filoviridae*, *Nyamiviridae*, *Paramyxoviridae* and *Rhabdoviridae* families. Other viruses such as the *Arenaviridae*, *Bunyaviridae*, *Ophioviridae* and *Orthomyxoviridae* families which all possess segmented RNA genomes are grouped within a second unassigned order.

The *Orthomyxoviridae* family comprises six genera (International Committee on Taxonomy of Viruses, <http://www.ictvonline.org/>):

- ***Influenza virus A***
- ***Influenza virus B***
- ***Influenza virus C***
- ***Thogotovirus*** which includes *Thogoto virus* and *Dhori virus*
- ***Isavirus*** which includes the Infectious Salmon Anemia virus (ISAV)
- ***Quarantavirus*** which includes the *Johnston Atoll virus* and the *Quarantavirus*

Until very recently, with the appearance of a new type of influenza virus, known as influenza D which infects cattle (Hause et al., 2014), three major types were categorised: influenza A, B and C. Their classification was based on their core protein content and their epidemiological characteristics (Ritchey et al., 1976a). Influenza A is classified by antigenic classes determined by two of its surface glycoproteins: the haemagglutinin (HA) and neuraminidase (NA). To date, at least 18 subtypes of HA (H1-H18) and 11 subtypes of NA (N1-N11) can be distinguished. This large reservoir of HA (15 out of 18) and NA subtypes (9 out of 11) can be

sourced mostly within waterfowl hosts (Treanor, 2004). The last described antigenic variants such as H17/H18 and N10/N11 were discovered within fruit bats from central and south America (Tong et al., 2013). Influenza B only presents one subtype of HA and NA whereas influenza C possesses only one kind of surface glycoprotein known as a haemagglutinin-esterase-fusion (HEF) (Herrler et al., 1979, 1981; Kendal, 1975). The Current World Health Organisation nomenclature (WHO, 1980) classifies the influenza virus with regards to type (A, B or C), original host (except for human strains), geographical point of origin, strain number, year of isolation and the antigenic subtype for influenza A strains (HxNx). For this thesis project, two major strains were used: A/Victoria/3/1975 (H3N2) and A/Vietnam/1203/2004 (H5N1); most of the introduction to this thesis therefore focuses on the influenza A subtype characteristics.

1.1.2 ZONOSIS OF INFLUENZA A VIRUSES

Influenza A virus differentiates itself from the two other types with regards to its very large host pool, which notably includes birds, humans, pigs, horses, dogs and bats (Vahlenkamp and Harder, 2006; Chan et al., 2013). They have also been shown to infect sporadically other host species such as those belonging to the *Felidae* family (cats, tigers, leopards) but also seals, whales and several rodent species (figure 1). Serological studies have also suggested that influenza A can occasionally infect several ruminant species, together with reptiles and amphibians (Vahlenkamp and Harder, 2006; Reperant et al., 2009; Herfst et al., 2014). Influenza A therefore has an important spectrum range within the animal kingdom compared with other viruses.

However, its main reservoir host pool is comprised of birds belonging to the *Anseriformes* (mallards, geese and swans) and *Charadriiformes* (sea gulls, waders, butonquails) orders (Kaplan and Webby, 2013; Yoon et al., 2014). In total, influenza A viruses have been isolated in at least 105 bird species spread out over 26 different families (Yoon et al., 2014). Practically every subtype of HA and NA has been detected within the *Anas* genus, also commonly known as the dabbling duck.

Within those natural hosts, most viral strains evolve very slowly, indicating that they are highly adapted to these species (Webster et al., 1992; Yoon et al., 2014). This adaptation process towards birds has been broken with domesticated poultry species, notably in the mid 1990s with the emergence of highly pathogenic H5N1 strains (Shortridge et al., 1998; Shortridge, 1999). Aside from affecting poultry farms, these epidemics also led to a minority of avian influenza cases to be

transmitted to humans with often fatal consequences (de Jong et al., 1997; Claas et al., 1998a, 1998b; Subbarao et al., 1998).

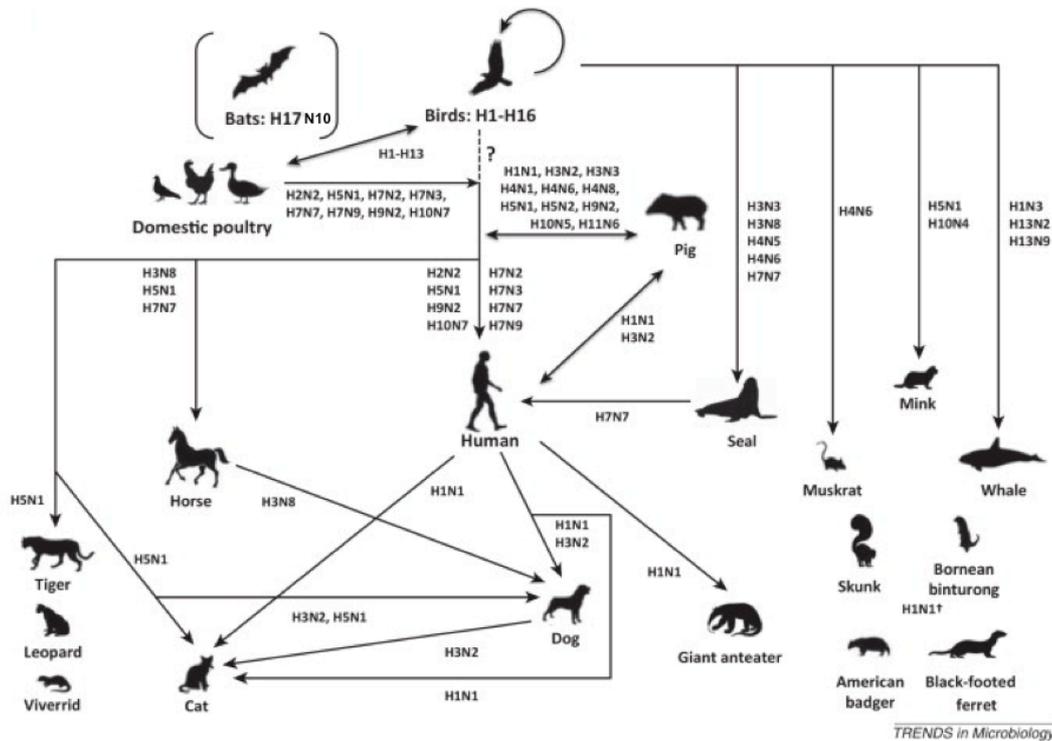


Figure 1: Host distribution of type A influenza stains. (Adapted from Chan et al., 2013)

Phylogenetic analyses, coupled with HA and NA subtype detection within the natural waterfowl species suggest that all current mammalian influenza A viruses originate from avian influenza species (Gorman et al., 1990a, 1990b, 1991; Webster et al., 1992; Yoon et al., 2014)

Flu B infects almost exclusively humans and is responsible with type A of seasonal influenza epidemics. Influenza C is essentially present within humans and provokes mild respiratory infections. All A, B and C viruses are probably related through a common ancestor (Wright et al, 2013).

1.1.3 HISTORICAL BACKGROUND

Flu viruses are unique with regards to their ability to generate both seasonal epidemics with variable intensities and pandemics. It is through the combination of epidemic and pandemic episodes that the influenza virus can be traced back throughout history.

Historical traces of influenza can be traced back to Late Antiquity. Early accounts in 412 BC by Hippocrates described comparable symptoms to those witnessed today during an influenza infection. Comparatively, several accounts of pandemic disease propagation have been attributed to influenza during the 16th and 17th centuries. One such account, dating from 1510 described an epidemic of “sweat” (*Sudor Anglicus*; “English sweating sickness”) which could plausibly be attributed to influenza (Morens et al., 2010). Sero-archeological studies have confirmed influenza pandemics during the 19th century, most notably the one of 1889 (also known as the “Russian flu”) which was able to spread worldwide within four months despite the much slower and more limited traveller flow of the time (Valleron et al., 2010). With the progress of medicine and especially with the advent of modern virology, researchers were able at the beginning of the 20th century to isolate influenza virions. In 1930 the first influenza virus was isolated (A/Swine/Iowa/30) from swine (Shope, 1931). Then in 1933 the first human strain was isolated (Smith et al., 1933) and categorised as a type A influenza. Seven years would pass before the characterisation of a new virus with different antigenic properties, classified as an influenza B virus (Francis, 1940). Influenza C was later discovered in 1947 (Taylor, 1949).

1.1.4 SEASONAL INFLUENZA

From an epidemiological standpoint, seasonal epidemics of influenza occur during the winter months within the temperate regions of the planet, thus generating two epidemic peaks per year worldwide. In the northern hemisphere, the epidemic peak can be observed from December and February. In 2011 in the USA, the highest viral activity was recorded on the 5th week of the year (Figure 2) (CDC, 2012). Since 1977 and the re-emergence of the H1N1 virus from the 1918 pandemic, seasonal strains of influenza (i.e., A/H1N1 and A/H3N2) circulate together with influenza B (World Health Organisation; Wright, P.F. et al, 2013). As of 2009, the seasonal epidemic strain of influenza H1N1 has been replaced by the adapted pandemic strain H1N1 originated from the “swine flu” episode of that same year (Broor et al., 2012, Wright et al, 2013). The distribution prevalence of both strains varies geographically making epidemiological studies complex to model. Also, epidemic influenza follows a slow but constant evolution process known as “antigenic drift” (detailed in chapter 1.3.1) which drives seasonal influenza strains to continually change in antigenic signatures in order to remain infectious within otherwise immunised populations.

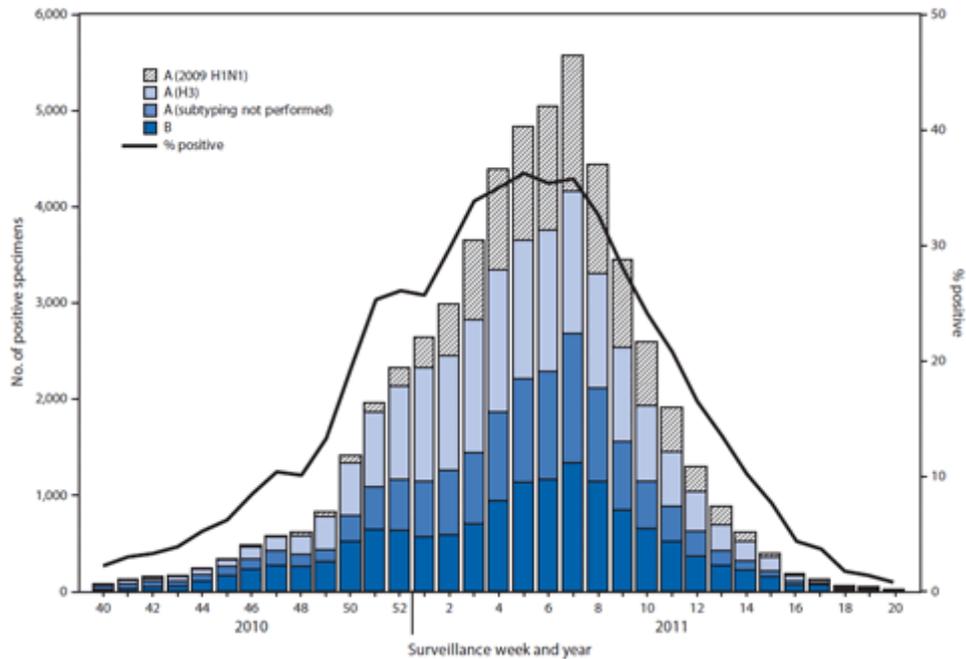


Figure 2: Number and percentage of respiratory specimens testing positive for influenza in the year 2010/2011, by type and week. (Centre of Disease Control and Prevention, 2012).

Antigenic and genetic studies undertaken for several years have highlighted the common geographical source of the epidemic strains H3N2 and H1N1 as originating from South East Asia (Russell et al., 2008; Chan et al., 2010; Bedford et al., 2010). South East Asia, by its geographical situation at the crossroads of important commercial routes and with its high population density maintains a continuous epidemic linkage between the Northern and Southern hemispheres of the globe. Climate plays an important role in the epidemic spread of influenza and explains this situation. Tropical parts of the world such as South East Asia have mild seasonal variations and epidemic strains of influenza are active all year long in these regions of the world (Viboud et al., 2006; Russell et al., 2008). Continental climates, on the other hand, generate important humidity and temperature variations during the year. These variations limit or enhance the air transmissibility (Noti et al., 2013) and surface survivability (McDevitt et al., 2010) of the virus, generating a peak distribution throughout the year as shown for the USA (figure 2). Human transmission of the virus is essentially the effect of inhalation of expelled water drops from the coughs of infected people and direct contact with contaminated surfaces. It has therefore also been proposed that an increased crowding during the cold seasons could enhance the transmissibility of the virus (Souza et al., 2003). Lastly, another factor that can also explain the seasonality of the influenza are variations in the host immune systems. In this regard, vitamin D levels —whose metabolism is correlated with ultraviolet B exposure— may affect the onset of influenza infections (Cannell et al., 2006).

From a clinical standpoint, seasonal influenza generates a mild or benign illness within the majority of otherwise healthy humans. Non-specific symptoms, such as strong fever, dry coughs, muscular and joint soreness, headaches and asthenia characterise the infection. On average, these symptoms persist 2 to 7 days. Response variation between patients is influenced both by the strain virulence and the innate immune and inflammatory response that is responsible for the infection symptoms (Junge, 2011). In the most severe cases, the initial infection can trigger complications such as bacterial secondary infections resulting in pneumonia and/or severe dehydration that can lead to death. The most sensitive age groups towards seasonal influenza are children below 5 years of age (YOA) and elderly people above 65 YOA. Virus shedding has been shown to precede the appearance of symptoms by one day in clinically induced infections (Carrat et al., 2008); on average viral shedding will peak on day 2 and then decrease to finally end after 5 days. In some cases the shedding period can extend to 9 days. It has also been shown that immune-deficient patients, elderly people and children can have extended viral shedding, lasting in some cases for several weeks (Mitamura and Sugaya, 2006; Gooskens et al., 2009).

Although difficult to precisely quantify, the consequences of seasonal epidemics in the world are quite important. Albeit socially accepted as a “norm”, seasonal influenza is still responsible every year of important economic and human losses. WHO estimates that every year, the combination of both Northern and Southern hemisphere epidemics account for 250000 to 500000 deaths and 3 to 5 million cases of severe illness requiring in many cases hospital treatment (WHO, 2015). In France, influenza kills between 1500 and 2000 people a year and infects an average of 2.5 million people (INSERM, 2012). In 2015, a strong epidemic of H3N2 influenza has been shown to correlate with a 19% increase in mortality (8000 additional fatalities) during the winter period in France (InVS, 2015).

It is estimated that the economic burden generated by seasonal influenza is also considerable. In an economic study undertaken in France for the 1988/1989 epidemic (Levy, 2012), research estimates a net direct cost of seasonal influenza of 2.2 billion euros for the French social security system in today's currency value. Added to this is the indirect social cost of 16.6 billion euros attributed to the productivity loss. When compared to today's Gross Domestic Product (GDP), 18 billion euros represents 0.6% of the French GDP (2,8 Trillion euros for the year 2013). It is indeed estimated that every year, in France, between 2 and 12 million days of work are lost due to the influenza (GROG).

Although an important burden, both socially and economically, the emergence of seasonal influenza follows a regular and repetitive pattern which enables for preventive approaches to be put into place. This pattern has been broken in the past and will be broken in the future by occasional antigenic jumps, also known as “antigenic shifts” (detailed in chapter 1.3.2) which enable the appearance of Pandemic influenza strains.

1.1.5 PANDEMIC INFLUENZA

When looking at catastrophes, the twentieth century is still remembered for its two major conflicts, World War I and World War II which claimed the lives of around 70 to 90 million people in total (Micheal Clodfelter, 2002). Little is remembered though, that in the space of one year (1918-1919), 50 million to arguably 100 million people lost their lives to the “Spanish influenza” pandemic (Johnson and Mueller, 2002; Patterson and Pyle, 1991). This H1N1 outbreak remains the most severe influenza pandemic humanity ever experienced. At least 30% of the world population was infected with the virus which had an estimated case fatality rate > 2.5% (Taubenberger and Morens, 2006). Of course one cannot impute the death toll exclusively to the virus, specific factors of the time came into play. The pandemic onset started near the end of World War I and maintaining hygienic conditions in the front lines was difficult; also crowding due to troop gatherings enabled rapid contagion pockets to emerge. Malnutrition due to rationing and overall stress of war could also have been factors of aggravation (Song, 2014). Most strikingly was the fact that most deaths were the result of a consistent secondary bacterial infection causing respiratory pneumonia (Morens et al., 2010; Sheng et al., 2011). Contrary to seasonal influenza, the “Spanish influenza” pandemic was also very virulent within the young adult (20-35 YOA) age group (Simonsen et al., 1998; Frost, 2006). The mechanisms underlying such a specific virulence are still not fully understood, it has however been suggested that the genesis of the virus which resulted in the reassortment of previous H1 pandemic strains and an Avian strain possessing the N1 gene may be linked (Worobey et al., 2014). People born before 1890-1900 may have acquired partial immunity to the virus through past infections with precursor influenza strains. Another explanation for such a distinct age distribution may lie in the nature of the immune response at that age: it has indeed been proposed that deregulated cytokine and chemokine responses within young adults could be one of the reasons for such a strong pathogenicity (Loo and Gale, 2007; Gao et al., 2013). Remarkably, the pandemic followed in some countries like England, with three distinct epidemiological bursts of decreasing intensity (morbidity and fatality rate) in time and in geographical spreading, hinting a rapid adaption mechanism of the virus

to its human host (Taubenberger and Morens, 2006). In the years following 1919, the H1N1 “Spanish influenza” became seasonal influenza virus, with a normal age/pathogenicity distribution. This strain remained in circulation through the “antigenic drift” process until 1957 where it was replaced by the H2N2 subtype. It however later re-emerged in 1977 through what is believed to be an accidental release of a conserved virus within a laboratory. Genetic analysis have indeed revealed that the strain is closely related to the pre-1957 H1N1 strain (Nakajima et al., 1978; Scholtissek et al., 1978).

Later pandemics emerged during the twentieth century. In 1957, an H2N2 reassortment known as the “Asian influenza” originated from China, spread through Asia and caused the deaths of an estimated 1 million people worldwide. The United States were strongly hit by the pandemic, causing around 70 000 deaths, mostly among elderly people (Simonsen et al., 1998). Vaccination, which came into practice within the industrialised countries after the second world war was produced against the H2N2 virus from 1957 to 1960. The virus eventually disappeared from the human reservoir in 1968. Concern however persists as only individuals above 50 years old remain immunised against the strain (Nabel et al., 2011) which still circulates within wild and domestic animals (Ma et al., 2007). In 1968, a new strain emerged from the reassortment of A/H2N2 and A/H3 strains which produced the A/H3N2 strain also known as the “Hong Kong” pandemic. This pandemic which was severe in Asia, claimed the lives of an estimated 700 000 people and came to replace A/H2N2 as the dominating seasonal influenza strain in the later years (Simonsen et al., 1998).

With the end of the 20th century and the beginning of the 21st century came new fears of pandemic influenza. In 2009, a social, political and media hype was triggered by the fast emergence of a novel H1N1 virus generated from the reassortment of bird, swine and human influenza A viruses (Garten et al., 2009; Trifonov et al., 2009). Considerable physical and financial means were invested to contain the spread. It is estimated that in France, 1.8 to 2.2 billion euros were invested to prevent the pandemic spread alone, by means of vaccines, antivirals, breathing masks and communication campaigns (Sénat Français, 2011). An entire medical infrastructure was put into place worldwide and coordinated by national (Ministère de la santé, in France) and international institutions (WHO, Center for Disease Control) to cope with the ongoing crisis. Fortunately, this influenza strain, despite having a very high transmissibility, presented a mild pathogenicity towards humans. After one year, it was estimated that in France, 8 to 15 million people were infected by the first wave of H1N1 pandemic influenza; in total 1300 severe infections were reported with 312 confirmed deaths due to the virus (INSERM, 2010). Comparable to the situation in 1918, young adults within the 20-30 years of

age group were more afflicted by the pandemic strain compared to seasonal strain epidemics. Personal medical histories such as chronic respiratory diseases and obesity were linked as being aggravating factors (CDC, 2010; INSERM, 2010).

On the sidelines of these pandemic episodes, short, sporadic bursts of highly pathogenic viruses have been witnessed, especially in Asia. These were reported on several occasions starting in the late 90's (1997) and shown to be caused by an avian influenza A/H5N1 virus (Claas et al., 1998b; Peiris et al., 2004). In most cases transmission was provoked by direct contact with contaminated birds or contaminated bird meat. Some rare cases of human to human transmission have however been documented (Ungchusak et al., 2005; Normile, 2006). More recently, H7N9 strain has been shown to infect humans as well, also with high pathogenicity. For all these episodes, only a very limited number of cases were reported, 718 cases for A/H5N1 since 2003 with a total of 413 deaths (WHO) placing the case fatality rate > 50%. These new emerging viruses are the cause of great concern, although they are weakly transmissible for the moment. A major fear is that with time, these viruses may mutate and/or reassort with other viruses to generate both highly pathogenic and transmissible strains. Several independent laboratories drew up controversy in 2012 by either modelling (Russell et al., 2012) or by proving in ferrets that certain mutations can lead to an airborne transmissibility of the H5N1 virus (Herfst et al., 2012; Imai et al., 2012).

1.1.6 CURRENT TREATMENT STRATEGIES

Immunisation, through the use of vaccination, is currently the most efficient method of preventing the onset of an influenza infection. Early studies in the aftermath of the "Spanish influenza" pandemic showed that using viruses grown *in vitro* could drastically increase the vaccinee's antibody titer (Thomas and Magill, 1936). In the 1940's the first large scale trials of whole-virus inactivated vaccines were undertaken on military recruits and college students (Stokes et al., 1937; Salk et al., 1945). By 1945 a commercial vaccine was available in the United States and applied by intra-muscular injection. Whole-virus vaccines were prepared using inoculated eggs to grow the virus. After extraction, the viruses were inactivated by using formalin or β -propiolactone (Fiore et al., 2009). Although less immunogenic, most vaccines manufactured after 1970 were solvent disrupted subvirion preparations. Today, trivalent combinations of A/H3N2 A/H1N1 and B strains are manufactured every year in order to counter the "antigenic drift" of Haemagglutinin and Neuraminidase surface proteins of seasonal influenza. The WHO issues virus formulations directed at the preparation of the vaccine; these are based on the

observed strains of the seasonal influenza affecting the opposite hemisphere. Therefore, every year two separate formulations are issued. Depending on the “antigenic drift” outcome, these vaccines are more or less efficient. For seasonal influenza, the cost benefit ratio of vaccination is profitable within the very young (6 months to 5 YOA), older age groups (> 65 YOA) and individuals presenting a high-risk medical background such as immunocompromised patients. One of the main downfalls of vaccination is its inherent incapability to protect against sudden pandemic strain outburst. This is in part due to the current impossibility of predicting the outcome of antigenic shift, but mostly because of the necessarily long production period (6 months) required for vaccine production. In this rather long period, other means to actively fight the virus are needed. For these reasons, a lot of effort and means have been employed to develop antiviral inhibitors.

To date, there are only four commercially available influenza inhibitors, which can be divided into two groups:

- M2 ion channel inhibitors (Table 1) developed during the mid-sixties (Davies et al., 1964; Dolin et al., 1982) known as Amantadine and Rimantadine and today commercialised respectively under the names Symmetrel (Endo pharmaceuticals) and Flumadine (Forest pharmaceuticals). These chemical compounds are amine derivatives of the Adamantane class, which intercalates within the M2 ion channel and blocks the H⁺ proton influx required viral disassembly (see chapter 1.2.2). These drugs were previously used exclusively as a prophylaxis treatment and were only effective against influenza A viruses. They are however largely out-dated due to the rapid emergence of resistant influenza A strains (Belshe et al., 1989; Hayden and Hay, 1992). By the mid 2000s, 99% of H3N2 seasonal strains were resistant to this inhibitor class (CDC, 2006).
- Neuraminidase inhibitors (Table 1) known as Zanamivir and Oseltamivir are commercialised respectively under the names of Relenza (GlaxoSmithKline) and Tamiflu (Roche). These compounds are also the product of years of research and derive from sialic acid analogues which were first investigated in the late 1970s (Schulman and Palese, 1975). Their mode of action is to target the neuraminidase active site, by sterically outcompeting the interaction with the sialic acid present on the infected cell membrane. Neuraminidase activity is required to cleave the sialic acid moieties involved in the release of newly assembled virions (see chapter 1.2.2). Due to high conservation of the neuraminidase active site, these inhibitors present a broad spectrum activity in most strains of influenza A and B viruses (Yen et al., 2006). Since the 90s

these drugs have been used as a prophylaxis and active treatment of both seasonal and pandemic influenza with varying efficiency depending greatly on the virus strain and the infected individual. Although much more potent than M2 ion channel inhibitors, neuraminidase inhibitors present more severe side effects (Dutkowski et al., 2003). Resistant strains of influenza have also appeared for these drugs with much lower frequencies than for M2 ion channel inhibitors (Samson et al., 2013), due to the inherent loss of infectivity of NA mutated viruses. Still, it has recently been reported that strains resistant to both Oseltamivir/Amantadine are emerging (Sheu et al., 2011).

Due to the limited scope of action provided by these few inhibitors and considering the rapid and unavoidable rise of poly-resistant strains, new strategies to inhibit the influenza virus proliferation are in the works (Monod et al., 2015). Several new inhibitors are in phase III clinical trials in the USA or clinically approved within different countries. Peramivir and Laninamivir, novel NA inhibitors have shown promising potential and remain potent on Oseltamivir resistant strains (Babu et al., 2000; Koyama et al., 2009; Samson et al., 2013).

Broad spectrum antivirals are also used to treat the influenza such as Arbidol, a haemagglutinin inhibitor (Teissier et al., 2011) whose use is only approved in China and Russia. Ribavirin, a nucleoside analogue mostly used against Hepatitis C (HCV) has been shown to be effective against influenza. It is however responsible of important secondary effects such as haemolytic anaemia. Favipiravir (Furuta et al., 2005), a nucleoside analogue, is approved in Japan for the treatment of influenza. It was also shown to be effective against other negative-stranded RNA viruses including Ebola virus (Oestereich et al., 2014).

Novel druggable targets have also emerged in the last ten years. In the future, multiple viral pathways will be blocked in order to limit the escape mechanism provided by the high mutational rate of the virus. Within this current strategy, several groups including the one in which I performed my thesis, have focused on the RNA dependent RNA polymerase (RdRp). This enzyme could constitute a major drug target due to its pivotal role in the viral replication process, but also because of its role in enabling a high mutation rate which is required for evading viral inhibitors.

Further reading on the topic can be provided by the **supplementary publication n°2**.

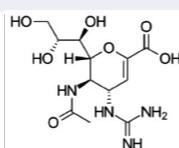
Table 1: General overview of currently used influenza inhibitors. (Compound depiction were adapted from Monod et al., 2015)

Target Protein	M2 ion channel	Neuraminidase	Haemagglutinin	RNA dependant RNA Polymerase
Targeted Functions	<ul style="list-style-type: none"> M1 and vRNP dissociation during viral uncoating 	<ul style="list-style-type: none"> Viral release 	<ul style="list-style-type: none"> Viral interaction and endocytosis within host cell 	<ul style="list-style-type: none"> Replication Transcription
Inhibitors	Adamantanes derivatives	Sialic acid derivatives	Neutralizing IgG antibodies Arbidol	Nucleoside analogues

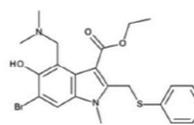
Approved
therapeutical use



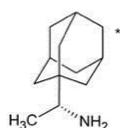
Amantadine



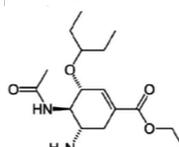
Zanamivir



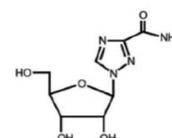
Arbidol



Rimantadine



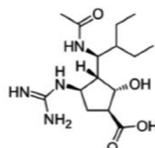
Oseltamivir



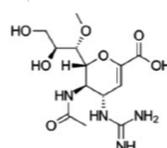
Ribavirin

* Only in Russia and China

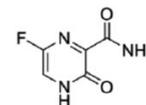
In phase III human
clinical trial



Peramivir



Laninamivir



Favipiravir

1.2 STRUCTURAL FEATURES AND INFECTIOUS CYCLE OF THE INFLUENZA A VIRUS

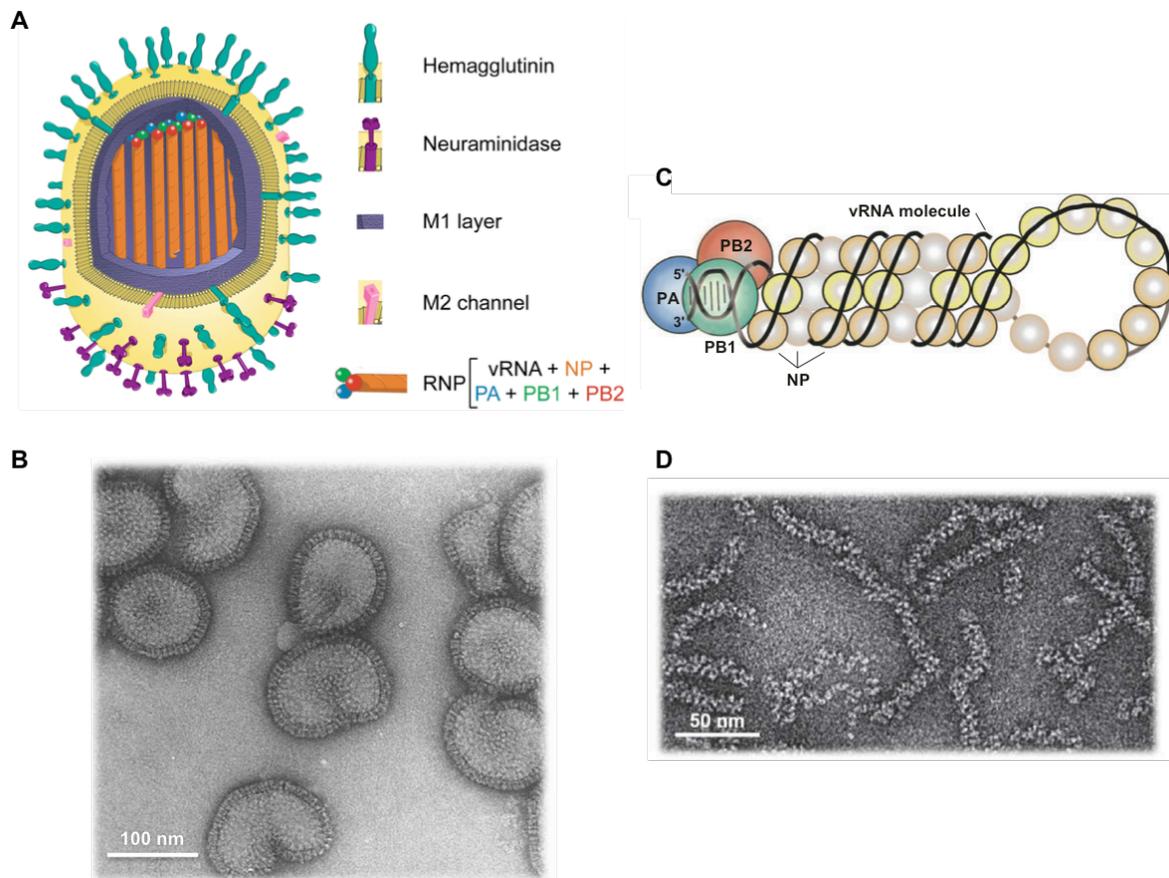


Figure 3: Influenza A viral architecture. (A) Schematic view of the virion with core components. (B) Negative stain EM picture of virions (Rob Ruigrok). (C) Schematic representation of the RNP assembly (adapted from Naffakh et al, 2008). (D) Negative stain Electron Microscopy (EM) picture of purified RNPs. (Adapted from Ruigrok et al., 2011)

1.2.1 STRUCTURE OF THE VIRAL PARTICLE

The influenza A virus is an enveloped virus; virions are pleomorphic but usually adopt a spheroid shape measuring approximately 100 nm in diameter (Fujiyoshi et al., 1994), although much bigger virions have also been reported (Chu et al., 1949; Kilbourne and Murphy, 1960). The envelope is derived from the host cell membrane which is specifically enriched in cholesterol and glycosphingolipids (Scheiffele et al., 1999). In this membrane are inserted HA and NA surface glycoproteins and the M2 ion channel protein (Figure 3A). These surface glycoproteins can easily be seen on a negative stain electron microscopy image of full viruses (figure 3B). They can also be distinguished by shape: NA adopts a “nail” shape and HA a “spike” shape. They are generally distributed as 1 NA for 4 HA (Murti and

Webster, 1986). Inside the virion, the M1 (matrix) protein scaffolds the virion and is at the interface between the lipidic membrane and the ribonucleoproteins (Sha and Luo, 1997). In addition, the Nuclear Export Protein (NEP, previously described as NS2) is found within purified virion (Richardson and Akkina, 1991; O'Neill et al., 1998).

Within the viral core lies the segmented, negative and single stranded (ss) RNA genome divided into 8 separate segments. Each vRNA is bound to multiple copies of nucleoprotein (NP) and the highly conserved 5' and 3' extremities are bound to the RdRp complex heterotrimer composed of PA, PB1 and PB2 proteins. This macromolecular assembly (vRNA/polymerase/NP) forms the ribonucleoprotein (RNP) complex (Figure 3C and 3D) which is observed as a closed helical structure (Pons et al., 1969; Compans et al., 1972).

1.2.2 SEGMENTED GENOME ORGANISATION

One of the distinct characteristics of the influenza virus is its genomic organisation. Differently from other negative and single strand RNA viruses, influenza has a segmented division of its genome. This distinct feature enables a complex reassortment process between strains (detailed in chapter 1.3.2). Influenza A and B possess eight vRNA segments (Palese and Schulman, 1976; Ritchey et al., 1976), differently from influenza C which only possess seven. Segments vary in size and are ordered from the longest (**segment 1**, 2341 nucleotides) to the shortest segment (**segment 8**, 890 nucleotides) for influenza A (figure 4). Each segment codes for a minimum of one protein; to date, segments 1, 4, 5 and 6 are known only to code for one protein. Other segments code for both “canonical” genes expressed with the original open reading frame (ORF) and additional proteins, which are either the product of a frame-shift reading (ORF+1) and/or produced after splicing of the mRNA. Overall, the combined segments can express 10 different proteins found in all viruses (Figure 4). Additionally, 7 accessory proteins have been described to exist in many strains. All of the segments share a common nucleotide (nt) sequence pattern. ORFs are flanked on both 5'tri-phosphate (5') and 3'-hydroxyl (3') extremities by untranslated regions (UTR) of 19 to 58 nt in length. These UTR regions harbour both the 5' and 3' terminal promoters measuring 12 and 13 nt in length are both fully conserved among the 8 segments, but also among all influenza A subtypes. Due to their *quasi* complete complementarity, these promoter regions have been shown to form a distinct secondary structure conformation called the panhandle when alone in solution (Baudin et al., 1994; Hsu et al., 1987). Following these promoter regions are segment specific UTR regions which have been shown

to be important in the regulation of viral gene expression and in packaging RNP within the progeny virions (Luytjes et al., 1989; Fujii et al., 2003, 2005).

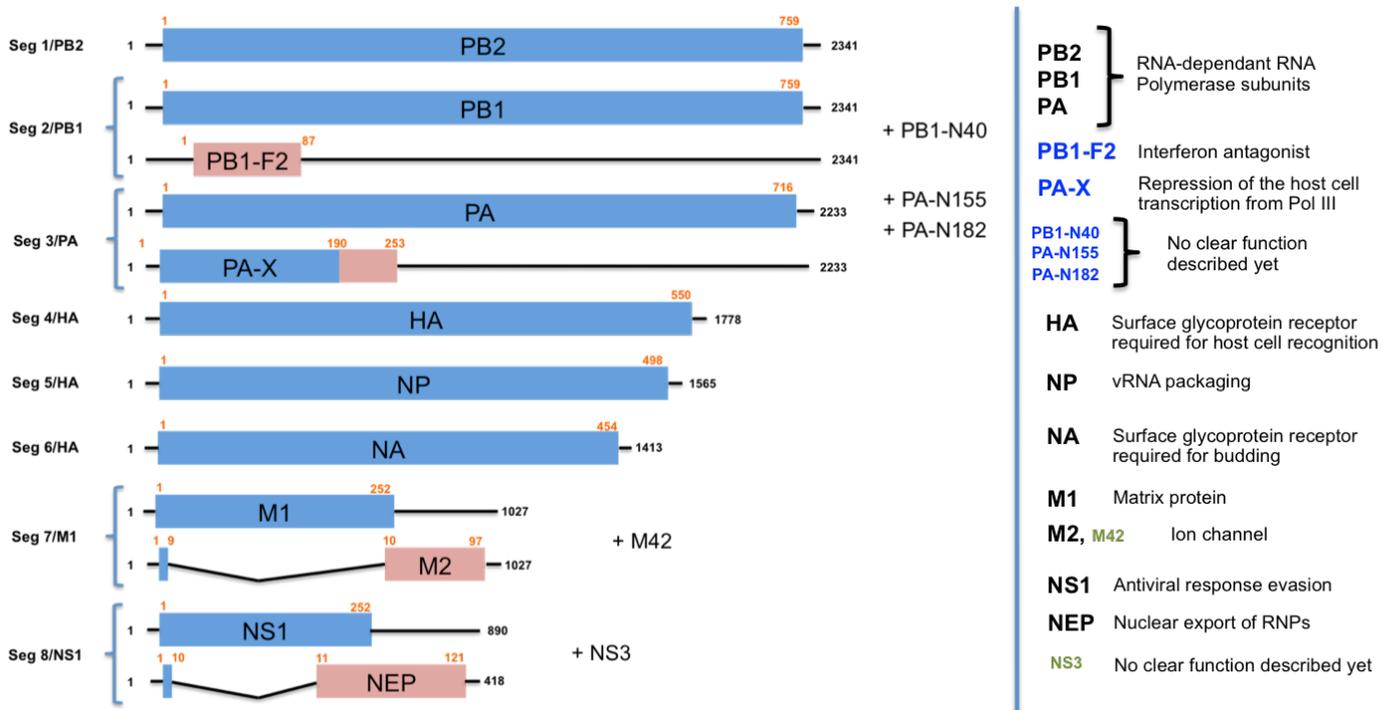


Figure 4: Organisation of the influenza A genomic segments and the respective coded viral proteins once transcribed into positive sense RNA. Segment length, in nucleotides, is numbered in black whereas protein length, in amino acid residues, is numbered in orange. Canonical ORF peptides are coloured blue, frame shift ORFs are coloured in salmon. Spliced-out sequenced are represented with the V shape. In the right column core proteins are spelled in bold black, abundantly distributed accessory proteins in blue and rare accessory proteins in green. (Adapted from Wright et al., 2013)

1.2.2.A SEGMENT 1

Segment 1 is only known to code for **PB2** (759 aa), which is one of the 3 subunits constitutive of the RdRp whose function and structure is discussed in **chapter 1.4**.

1.2.2.B SEGMENT 2

Segment 2 primarily codes for the expression of PB1 (757 aa), another subunit of the RdRp responsible of the polymerase activity (see **chapter 1.4**). In addition two auxiliary proteins **PB1-F2** and **PB1-N40** (Wise et al., 2009).

PB1-F2 is a ≥ 78 aa peptide expressed from an alternate ORF whose discovery dates back to 2001 (Chen et al., 2001). This peptide has not been shown to integrate within the virions constitution (Krumbholz et al., 2010) but seems to play

multiple roles during infection. Research has shown a modulation of the viral polymerase activity (Mazur et al., 2008) through the interaction with the PB1 protein. It has also a pro-apoptotic activity within infected lymphocyte B cells through the disruption of mitochondrial membranes and the release of cytochrome C, potentially contributing to the evasion of the virus from the host immune response (Chen et al., 2001; Henkel et al., 2010). PB1-F2 is also extensively studied for its pro-inflammatory activity induced by its modulation of cytokine and interferon expression (Le Goffic et al., 2010; Henkel et al., 2010; Varga et al., 2012). The strain distribution of actively folded PB1-F2 is mostly found within avian influenza subtypes (Chakrabarti and Pasricha, 2013). The loss of biologically functional PB1-F2 seems to be beneficial in the adaptation to mammalian hosts (Zell et al., 2007; Dundon, 2012), including humans (McAuley et al., 2010).

As of 2009, an amino terminal (N-terminal) truncated version of PB1 (**PB1-N40**) originating from an alternate AUG initiation site has been described (Wise et al., 2009) and seems important for viral proliferation (Tauber et al., 2012). Although able to interact with the polymerase complex, the role played by this protein remains elusive, as it has been shown to lack a transcriptase activity (Wise et al., 2009).

1.2.2.C SEGMENT 3

Segment 3 harbours the gene coding for PA (716 aa), the third subunit of the RdRp (detailed in chapter 1.4) that possesses the endonuclease domain. Also expressed is PA-X (253 aa), this protein discovered in 2012 (Jagger et al., 2012), is the combination of both an N-terminal part of PA (residues 1-190) and an extra carboxy terminal (C-terminal) domain expressed from a shifted reading frame, called the X domain (res 190-253). In this particular case, the shift originates from the RdRp slippage on a UUUGGUC motif during transcription (Vasin et al., 2014). The X domain is highly conserved among influenza A subtypes suggesting an important role in the viral life cycle (Shi et al., 2012). The functional roles played by PA-X seems to be that of host response modulation during infection, through the degradation of host messenger RNAs (mRNAs), specifically those transcribed by the RNA polymerase II (Jagger et al., 2012). In addition to PA-X, PA-N155 and PA-N182 are also expressed by downstream alternate AUG initiation codons (Muramoto et al., 2013). In essence, these proteins are C-terminal domains of PA without the endonuclease domain. As later discussed, this C-terminal domain is tightly bound to the PB1 subunit (see chapter 1.4.3) hinting a potential alternate polymerase assembly with these proteins. Little is known of their function except that their deletion generates slower replicating viruses (Muramoto et al., 2013).

1.2.2.D SEGMENT 4

Segment 4 houses the gene coding for the **haemagglutinin** surface glycoprotein (HA), most important for the specific viral recognition/attachment to the host cells and membrane fusion process. Haemagglutinin is a class I viral fusion protein, a class which also contains the HIV GP41 receptor and the Ebola virus GP2 receptor (Wright, P.F. et al, 2013). The protein is first synthesised in a precursor form HA0 which is then cleaved by host cell proteases. This cleavage enables the formation of two ectodomains HA1 (N-terminal domain of HA0) and HA2 (C-terminal domain of HA0); this step is required to expose the hydrophobic HA2 ectodomain whose function is to enable virus/endosomal membrane fusion (Weissenhorn et al., 2007; Gamblin and Skehel, 2010). Haemagglutinin adopts a trimeric oligomerization state on the viral surface (**figure 5A**).

HA1 forms a globular domain whose role is to recognize and interact with sialic acid glycans. Research has shown a viral strain prevalence towards certain glycan signatures, human influenza strains interact with sialic acid bound to galactose with an $\alpha 2,6$ (S $\alpha 2,6$) linkage whereas avian strains interact with S $\alpha 2,3$ motifs (Connor et al., 1994; Viswanathan et al., 2010). Both motifs can be found in humans but within different cellular compartments. S $\alpha 2,6$ is found in the upper respiratory track epithelial cells whereas S $\alpha 2,3$ is found within the alveoli epithelial cells.

HA2 adopts characteristic stem architecture with a triple α -helix segment linked to HA1. Upon the proton influx increase in the endosomal vesicle during endocytosis, the HA2 domains within the trimeric context are submitted to multiple phases of structural changes which result in pulling of viral and endosomal membranes together leading to membrane fusion (**figure 5B**). It has been reported that a minimum of three haemagglutinin trimers are required to initiate membrane fusion (Danieli et al., 1996; Floyd et al., 2008).

HA is also directly involved in the budding process, like NA through the coalescence of lipid raft domains generating favourable budding zones (Schmitt and Lamb, 2005).

Finally, HA is one of the two main (with NA) surface antigens of influenza. During infection, the entering virus HA, especially the HA1 domain will stimulate neutralising antibody production by B-lymphocytes. These antibodies will then apply a selective pressure on the progeny virions after replication in the host cell. By actively pressuring the “antigenic drift” process, variations in HA epitopes are

important features considered in vaccine design (Smith et al., 2004; Wan et al., 2014).

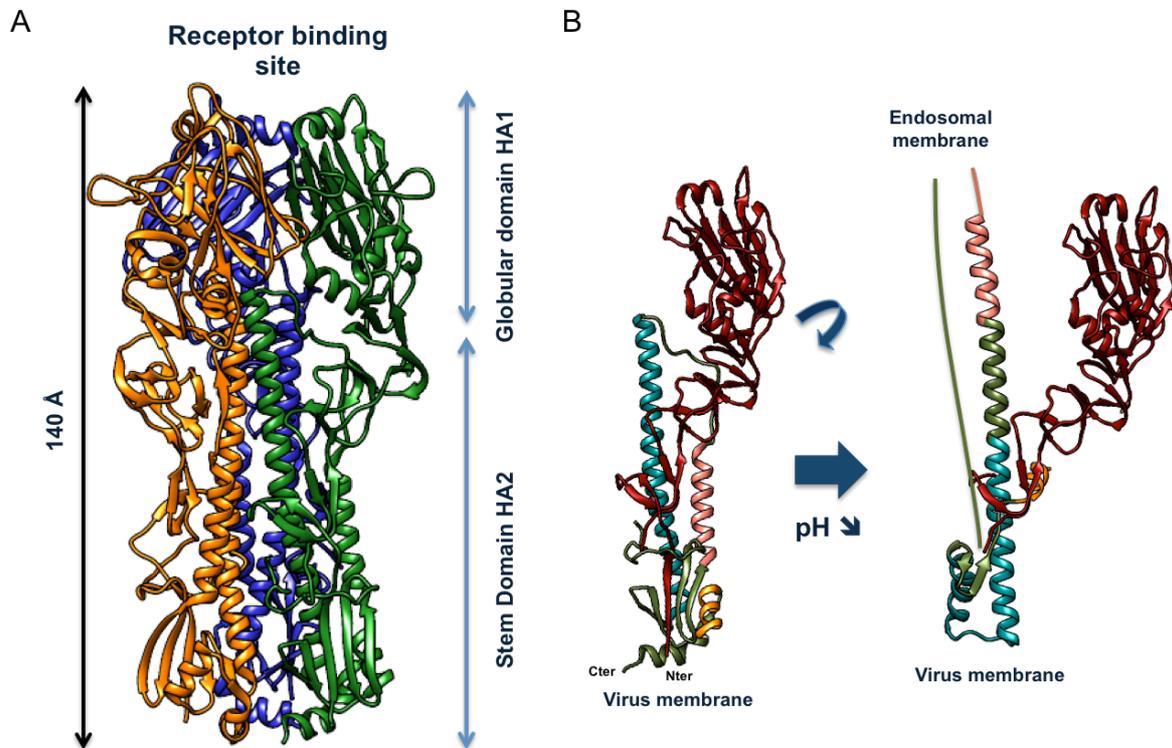


Figure 5: Structural overview of haemagglutinin from an H3N2 influenza strain. (A) Trimeric oligomerization form (PDB 2YPG) featuring yellow, blue and green monomers. (B) Monomeric depiction of the structural changes upon acidification depicted with the HA1 ectodomain in red and the HA2 ectodomain in rainbow colours distinguishing the α -helices and their conformational changes from the pre-fusion state to the late intermediate state (PDB 1HTM). (Depictions were generated with Chimera)

1.2.2.E SEGMENT 5

Segment 5 codes for the **nucleoprotein** (NP) (498 aa), the most abundant protein component of the RNP assembly (Pons et al., 1969). NP plays multiple roles within the RNP complex: it structures the RNP complex into a double helix (Heggeness et al., 1982; Jennings et al., 1983; Ruigrok and Baudin, 1995), by coating most of the vRNA and complementary (+)vRNA (cRNA), except the 5' and 3' promoter regions which interact directly with the RdRp. NP is proposed to modulate both replication and transcription of the vRNA (Beaton and Krug, 1986; Shapiro and Krug, 1988; Naffakh et al., 2008) and could also be the key regulatory element in the switch between replication and transcription (Skorko et al., 1991). It has been shown that the residue D88 is directly involved in the binding process to PB2 (Gui et al., 2014) which corroborates previous works which showed the N-terminal region of NP

to interact with PB2 (Albo et al., 1995; Biswas et al., 1998; Turan et al., 2004). NP is also proposed to interact with PB1, M2 and a number of host cell proteins including RAF-2p48/UAP56 and Tat-SF1 (Portela and Digard, 2002).

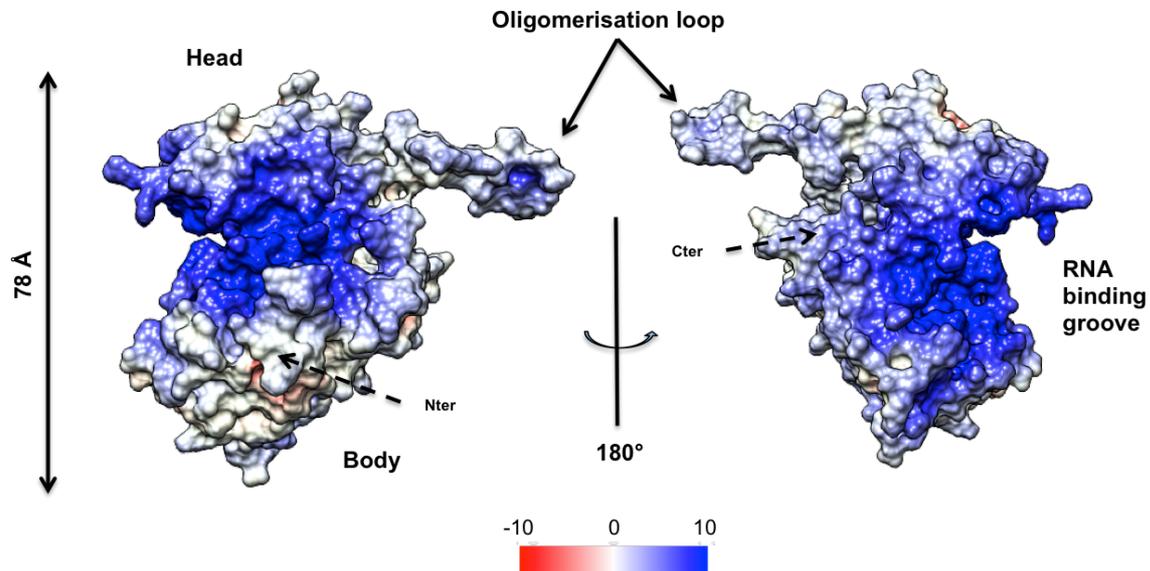


Figure 6: Surface representation of H1N1 NP (2IQH) coloured by predicted polarity using. Blue indicates positively charged regions whereas red indicates negatively charged regions. (Depictions were generated with Chimera)

A total of four crystallographic structures have described the influenza NP, 3 of influenza A (Ye et al., 2006; Ng et al., 2008; Chenavas et al., 2013) and one of influenza B NP (Ng et al., 2012), all of them without RNA. Interestingly, most of the structures were crystallised as trimers for influenza A and tetramers for influenza B highlighting the inherent tendency of NP to oligomerize in the absence of vRNA. This property is enabled by the oligomerization loop (res 402 to 428), which performs a domain swap with the neighbouring NPs. This oligomerization process was broken by mutating the residue R416A in influenza A and generated a monomeric version of NP (Chenavas et al., 2013). This oligomerization property has been described as dynamic and influenced by salt content (Tarus et al., 2012). Monomerised NP can then form higher order oligomers with the vRNA. On the other hand, the loss of the oligomerization of NP mutants causes ineffective protein translation in the mini RNP model (Ye et al., 2012). The entrapment of NP in a non-functional oligomeric state by the use of compounds has been shown to block the replication of the influenza virus in infected cells (Kao et al., 2010; Gerritz et al., 2011).

Another potential regulatory element is phosphorylation. Early on studies have shown different patterns of phosphorylation within NP depending on the influenza strain and time of infection (Kistner et al., 1985, 1989). More recently it was shown

that mutating key phosphorylated residues (S165, S407, S486) could disrupt both oligomerization and vRNA binding (Chenavas et al., 2013; Mondal et al., 2015). Also, induced phosphorylation during cell culture could promote oligomerization of purified NP (Mondal et al., 2015). Another study has also shown the role of phosphorylation in regulating NP shuttling between the nucleus and the cytoplasm (Zheng et al., 2015). S9 and Y10 phosphorylation seem to affect the binding with importin- α and Y296 phosphorylation is assumed to affect CRM1 binding.

Interaction with vRNA is performed by a highly positively charged groove, which is at the opposite side of the oligomerization loop (figure 6). Although still in debate, current models predict NP to bind to an average of 22 to 28 nucleotides (Ortega et al., 2000; Martín-Benito et al., 2001) with an affinity of 20 nM (Baudin et al., 1994; Ng et al., 2012; Portela and Digard, 2002). This number of bound nucleotides to NP differs greatly with other negative strand RNA virus NPs described such as rhabovirus (Albertini et al., 2006), RSV (Tawar et al., 2009) and paramyxovirus (Rudolph et al., 2003; Reguera et al., 2013). Chemical labelling studies have described the binding to occur upon the phosphate backbone (Baudin et al., 1994), exposing the bases to solvent.

1.2.2.F SEGMENT 6

Segment 6 is dedicated to code for the second major glycoprotein, the **neuraminidase** (454 aa). Although jointly present with HA, NA enables for the detachment of novel virions from the host cell. Organised as tetramer (figure 7A), NA is mostly structured as a globular head domain anchored within the lipid raft by its N-terminus (residues 1- 90) (Blok et al., 1982). The head domain houses a pocket dedicated to the recognition and cleavage of sialic acid motifs (Colman et al., 1983). The cleavage of host cell sialic acid from HA is performed by NA, releasing the virion from the host cell membrane (Schmitt and Lamb, 2005; Gamblin and Skehel, 2010). Neuraminidase activity is also required in order to cleave off sialic glycans from the progeny virion to avoid self-agglutination which would diminish viral proliferation (Palese et al., 1974; Palese and Compans, 1976).

NA activity is a key druggable target. Currently, several sialic acid derivatives, and especially Oseltamivir are effective in reducing viral activity by binding to the sialic acid cavity of NA (figure 7B), thus sterically blocking sialic acid recognition and cleavage. This has the effect of leaving progeny virions bound to their host cell through the HA interaction (Palese et al., 1974; Palese and Compans, 1976). Although effective, this inhibition strategy still enables the virus to replicate its

genomic material and therefore enables the mutation process to occur, giving rise to resistant strains.

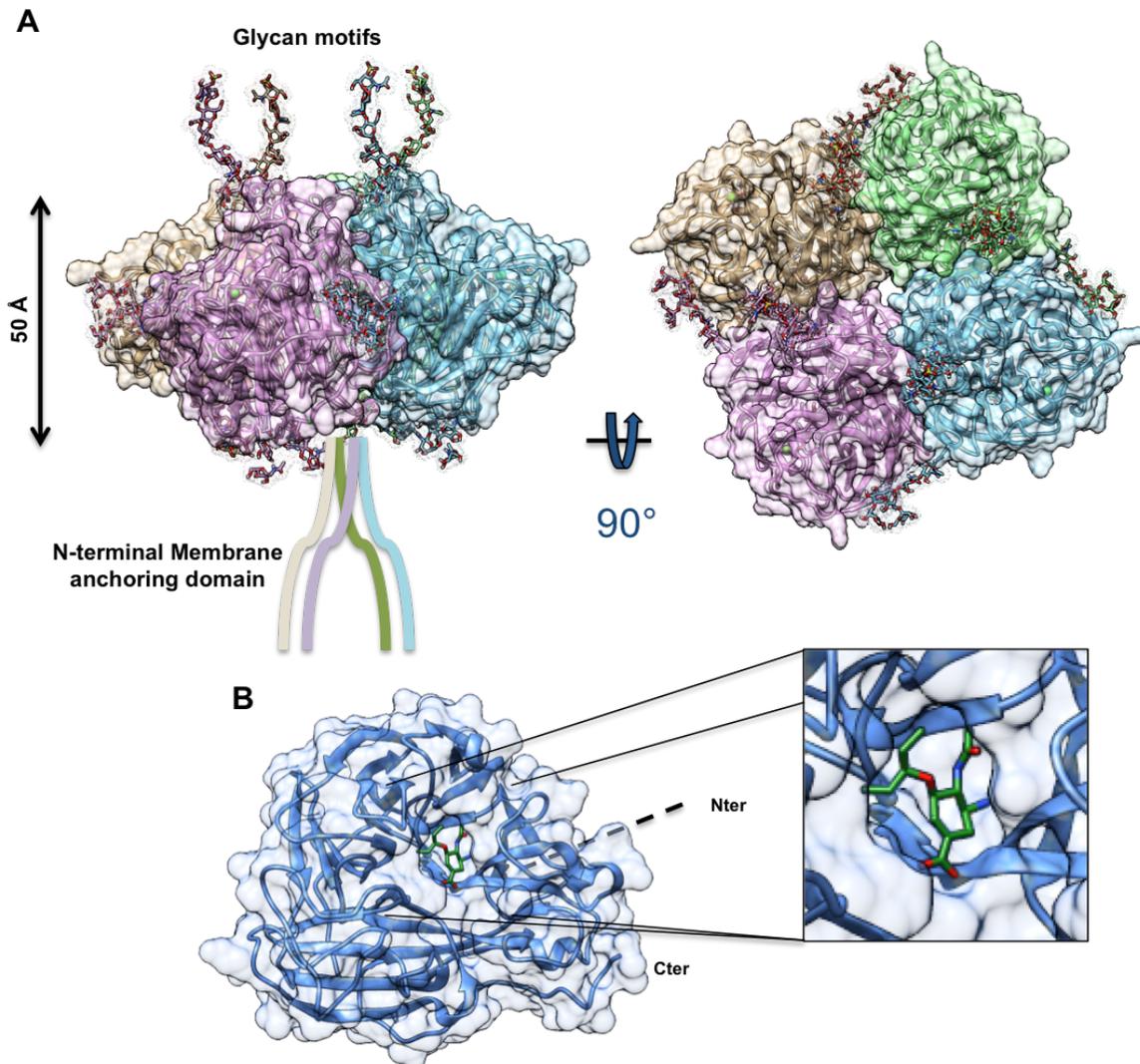


Figure 7: Neuraminidase structural features. (A) Tetrameric organisation of the neuraminidase (PDB 1NN2). (B) Monomer bound to Oseltamivir (green, PDB 4GZP). (Depictions were generated with Chimera)

Like HA, NA is also one of the major surface antigens of the virus and is subject to the same antigenic drift as HA. However, antibodies targeting NA do not neutralise the virus (Sylte and Suarez, 2009). They seem to favour viral aggregation and therefore only limit the viral spread.

1.2.2.G SEGMENT 7

Segment 7 contains the gene of **M1** (or matrix protein), of **M2** (ion channel), and **M42** which are produced through a splicing mechanism.

The **M1** matrix protein (252 aa) is the most abundant protein within the viral particles as it constitutes an interface between the outer envelope with the anchored glycoproteins and the internal 8 RNPs. EM observations have described M1 as a 6 nm long rod with one end facing the membrane and the other facing the internal part of the virion (Ruigrok et al., 1989, 2000). The protein is structurally divided into two domains which both have separate biochemical properties. The N-terminal domain (res 1-164) has been shown to bind non specifically to vRNA but not to RNPs (Watanabe et al., 1996; Baudin et al., 2001). The vRNA binding has been proposed to originate from a strongly positively charged surface (**figure 8A**) which would interact with the negatively charged phosphate backbone of the RNA (Elster et al., 1994; Watanabe et al., 1996). An alternative explanation is that this surface is dedicated to the interaction with negatively charged liposomes (Ruigrok et al., 2000). The C-terminal domain on the other hand binds to RNPs without interacting with the vRNA (Baudin et al., 2001) suggesting potential protein interactions with NP.

M1 is proposed to play important roles in the transport and assembly of RNP particles to the budding zone (Schmitt and Lamb, 2005; Nayak et al., 2009), notably through the interaction with NEP for the nuclear export of RNPs (Cros and Palese, 2003) and with cytoskeleton components, such as microfilaments (Avalos et al., 1997). Furthermore, M1 has been reported to play an active role in membrane separation in the terminal budding phase (Nayak and Hui, 2004; Nayak et al., 2004).

M2, the ion channel protein (97 aa), is obtained through the splicing of the segment 7 mRNA. This membrane anchored protein is organised as a tetramer (**Figure 8B**) on the viral surface (Holsinger and Lamb, 1991; Sugrue and Hay, 1991), enabling the passage of protons during endocytosis and thus the acidification of the virion core (Hay et al., 1985; Lamb et al., 1985). This feature is crucial to trigger the HA conformational change which in turn triggers the membrane fusion process. Acidification enables vRNP release from its interaction with M1 (Helenius, 1992).

M2 can be divided into three separate domains: an N-terminal antigenic outer domain (res 1-24), a central AM2 trans-membrane domain of 23 residues organised as a α -helix and finally an internal C-terminal domain of 50 residues (Pielak and Chou, 2011). Two key residues (H37 and W41) have been determined to control both selectivity and the unidirectional flow of protons into the viral particle (Wang et al., 1995; Tang et al., 2002). Adamantanes derivatives, such as rimantadine, maintain the W41 residue within a closed conformation (**figure 8B**), thus blocking the acidification process (Schnell and Chou, 2008). Additionally, the large C-terminal domain, also known as the cytoplasmic domain, has been shown to be important in

the viral packaging process through its interaction with the M1 protein which serves as an intermediate with the vRNPs (McCown and Pekosz, 2006; Chen et al., 2008).

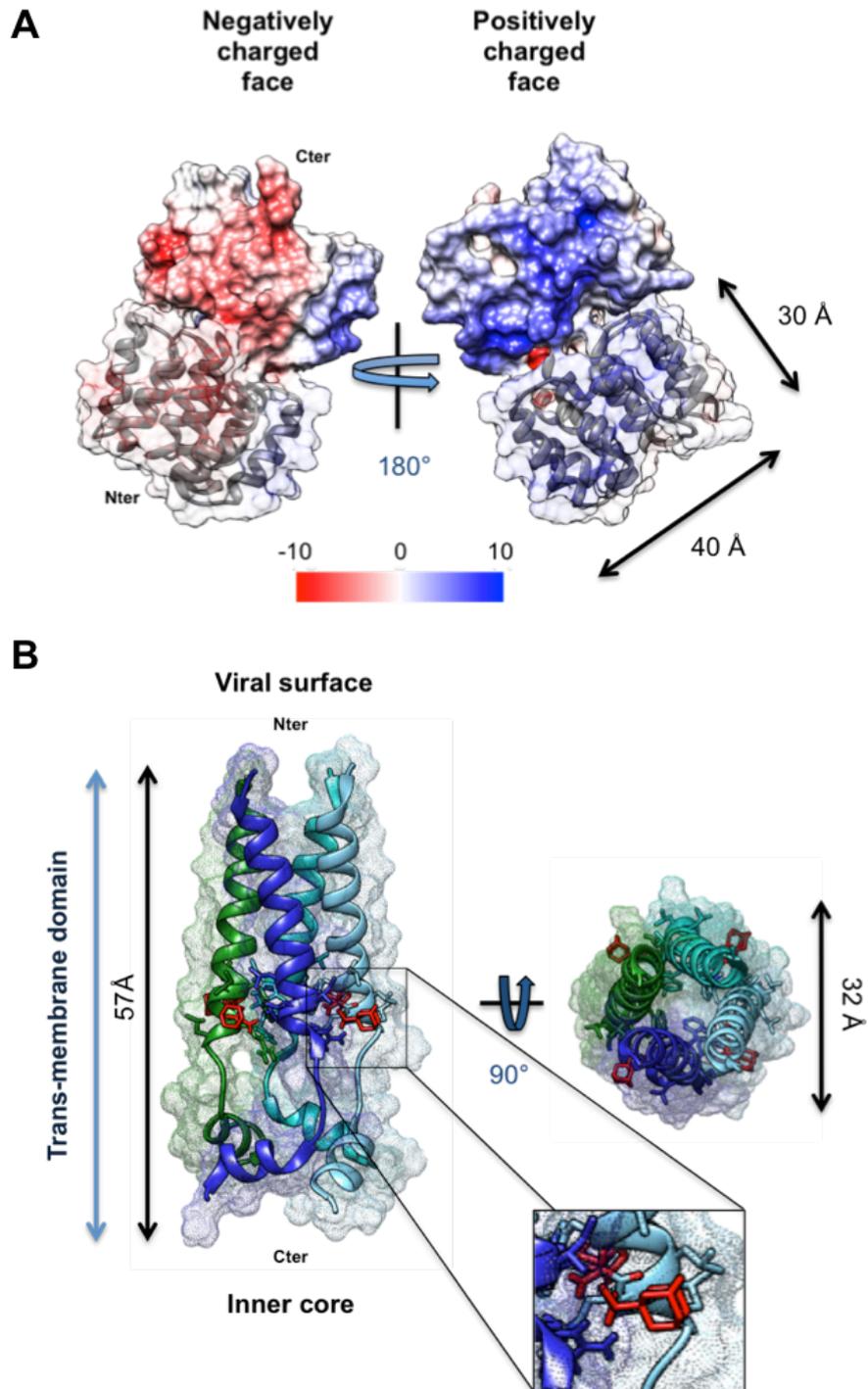


Figure 8: Structures from segment 7 proteins. (A) Crystal structure of N-terminal domain of M1 (res 1-164, PDB 1EA3) obtained at pH 7 with a P1 space group. Monomer shown in blue cartoon depiction with its neighbouring symmetry generated translation coloured by surface charge. Positive charges are shown in blue, negative charges in red. (B) NMR structure (PDB 2RLF) of the tetrameric trans-membrane domain of M2 in complex with the rimantadine inhibitor (red). (Depictions were generated with Chimera)

M42 shares a homologous trans-membrane and cytoplasmic domain with M2. The N-terminal antigenic domain is however completely different due to an alternate splicing site (Wise et al., 2012). Little is known about the function of M42 other than its potential role as an alternate ionic channel and that it is distributed on Golgi apparatus membranes instead of the cytosolic membrane for M2 (Wise et al., 2012).

1.2.2.H SEGMENT 8

Segment 8 contains the genomic information to encode **NS1** through a direct sequence read-through and **NEP** (previously called NS2) or **NS3** through a spliced mRNA read-through.

NS1 is a multifunctional protein implicated in a plethora of mechanisms that enhance the viral proliferation of influenza. The length of NS1 is strain-dependent and varies between 230 to 237 residues (Wright et al., 2013). Some reports have also detected C-terminal truncations (15 to 30 res) of NS1 within infected cells (Suarez and Perdue, 1998).

Functionally, NS1 has been characterised to modulate different distinct pathways within the infected cell in order to favour viral replication. The most prominent of these pathways is the type 1 Interferon (IFN- α /IFN- β) response suppression (Hale et al., 2010a). During infections, one of the host cell defence mechanisms is the production of cytokines (Type I IFN). The triggering mechanism is induced by the recognition of molecular sensors dedicated to the recognition of exogenous vRNA, such as the Toll-like receptors (TLRs) (Kawai and Akira, 2011) and more specifically for influenza, the RIG-I like receptors (Yoneyama et al., 2004). Once these sensors are triggered, the IFN response promotes the up-regulation of up to >300 antiviral genes (Randall and Goodbourn, 2008) such as the Mx GTPases (Haller et al., 2009). To prevent these cytokine cascades from triggering, NS1 is expressed in important quantities within the cytoplasm and outcompetes TLRs and RIG-I by binding to the double stranded vRNA (Liu et al., 1997), masking the IFN trigger mechanisms. However, this is but one of many strategies of NS1 to counter host cell defence mechanisms. NS1 also directly interferes within the cytoplasm with antiviral proteins such as 2'-5'-oligoadenylate synthetase (OAS) and serine/threonine protein kinase R (Min and Krug, 2006; Min et al., 2007). Furthermore, NS1 has been shown to bind to RIG-I activator E3 ubiquitin ligase thereby down-regulating RIG-I activity (Gack et al., 2007).

Within the nucleus, NS1 modulates the transcriptome of host cell genes by interfering with host mRNA synthesis. This interference takes place notably by inhibiting the transcription terminator CPSF (Nemeroff et al., 1998) through the inhibition of splicing and mRNA nuclear export factors (Fortes et al., 1994; Qiu et al., 1995; Satterly et al., 2007).

From a structural standpoint, although named non-structural protein 1, NS1 can be divided into 2 structured domains separated by non-structured linker regions (res 73-87, **figure 9A**) (Carrillo et al., 2014), each domain presenting its own NLS motif (Greenspan et al., 1988). The N-terminal domain (res 1-73) is dedicated to the recognition and binding of specific dsRNA motifs (Cheng et al., 2009; Liu et al., 1997). Biochemical (Nemeroff et al., 1995) and structural characterisations (Liu et al., 1997; Cheng et al., 2009) have shown that a dimerisation process is required for RNA binding (**figure 9B**).

The C-terminal domain of NS1 (res 87-230), also known as the “effector domain”, is involved in the binding of cellular host factors and stabilises the RNA binding domain (Bornholdt and Prasad, 2006; Hale et al., 2008, 2010b; Kerry et al., 2011). The interplay between both domains of NS1 is still not fully understood. Both independent domains have a tendency to dimerise through different interfaces, suggesting multiple oligomerization states (Kerry et al., 2011).

NEP (previously NS2), is a 121-residue protein translated from the spliced mRNA of segment 8. The NEP protein has been shown to integrate within the virion particles (Richardson and Akkina, 1991; Ward et al., 1995; Yasuda et al., 1993). With M1, NEP is implicated in the nuclear export of RNPs (O’Neill et al., 1998; Neumann et al., 2000; Cros and Palese, 2003). In the cytoplasm, NEP contributes to the budding process through a mediated interaction with a cellular ATPase (Gorai et al., 2012).

Structurally, NEP is predicted as mostly disordered in its N-terminal domain. Only the C-terminal domain (res 63-116) has been crystallised (Akarsu et al., 2003): it is structured as two antiparallel α -helices of 20 residues each generating an amphiphilic structure believed to interact with M1 (Akarsu et al., 2003). Recent studies have also highlighted the potential role of NEP in regulating transcription and replication of RNP complexes (Bullido et al., 2001; Mänz et al., 2012; Paterson and Fodor, 2012).

Also, an alternative mRNA splicing can generate **NS3**, which shares a common C-terminal domain with NEP. NS3 has been reported to enhance the viral

replication and could be linked to host adaptation mechanisms. It is only found within 33 strains which have all a strong propensity to jump host (Selman et al., 2012).

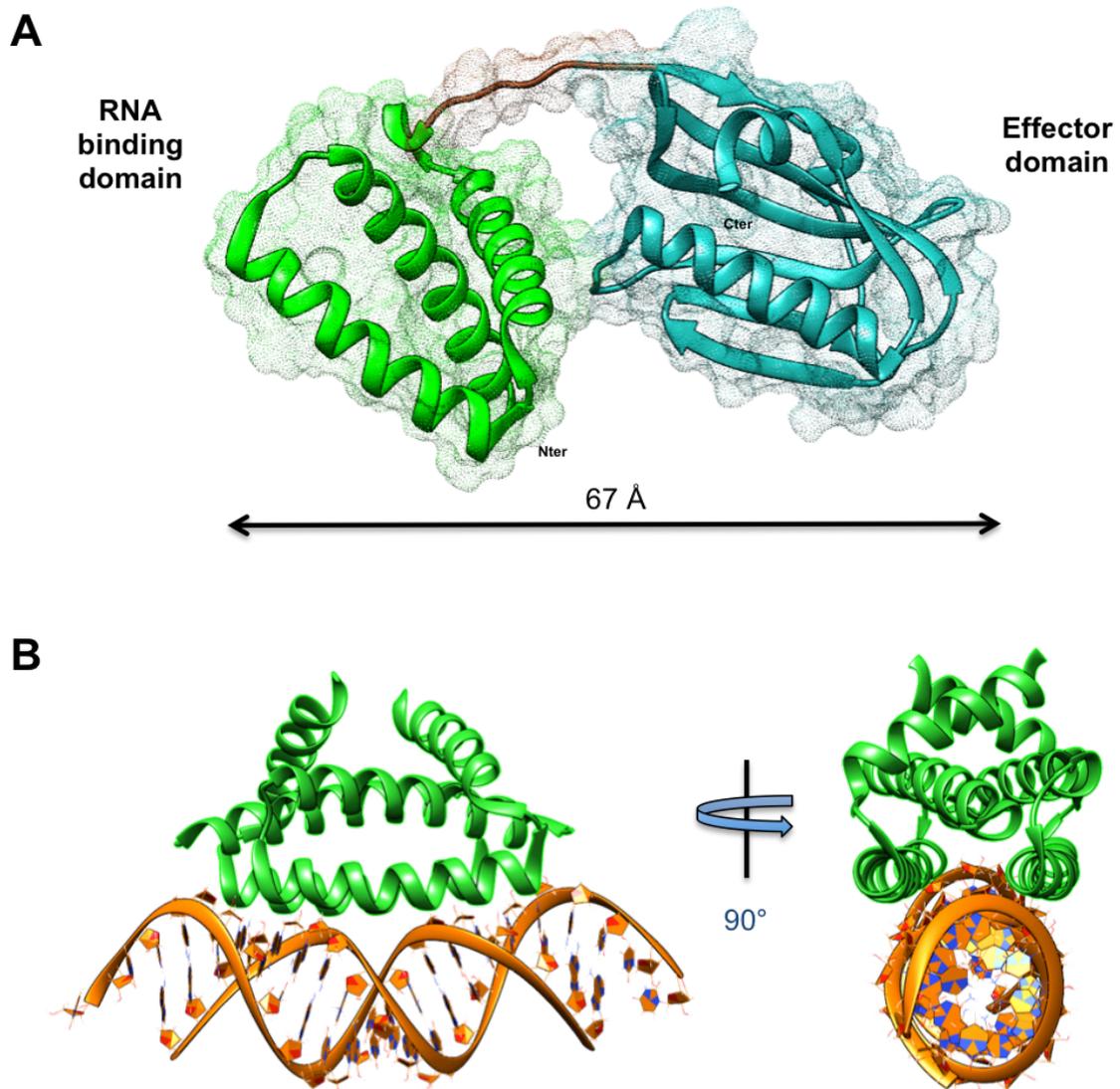


Figure 9: Structural features of NS1. (A) Full NS1 protein (PDB 4OPH) depicted in cartoon backbone and dotted surface depiction. The N-terminal RNA binding domain is coloured green and the C-terminal domain is coloured in blue with the linker region in brown. (B) N-terminal RNA binding domain (green) interacting with dsRNA (orange) through a dimerization process (PDB 2ZKO). (Depictions were generated with Chimera)

1.2.3 VIRUS LIFE CYCLE

The influenza life cycle can be divided into five major steps where all the previously described viral proteins play multifunctional roles.

1.2.3.A ENTRY WITHIN THE HOST CELL

(Stages 1 to 4, Figure 10)

HA is the main driving protein within this process, through its recognition of Saa2,6 (human influenza strains) and/or Saa2,3 (Avian influenza strains) glycan motifs (Connor et al., 1994; Viswanathan et al., 2010). Virions directly bind to the surface of host epithelial cells, which triggers the internalisation of the viral particle within an endosome. The low pH within the endosome (pH 5-6) then acts as a chemical switch on which membrane fusion is dependent. Upon acidification, the HA undergoes an important structural conformation change releasing the HA1 ectodomain and exposing the fusion peptide of the HA2 ectodomain (Harrison, 2008; Gamblin and Skehel, 2010). This fusion peptide is then inserted within the endosomal membrane and brings both the viral membrane and endosomal membrane together.

In parallel, the ion channel **M2** enables the passage of protons within the viral particle which in turn acidifies the viral core, releasing the **vRNPs** from their interaction with the **M1** matrix protein (Matlin et al., 1981). The released RNPs are then transiently (10 min) found within the cytoplasm (Martin and Helenius, 1991).

1.2.3.B NUCLEAR IMPORT OF THE vRNPs

(Stage 5, Figure 10)

One of the main characteristics of the influenza virus is that it performs its genomic replication and transcription within the nucleus of the host cell. Viral RNPs, due to their important size cannot diffuse through the nuclear membrane. This characteristic therefore imposes the hijacking of nuclear import/export mechanism through the nuclear pores (Cros and Palese, 2003; Boulo et al., 2007). This is required not only to import the incoming vRNPs, but also in later infection stages in order to assemble novel RNPs in the nucleus. Core components of the RNP are: the vRNA, which is always coated by NP and by the polymerase complex. NP, PA, PB1 and PB2 all possess one or several nuclear localisation signal (NLS) (Jones et al., 1986; Nath and Nayak, 1990; Nieto et al., 1994; Cros and Palese, 2003; Boulo et al., 2007). These NLS signals enable the recruitment of Karyopherin import/export complexes (also discussed in chapter 1.4.5). However, it has been shown that the NLS on

NP is the most important signalling element for initial RNP import within the nucleus (O'Neill et al., 1995; Cros et al., 2005; Ozawa et al., 2007). The NLS of NP recruits a karyopherin- α class protein which then serves as an adaptor to recruit the karyopherin- β class importin. Karyopherin- β then initiates passage through the nuclear pore complex (NPC) and releases the RNPs upon binding to RanGTP (Boulo et al., 2007; Cros et al., 2005). Ran is a RAS family GTPase that cycles from the nucleus to the cytoplasm through its binding with a β -importin or β -exportin complexes when it is bound to GTP, which is mostly found in the nucleus. Once in the cytoplasm, Ran recruits a GTPase activating factor GAP which provokes Ran to hydrolyse the GTP into GDP. This triggers the disengagement of Ran from the β -importin or exportin and enables Ran to diffuse back to the nucleus in order to re-engage a GTP molecule (Görlich and Kutay, 1999).

1.2.3.C TRANSCRIPTION AND REPLICATION OF THE RNPs (Stages 6 to 10, Figure 10)

Once in the nucleus, the RdRp performs two distinct activities (within the RNP context) which are the **transcription** and **replication** of the vRNA genome (detailed in chapter 1.4.4).

Transcription, as further described in chapter 1.4.4.A, relies on the very unique mechanism of “cap-snatching” to transcribe the vRNA template into mRNA which will then be exported into the cytoplasm and translated into viral proteins (Krug et al., 1979; Plotch et al., 1981). Some of these viral proteins are imported into the nucleus after translation; such is the case for polymerase subunits (PA, PB1, PB2, see chapter 1.4.5), NP, NS1 and M1 through the importin α/β pathway (Boulo et al., 2007; Hutchinson and Fodor, 2012). NEP, on the other hand is small enough (14,5 kDa) to diffuse directly into the nucleus. Polymerase subunits and NP will be used in the replication of vRNA to generate novel RNPs (Moeller et al., 2012; Ye et al., 2012). NS1 will interfere with the host cell transcription machinery and also play a role in the regulation of vRNP replication (Marión et al., 1997; Robb et al., 2009). NEP will participate —together with M1— in the export of RNPs from the nucleus (Neumann et al., 2000; O'Neill et al., 1998).

Replication is a primer independent process which leads to the accumulation of novel (-) vRNA (Hay et al., 1982). To do so, two main stages have been identified as required (detailed in chapter 1.4.4). First the transcription of the (-) vRNA into a positive strand template (+) cRNA which itself becomes a template to generate novel (-) vRNA. Through this process, cRNA is also coated by newly synthesised NP and the RdRp within a transient (+) RNP complex (cRNP).

Although the switch between a transcribing and replicating polymerase is not yet fully understood, presence of free NP and polymerase for a successful replication indicate that a first phase transcription process is required (Huang et al., 1990). A self-amplifying loop between replication and transcription will then occur: as more template vRNA is produced, more mRNA can be transcribed.

1.2.3.D EXPORT OF THE NEWLY SYNTHETISED vRNPs TOWARDS THE BUDDING ZONE

(Stage 11, Figure 10)

Once formed, novel RNPs will eventually be transported out of the nucleus (Shapiro et al., 1987). At first, NP was believed to be responsible of Crm1 binding through its two nuclear export signal (NES) domains. However, although Crm1 does bind to NP (Elton et al., 2001), no GTPase activity from Ran has been detected upon interaction with NP (Akarsu et al., 2003), which is required to trigger the migration of Crm1 towards the cytoplasm (Boulo et al., 2007). NEP on the other hand does have a NES domain and does activate Ran GTPase, suggesting that NEP links M1/RNP to Crm1 for nuclear export. This model points out the key role played by M1 and NEP in regulating this phenomenon. Interestingly, post translation modifications may be important as showed by the hyper-phosphorylation of M1 which sequesters M1 within the nucleus (Whittaker et al., 1995).

After being translocated through the nuclear pores, Ran GTP recruits RanGAP (Ran GTPase Activating Protein), which then hydrolyses the bound GTP into GDP. This provokes a conformational change within Ran which releases the GDP and itself from Crm1, triggering the release of the entire exportin bound cargo (Boulo et al., 2007).

RNPs are then migrated towards the “budd zone” through the interaction with Rab11, a GTPase which associates with endoplasmic reticulum (ER) (Bruce et al., 2010; Einfeld et al., 2011). The interaction with Rab11 seems to be mediated by the RdRp, possibly through the PB2 subunit. Rab11 in turn interacts with the ER, microtubules and a series of host partners in order to migrate the progeny RNPs along the ER and towards the apical plasma membrane (Amorim et al., 2011; Avilov et al., 2012). Interestingly, the migration along the ER also co-localises RNPs together, preparing them for final packaging within the virion (Takizawa et al., 2010; Chou et al., 2013).

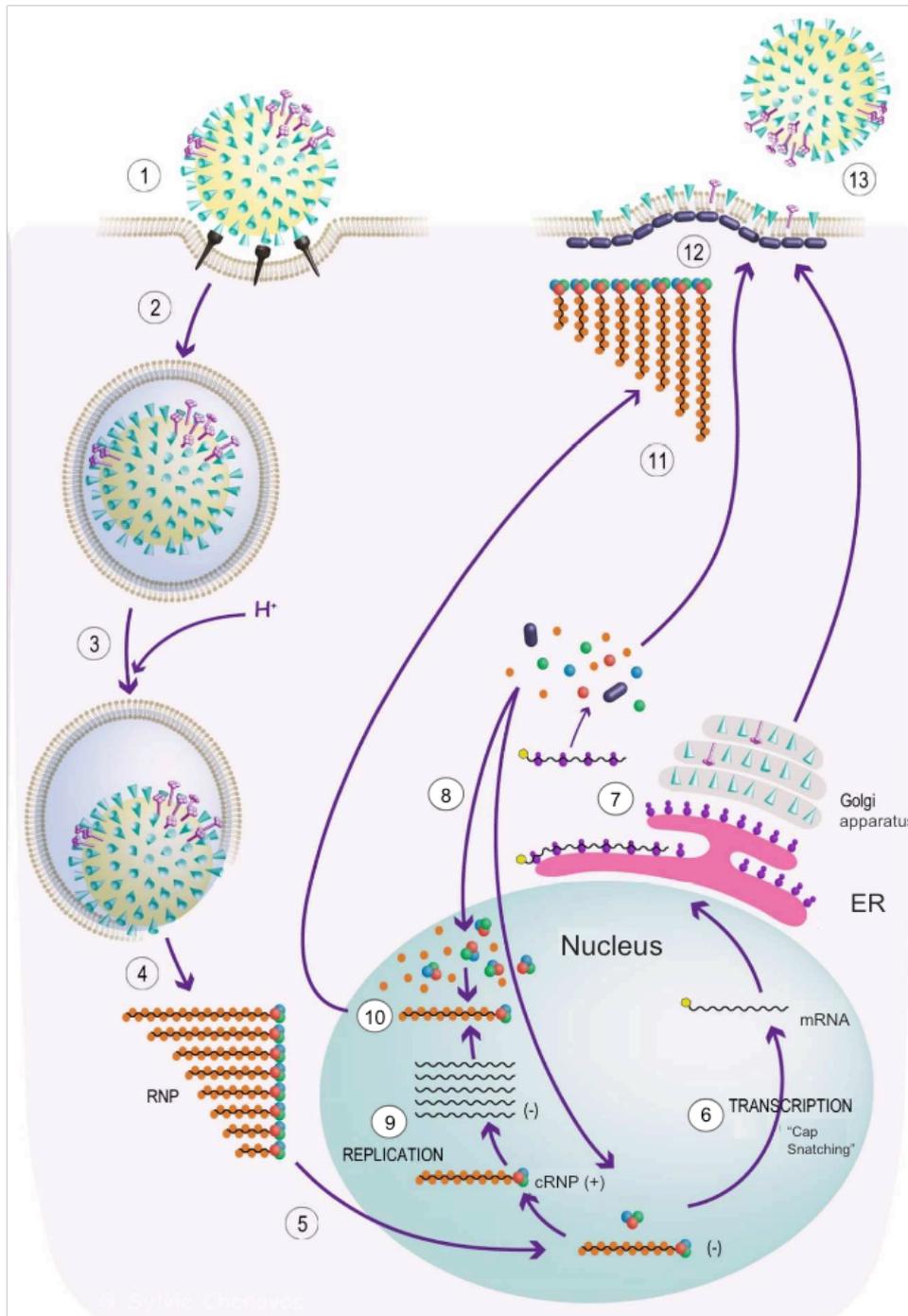


Figure 10: Influenza A virus life cycle. (1) HA mediated viral recognition of sialic moieties on the host cell surface followed by internalisation within the endosomes (2). (3) Acidification of the endosome triggering the viral and endosomal membrane fusion, which releases the RNPs within the cytosol (4). Importin mediated nuclear import of the RNPs (5). Cap-dependent transcription (6) is performed on vRNA to generate, through "cap snatching" chimeric mRNA coding for the different viral proteins which are then expressed in the cytoplasm (7). Core RNP viral proteins (NP, PA, PB1, PB2) are then imported back in the nucleus (8). (9) Replication of the vRNA using a cRNP intermediate to generate novel vRNA strands which are then assembled within new RNPs (10). Upon assembly, novel vRNPs are exported from the nucleus (11) to the viral budding zone (12) for final assembly with the other viral proteins. (Adapted from Guilligay et al., 2010)

1.2.3.E ASSEMBLY AND BUDDING

(Stages 12 to 13, Figure 10)

As RNPs are formed and make their way to the plasma membrane, HA, NA, M2 are also expressed and addressed to the apical side of the plasma membrane where they concentrate in lipid raft microdomains (Scheiffele et al., 1999; Nayak and Hui, 2004). M1 interacts with all the cytoplasmic tails of NA, HA and M2 and anchors them on the nucleocapsid (Rossman and Lamb, 2011).

RNP segment packaging is still currently debated. To date two models exist: a random packaging model *versus* a selective packaging model.

The first model is sustained by the observation that some influenza viruses can carry more than the eight required segments (Enami et al., 1991; Bancroft and Parslow, 2002; Gao et al., 2010). This model would favour a very important number of potential RNP reassortments and only the viruses with the correct combinations would be viable.

However, recent studies tend to favour the latter model of selective incorporation. Electron tomography experiments have shown a consistent ordering by size of the RNP segments within budding virions (Noda et al., 2006; Fournier et al., 2012; Noda et al., 2012), with only a limited number of conformations observed.

The signal enabling discrimination RNPs is believed to be present within the 5' and 3' 100-300 nt region which overlaps on both UTR and ORF regions of the viral RNA (Luytjes et al., 1989; Muramoto et al., 2006; Hutchinson et al., 2010). In order to discriminate among RNPs, segment specific sequences in 5' and 3' would be responsible for direct linkage between vRNA segments through Watson-Creek base pairing (Fujii et al., 2003; Goto et al., 2013). Also, viral assays have shown that mutations within these segment specific regions could modify the packaging of other segments (Muramoto et al., 2006; Hutchinson et al., 2008; Marsh et al., 2008; Hutchinson et al., 2009). This model therefore implies that vRNP structures enable direct access to some of the vRNA nucleotides and that evolution has pushed the influenza virus to integrate only a limited number of vRNP conformations. This question has great importance as it fundamentally defines some of the possible re-arrangements of influenza virions.

1.3 EVOLUTION OF THE INFLUENZA VIRUS

As mentioned earlier, the influenza virus can infect a wide variety of hosts, escape multiple layers of host defence mechanisms and use the host cell to its own advantage in very unique ways such as “cap snatching”. In modern history, it has also been able to evade vaccination whereas other viruses have *quasi* disappeared in humans such as smallpox. All these characteristics are explained by the astonishingly efficient adaptation mechanisms within the influenza and its ability to jump between species. Two main properties drive this adaptation, antigenic “drift” and antigenic “shift”.

1.3.1 ANTIGENIC DRIFT

Antigenic drift is a slow but constant process of adaptation of the influenza virus to counter the acquired host immunity. It is driven by point mutations within the genome of influenza and is the direct result of vRNA replication by the RdRp. It is estimated that the RdRp has an error rate of around 1 nucleotide/ 10 000 nucleotide insertions during replication as it does not possess any proofreading activity (Drake, 1993; Boivin et al., 2010). Therefore it is estimated that every progeny virion statistically contains one mutation. Every cell generates an estimate of 1000 to 10000 virions during the infection cycle (Stray and Air, 2001). During this process, viable virions with mutated HA and NA which are no longer recognised by neutralising antibodies escape from the host immunity and continue the infectious cycle (Bush et al., 1999; Plotkin and Dushoff, 2003).

The antigenic drift phenomenon is crucial to the virus for maintaining epidemic strains of influenza in circulation, notably to escape vaccine formulation. It is also worth noting that this mutagenic property is the main driver behind the emergence of antiviral resistant strains.

1.3.2 ANTIGENIC SHIFT

Antigenic shift is a sporadic and important change within the virion genome and is the direct consequence of its segmented nature. Indeed, every vRNA within the context of a vRNP behaves as an independent sub-genome, comparable to a chromosome within humans. This feature enables segmented viruses to reassort during a combined infection from two or more strains of viruses. Reassortment has

been observed within all influenza genera (A, B and C) but never between them. From an antigenic point of view, reassortments can generate novel HA and NA combinations that greatly differ from the circulating strains and therefore can potentially enhance the infectivity of the virus. As such, reassortments are the main drivers behind the emergence of novel pandemic influenza strains but also for interspecies adaptation (Wright et al, 2013; Schrauwen et al., 2014).

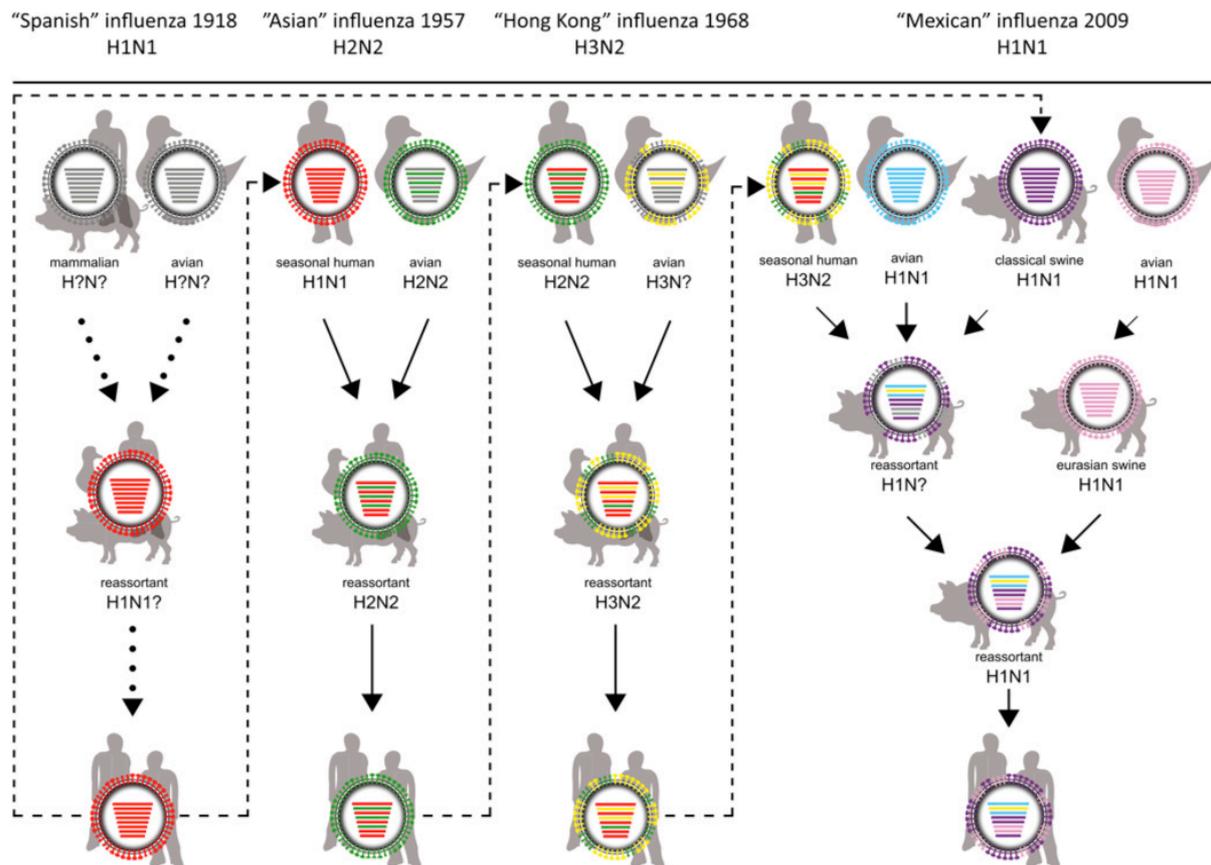


Figure 11: Reassortment events driving the emergence of pandemic influenza A strains. (Adapted from Schrauwen et al., 2014)

The importance of reassortment in modern pandemic emergence has been demonstrated by the last four major influenza pandemics (Figure 11). For instance the 1957 and 1968 pandemics were the result of human strains reassortments with avian segments coding for HA/PB1/NA or HA/PB1 respectively (Webster et al., 1992; Schrauwen et al., 2014; Yoon et al., 2014). The 2009 H1N1 pandemic was on the other hand generated by multiple reassortment stages between humans, birds and pigs. In the end, the H1N1 virus possessed (N1)/M segments from a Eurasian pig influenza strain, (H1)/NP/NS segments originating from another H1N1 swine influenza strain, PB2/PA segments from a North American avian strain (H1N1) and the PB1 segment from the human epidemic strain H3N2 (Munier et al., 2010). These

influenza reassortments that affect humans are but a fraction of the total amount of reassortment events occurring in nature. In an ever changing world where climate change will modify the habitat, crowding and interaction between all of these species, novel unpredictable reassortments are bound to emerge, once again pushing the need to better understand the mechanisms behind such reassortments.

1.4 THE INFLUENZA POLYMERASE STRUCTURE, FUNCTION AND ASSEMBLY PATHWAY

As previously described, three of the longest segments of the vRNA genome contain the coding sequence for the RdRp complex. Composed of **PA** (“Polymerase Acid”, 716 aa), **PB1** (“Polymerase Basic 1”, 757 aa) and **PB2** (“Polymerase Basic 2”, 759 aa), the heterotrimeric assembly of the polymerase is the largest protein complex of the viral particle (250 kDa). Within the context of a vRNP, the polymerase binds to the promoter regions of the vRNA (5' and 3' termini) and together with NP is the molecular machinery that drives **Replication** and **Transcription** of the vRNA genome.

1.4.1 FOUNDING WORK ON THE POLYMERASE SUBUNITS

For many years, the study of the polymerase complex was limited to a biochemical approach due to inherent difficulties in the structural characterisation of individual subunits of the polymerase complex.

Little was known of PA's function for a long time other than it was linked to the replication process, as this was shown through thermosensitive mutants (Krug et al., 1975; Mahy, 1983; Honda et al., 2002). Later on, some research showed a specific link with replication and transcription (Fodor et al., 2002; Regan et al., 2006), most notably through the inhibition of the endonuclease activity (Fodor et al., 2002). Other studies have also shown PA to be associated with a virus-induced proteolysis activity (Sanz-Ezquerro et al., 1996; Hara et al., 2001; Perales et al., 2000). This subunit was described to be the most soluble when expressed independently, hinting a stabilising role towards PB1 which cannot be expressed alone (Hara et al., 2006).

PB1 was identified as the central subunit at the core of the heterotrimeric complex (Digard et al., 1989) housing the polymerase activity. Early bio-informatics studies predicted the central region of PB1 to contain the conserved A-D motifs of RdRps (Poch et al., 1989) but also the pre-A and E motifs, which are exclusive to

segmented negative strand RNA viruses (Müller et al., 1994). These predictions corroborated earlier work using thermosensitive mutants which had already shown the pivotal role played by PB1 in the transcription/replication process (Mahy, 1983) and were in turn confirmed by point mutations (Biswas and Nayak, 1994). Differently from PA, no single domain of PB1 was ever expressed in a soluble form hinting its conformational dependency on other proteins, notably PA with which it interacts through a specific interaction domain (Pérez and Donis, 1995; González et al., 1996; Toyoda et al., 1996; Zürcher et al., 1996) and was shown to form a dimeric complex (Lee et al., 2002; Deng et al., 2005).

PB2 was shown early on to interact with capped RNAs (Blaas et al., 1982; Ulmanen et al., 1981) and to play a pivotal role in the initiation of transcription (Plotch et al., 1981). Through a thermosensitive mutant analysis it was confirmed to be important for transcription (Mahy, 1983). Later point mutational studies confirmed the importance of PB2 in the cap binding process (Fechter et al., 2003) but also in the replication process (Gastaminza et al., 2003; Jorba et al., 2009).

1.4.2 LOW RESOLUTION EM MAPS AND INDIVIDUAL DOMAIN STRUCTURES OF THE POLYMERASE

It was not before the beginning of the 2000s decade that structural information on the polymerase was obtained. Before that, classical structural biology techniques such as EM, protein crystallography and NMR were all subject to their own limitations. EM was just starting to be used in order to produce large assembly reconstructions and although RNPs could be purified, their high degree of flexibility made the process of class averaging difficult. X-ray crystallography and solution NMR required important quantities of highly pure soluble protein and before 2007 no soluble crystallising domains had been isolated.

The first structural breakthrough came with the design of artificially small and circular mini RNPs (Ortega et al., 2000). These mini RNPs remain replicase and transcriptase active. Using a 248 nucleotide vRNA coated by 9 NPs and one RdRp would ultimately generate a minimal mini vRNP rigid and homogenous enough for EM negative staining studies (Martín-Benito et al., 2001), producing the first ever low resolution model of a polymerase at 36 Å resolution within an RNP complex at 27 Å resolution (Figure 11A). This model was later improved in resolution to 24 Å and further characterised through the localisation of the polymerase subunits using subunit-specific monoclonal antibodies (Area et al., 2004). With the advent of modern cryo-EM imaging, the mini RNP was eventually modelled at higher resolution (Figure 12B),

giving unique structural insight into the NP oligomerization process (Coloma et al., 2009). NP protomers could be visualised with a 12 Å resolution enabling a precise fitting of the atomic structure of NP. The polymerase was characterised at 18 Å resolution. More recently, two groups were independently able to produce full RNP particle cryo-EM reconstructions using images from linearised RNPs, but also by separating the class averaging over the segments and the RNP (Arranz et al., 2012; Moeller et al., 2012). This produced 20 Å models of both native RNPs (Figure 12c) (Arranz et al., 2012) and recombinantly produced RNPs (Moeller et al., 2012) and gave rise to distinct interpretations of NP conformation with the supra-helical structure.

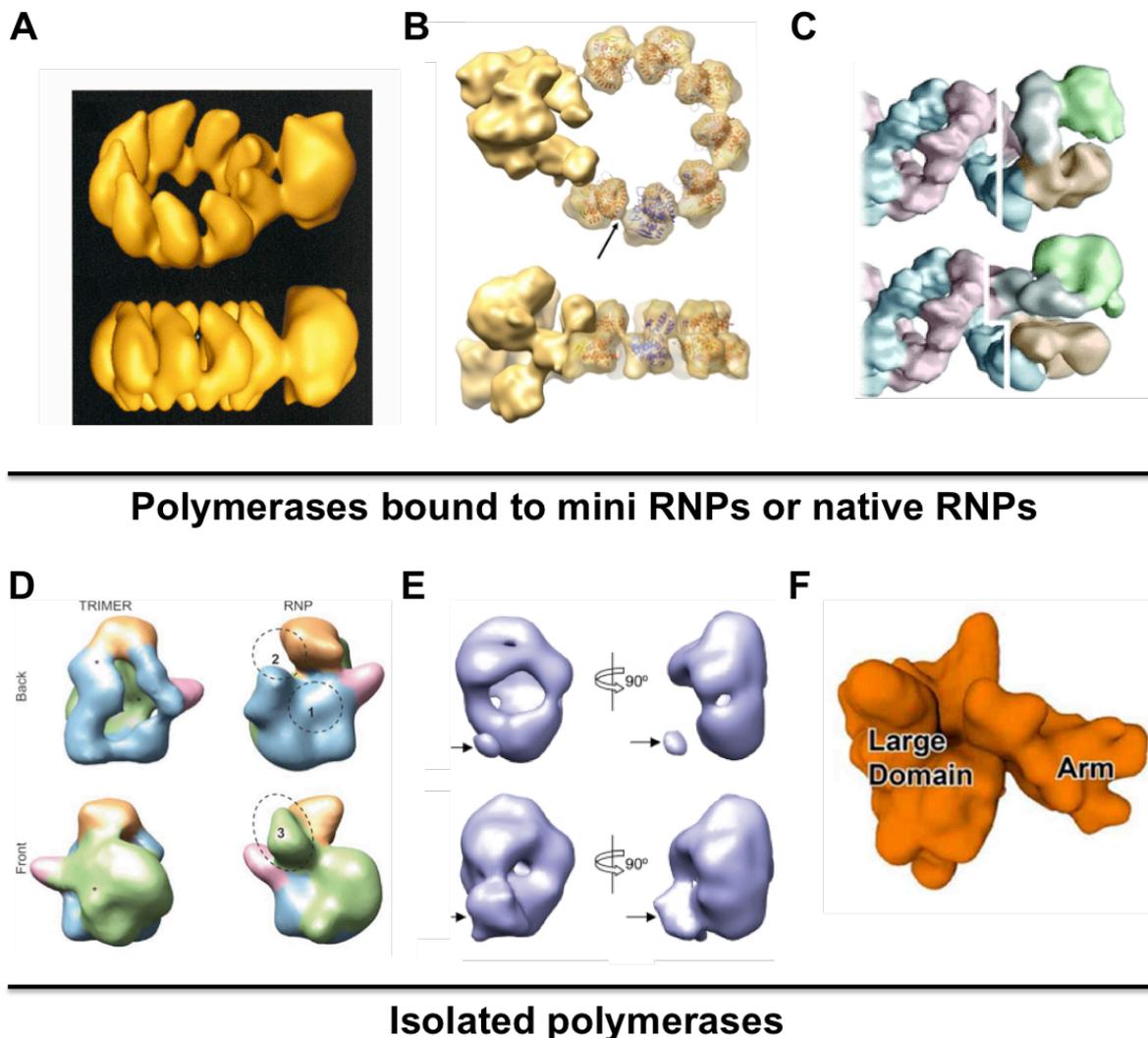


Figure 12: EM and cryo-EM models of RdRps isolated or within the context a vRNP. (A) Very first negative stain EM reconstruction of the mini-RNP particle at 27-36 Å resolution (Martin-Benito et al., 2001). (B) Cryo-EM reconstruction at 12-18 Å resolution with fitted NP atomic model (Coloma et al., 2009). (C) Native RNP polymerase end reconstruction at 20 Å resolution with three separated domains in brown grey and green (Arranz et al., 2012). (D) Negative EM models of unbound polymerase at 25 Å resolution with coloured domains (Torreira et al., 2007). (E) vRNA bound polymerase negative stain EM model at 24 Å resolution (Resa-Infante et al., 2010). (F) Cryo-EM reconstruction at 13 Å resolution of the unbound polymerase (Moeller et al., 2012).

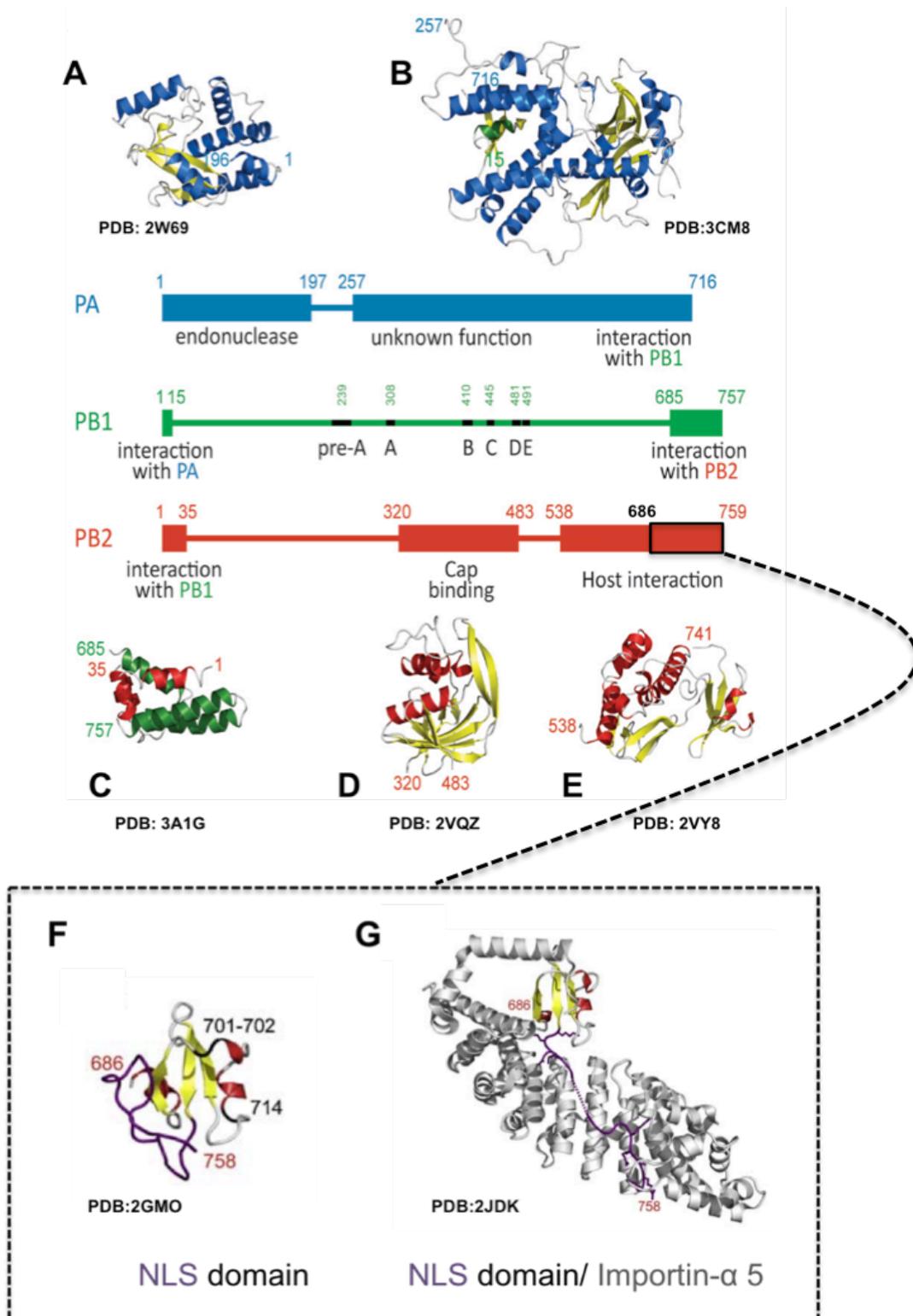


Figure 13: Domain structures of the influenza A RdRp. (A) N-terminal endonuclease domain of PA (Dias et al., 2009). (B) C-terminal domain of PA (He et al., 2008). (C) PB1c-PB2n interacting domain (Sugitama et al, 2009). (D) Cap binding domain of PB2 (Guilligay et al., 2008). (E) C-terminal host interaction and NLS dual domain of PB2 (Tarendeau et al., 2008). NLS domain of PB2 in a free state (F) or bound (G) to importin- α 5 (Tarendeau et al., 2007). (Adapted from Ruigrok et al., 2010)

In parallel to characterising the polymerase bound to vRNA within RNPs, recombinant expression of the polymerase was undertaken within mammalian cells and produced only enough material for negative staining reconstructions. Several models were generated at around 25-26 Å of resolution for unbound (Torreira et al., 2007) or bound (Resa-Infante et al., 2010) polymerase to a vRNA template (Figure 12D and 12E). More recently, a refined cryo-EM RNA free structure of the full recombinantly produced polymerase was also produced at 13 Å resolution (Moeller et al., 2012) (Figure 12F). The comparison of these structures between them and with the vRNP bound forms showed similarities in overall architecture but also many differences, revealing potentially important conformational changes upon RNA binding and an inherent mobility of some of the domains (Torreira et al., 2007; Martín-Benito and Ortín, 2013).

In conjunction with the EM successes in describing the overall architecture of the polymerase trimer, the very first atomic details of the polymerase were obtained on isolated soluble domains of PB2 by the ESPRIT screening methodology (Expression of Soluble Proteins by Random Incremental Truncation) (Tarendeau et al., 2007). An automated truncated fragment library search identified two separate domains on the PB2 protein, which are now describe.

First, the C-terminal NLS domain (res 678-759), containing a bipartite NLS peptide (736-RKRX₁₂KRIR-755), enables the recruitment of importin- α , itself an adapter to importin- β (Tarendeau et al., 2007). The structure of the NLS domain was determined both in solution using NMR (Figure 13F) and by X-ray crystallography in complex with importin- α 5 (Figure 13G). In these structures, the NLS signalling peptide adopts a completely different conformation. It is observed folded against the NLS domain when unbound to importin- α (NMR structure) and extended onto the importin- α binding within the complex (Tarendeau et al., 2007). In later studies, this domain was also crystallised within a N-terminally extended construct (res 538-759, Figure 13E) which included a new, individually folded domain that was also crystallised alone. Named the host interaction domain, it is highly implicated in host pathway interactions (Tarendeau et al., 2008; Kuzuhara et al., 2009). This host adaptation domain (res 538-678) has an exposed Glu 627 on its surface whose mutation into an lysine is crucial for the influenza virus adaptation within mammal hosts (Shinya et al., 2007; Naffakh et al., 2008). Interestingly, the double host adaption/NLS domain was observed folded in a manner which was incompatible with importin- α binding, suggesting that important conformational changes can take place between those two domains (Tarendeau et al., 2008; Kuzuhara et al., 2009).

The second major domain was a middle domain (res 320-483, **Figure 13D**) which is folded into a cap binding domain (Guilligay et al., 2008). This structure was solved using x-ray crystallography and was observed bound to m7GTP revealing the key residues implicated in cap recognition. Notably his357/phe404, which sandwich the methylated base, and glu361/lys376 that specifically interacts with the guanine nucleobase. Those same residues are also highly conserved among influenza viruses. Mutations on some of these residues within PB2 would then lead to a loss of transcription activity but not to a loss in replication within a mini-vRNP assay (Guilligay et al., 2008), ultimately showing the crucial role played by the cap binding domain in “cap snatching”. For these reasons, the cap-binding domain has been the focus of extensive antiviral compound screening by different research groups.

Differently from PB2, or PB1, PA is a highly soluble protein when expressed on its own but cannot be crystallised as a full protein. It was however shown that a trypsin digestion of PA generates two major degradations, a N-terminal 25 kDa domain and a C-terminal 55 kDa domain (Hara et al., 2006). Construct optimisation led both domains to be eventually crystallised by several groups.

The first domain to be characterised was the C-terminal domain (res 257-716) which was co-crystallised with a short part of the PB1 N-terminus (res 1-15, **Figure 13B**) by two independent groups at the same time (He et al., 2008; Obayashi et al., 2008). Both groups however, although suspecting a potential positively charged binding groove for the viral RNA, could not determine any biochemical activity from this domain. However, the presence of the PB1 binding domain opened the gateway to antiviral compounds targeting the PA-PB1 interaction (Li et al., 2012; Muratore et al., 2012).

In the following year, once again two separate groups solved the structure of the N-terminal domain (res 1-197, **Figure 13A**) of PA. It was shown to house the endonuclease domain of the polymerase and presented a fold comparable to class II endonucleases (Dias et al., 2009; Yuan et al., 2009). Both groups however presented a divergence in the metal ions present in the catalytic site. Yuan et al. crystallised the domain with one Mg²⁺ ion within the catalytic site and conducted point mutations on the residues implicated in its recognition, demonstrating that these conserved residues could affect replication in a recombinant RNP system. Dias et al. crystallised the domain with two Mn²⁺ ions within the catalytic core, with one of the Mn²⁺ ions placed at the same atomic position as in the structure from Yuan et al. This metallic configuration is not as common; however, Dias et al. were able to show *in vitro* the direct efficiency increase of Mn²⁺ for RNA and ssDNA cleavage when compared to Mg²⁺ of Co²⁺. Later on, a study undertaken by our group

(which is detailed within the **supplementary publication n°3**) showed that Mn^{2+} was bound to the endonuclease domain with a 500 fold superior affinity than Mg^{2+} and that point mutations within the key residues implicated in the recognition of manganese severely disrupted the endonuclease activity (Crépin et al., 2010). As for the cap binding domain, the endonuclease domain has recently been the focus of intense compound screening in the hope to design novel inhibitors of the influenza virus (Kowalinski et al., 2012; DuBois et al., 2012).

As I started my work on the polymerase, 90% of the PB1 protein was still structurally un-characterised. The only described structures were those of the interacting domains, res 1-15 for the interface with PA (He et al., 2008; Obayashi et al., 2008) and res 685-757 which forms the interface with PB2 (Sugiyama et al., 2009). However, towards the end of my PhD, a series of papers would finally unravel the overall structure of the RdRp.

1.4.3 FULL ATOMIC MODELS FOR THE RNA DEPENDANT RNA POLYMERASE

As you will later read in this thesis, our group has focused since a number of years to obtain the structure of the influenza A polymerase. In 2010, our group in collaboration with Imre Berger attempted a novel expression method to express the complex. This insect cell expression method (extensively discussed in the **supplementary publication N°1**) relies on the principle that all the subunits of a given complex are jointly expressed as one single polyprotein with TEV (Tobacco Etch Virus protease) cleavage sites in between every subunit. During expression, TEV, which is the first protein to be expressed, processes the polyprotein to generate a stoichiometric assembly of the complex. Although very promising compared to other recombinant expression methods, this method does not yet enable the full expression of an influenza A H3N2 (Human) or H5N1 (Avian) polymerase. The elements generating this problem and the optimization of these constructs will be discussed within the results chapter of this Thesis.

Our collaborators S.Cusack et al. have applied this same methodology successfully on distinct influenza strains and were able to express and crystallise a full influenza B polymerase (Reich et al., 2014) and influenza A H17N10 from the newly discovered bat strain (Pflug et al., 2014). Although sharing approximately 36% of sequence identity for PA and PB2 and 60% for PB1, both polymerases are structurally homologous in the overall architecture. The full polymerase adopts a U shape configuration with two protruding arms (**Figure 14A**), corresponding to the

endonuclease domain of PA and the cap-binding domain of PB2 that face each other and are at the entrance of an important tunnel.

Previous domain structures of the PA_{cter}-PB1_{nter} and PB1_{cter}-PB2_{nter} are found as expected. However, the complexity with which all subunits fold together is much more important than previously thought. PA when visualised alone (Figure 14B) is folded into two completely separate domains, linked by a very long linker called PA-hinge, which is completely wrapped around PB1 (Figure 14A). One small domain in the N-terminus corresponding to the endonuclease domain and a bigger C-terminal domain which is packed against PB1. PB1 is at the core of the complex and organised as one single domain (Figure 14B) interacting at a majority with PA on one face and with PB2 to a lesser extent on the other (Figure 14A). Interestingly, the region proposed to regulate the nuclear import of PB1, containing a bipartite NLS motif is folded into a very long and exposed β -ribbon (Figure 14A).

PB2, like PA, is spread out over the entire polymerase (Figure 14B) into what can be simplified as three main regions. The first N-terminal region (res 1-247) consists in a series of α -helices and a few β -strands which tightly intertwine with the C-terminal domain of PB1 (Figure 14A) and support the N-terminal endonuclease of PA, making this region unique, as all three subunits interact with each other. The second region is the globular cap-binding domain (res 320-480) which itself sits on the last C-terminal region of PB2 composed of several sub-domains (res 480-751).

Importantly, both influenza B and influenza A polymerases were crystallised bound to a specific combination of 3' and 5' vRNA promoter sequences (Figure 14C). Both strands are bound in a very unique conformation which answers the longstanding dilemma of whether the promoter ends form a "panhandle" structure (Hsu et al., 1987; Fodor et al., 1994) or a "corkscrew" (Flick et al., 1996; Tomescu et al., 2014) structure. In reality the two vRNAs form a mix of both models with both strands exiting the polymerase as a duplex (3' 10-UCUC-13 with 5' 11-AGAG-14), however when entering the polymerase both 3' and 5' vRNA are separated and bind in different ways. The 5' vRNA strand forms a hook structure as some models had already predicted (Pritlove et al., 1999) and is bound at the PA-PB1 interface. The 3' vRNA is packed as a single stranded arc and most interestingly binds in a specific RNA binding pocket which is composed of all three subunits. Importantly, the 3' vRNA appears to be directed out of the polymerase active site; this characteristic, also observed in influenza B, puts into question the initiation of polymerization on the 3' strand.

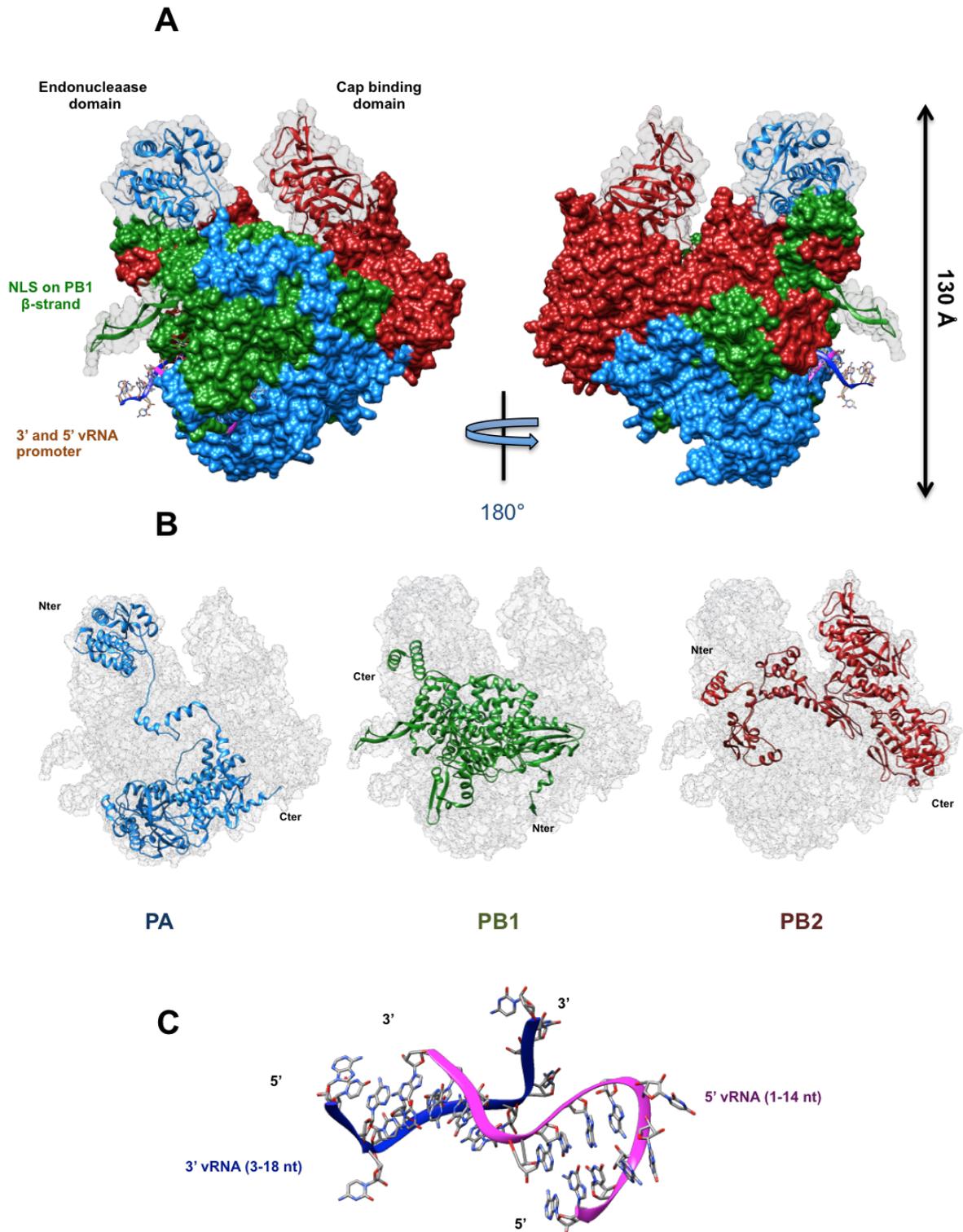


Figure 14: influenza A H17N10 overall structure. (A) Surface representation of the polymerase with PA in blue, PB1 in green and PB2 in red. The endonuclease of PA, the cap binding domain of PB2 and the NLS containing β -ribbon on PB1 are drawn in cartoon style. vRNA promoters are depicted in dark blue (3'vRNA) and magenta (5'vRNA) ribbons. (B) Cartoon depiction of the separate subunits with the surface envelope in the background. (C) Ribbon and stick diagram of the 3' and 5' vRNA promoter structure within the polymerase. (PDB: 4WSA). (Depictions were generated with Chimera)

As previously mentioned, PB1 is found as a single well-folded domain with several extruding features (Figure 15A). These features include an N-terminal PA binding region (res 1-15), the C-terminal PB2 binding region (res 669-750) and the NLS containing β -ribbon (res 177-212) motif, but also the β -hairpin motif (res 352-360) that intricately folds within the PA C-terminus in a location crucial for vRNA binding. Structurally, it presents the classical organisation of a RdRp. Analogous to a human right hand; it is configured successively in “finger, palm and thumb” domains. The “finger” and “thumb” domains present “fingertips” which are unstructured loops within the active site possessing the highly conserved polymerase pre-A/F and A to E motifs which catalyse the polymerisation reaction. One of these loops extended from two antiparallel β -strands (res 640-657) is believed to be a priming loop enabling primer independent replication. Structural alignments with other polymerases have determined the closest analogue structure to be from the HCV polymerase (Figure 15C) and Norwalk polymerase to a lesser extent (Figure 15B). Based on these homologues, a model for template entry, template/product exit and nucleotide entry has been proposed (Figure 15 D) and fits with the existing solvent tunnels within the structure.

Although structurally aligned, influenza B and bat influenza A crystals forms display a few distinct structural differences; the most important of which is the orientation of the cap-binding domain of PB2. In bat influenza A, the cap-binding site is found facing the endonuclease active site at a distance of 50 Å which is consistent with predicted cleaved mRNA length of 12-15 nt (Sikora et al., 2014) whereas in influenza B, it is observed tilted 70° down, which would lead the mRNA primer into one of the PB1 active site tunnels. In one of the two influenza B structures, there is even residual non-attributed electron density showing a ssRNA pattern bound to the cap-binding domain and directed within the active site. These fortuitous observations are but frozen states of the polymerase which probably adopts a multitude of conformations; however, they enrich the current models explaining cap-dependent transcription.

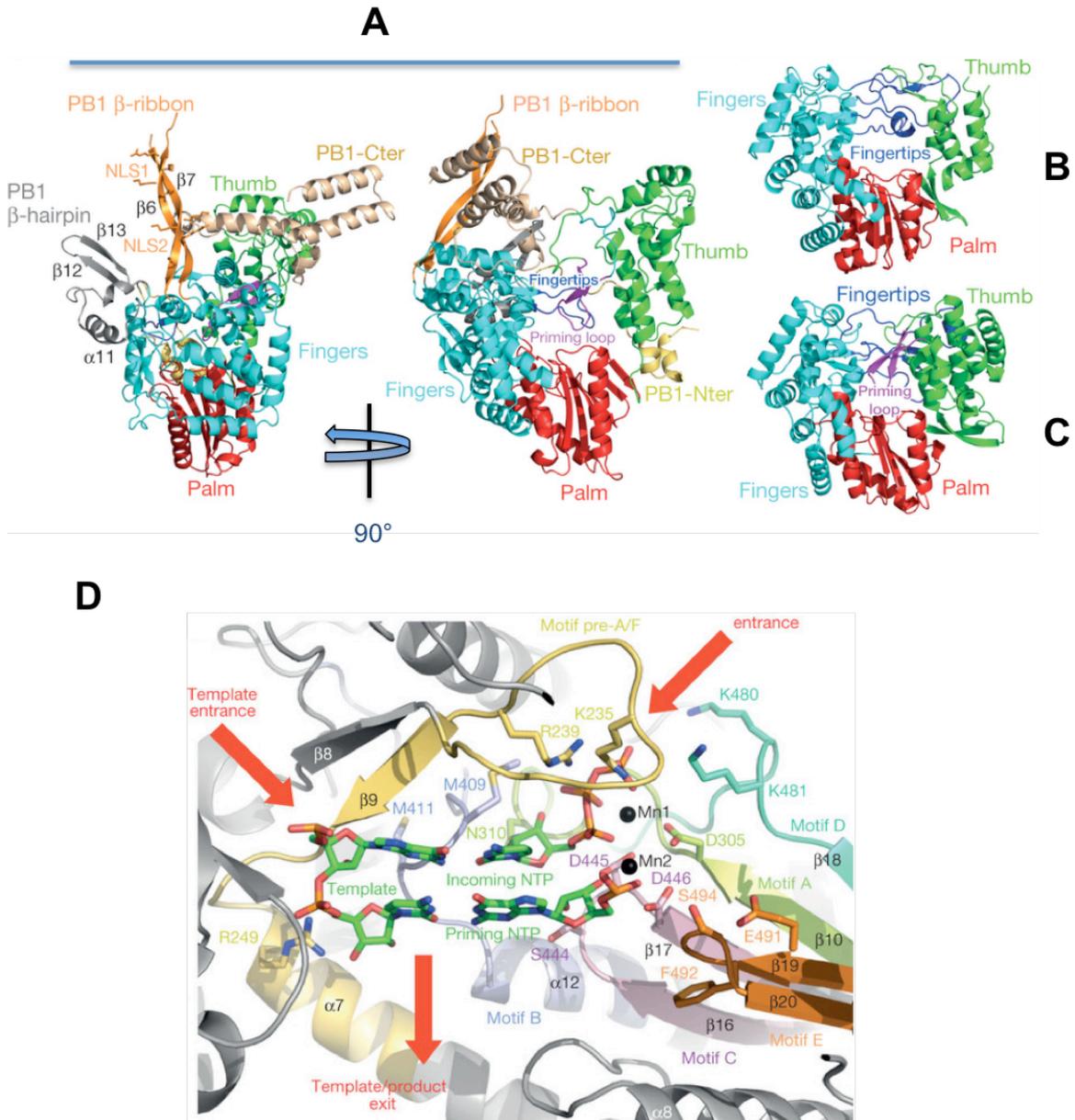


Figure 15: Structure of the PB1 polymerase active site (PDB: 4WSA). (A) Side and front view of PB1 drawn in a cartoon representation with a rainbow colouration of all sub-domains. (B) Superposed Norwalk polymerase (PDB: 3BSO) front view coloured with the same sub domain colour code as for PB1. (C) Superposed HCV polymerase (PDB: 2XI3) which possesses a priming loop coloured in purple. (D) Visualisation of the PB1 catalytic site in cartoon representation with a colouration distinguishing the different highly conserved pre A/F and A to E motifs. Based on the Norwalk polymerase catalytic (PDB: 3BSO) domain superposition, location of the yet undetermined metal ion configuration to initiate the polymerase reaction of incoming NTPs and product formation. Supposed NTP, template entrances and product/template exit channels are visualised with orange arrows. (Adapted from Pflug et al., 2014)

1.4.4 CURRENT MODELS FOR TRANSCRIPTION AND REPLICATION

1.4.4.A TRANSCRIPTION

As previously mentioned, transcription of vRNA enables the synthesis of positive strand mRNA, coding for viral proteins. Differently from other viral polymerase complexes, the influenza RdRp does not possess a guanylyl methyltransferase activity required for synthesising a 7-methyl-guanosine “cap” on the 5' extremity of the newly synthesised RNA. Yet the presence of a “cap” in 5' and of a poly-U tail in 3' is compulsory for translation of the novel mRNA. Therefore evolution has pushed the influenza virus to generate capped mRNAs through a cap dependent mechanism known as “cap snatching” (Bouloy et al., 1978; Plotch et al., 1979, 1981). This mechanism recruits all three of the polymerase subunits to perform different sequential tasks. First, the cap binding domain of PB2 will bind a nascent host cell mRNA through its cap-binding domain (Figure 16A). Experimental evidence seems to indicate that the RdRp recruits the host cell RNA polymerase II in order to be favourably exposed to nascent mRNA (Engelhardt et al., 2005; Mayer et al., 2007; Rodriguez et al., 2007), showing how the polymerase “hijacks” the host cells' own replication machinery. After binding the cap, the endonuclease on the N-terminus of PA cleaves the mRNA (Figure 16A) to a length of 12-13 nt with a specific preference for 3' ends of guanine bases (Datta et al., 2013). The cap-binding domain of PB2 is then thought to freely rotate the cleaved mRNA primer into the active site through a primer specific tunnel (Figure 16B), thus enabling a base pair formation to occur. This event triggers the elongation phase (Figure 16C) where the vRNA is read from the 3' extremity to the 5' end. Reaching the 5' end, the polymerase will then generate a poly A tail by a stuttering mechanism triggered by the template strand blocked on a poly U sequence by the 5' extremity, which maintains its interaction with the polymerase (Tiley et al., 1994; Poon et al., 1999; González and Ortín, 1999).

1.4.4.B REPLICATION

Replication is a series of sequential stages which ultimately generates multiple novel copies of vRNA. Mechanistic differences must exist during both transcription and replication stages in order for the polymerase to completely replicate the vRNA or cRNA back into vRNA, without capping those RNAs or adding a poly A tail. Also, during replication, NP must be recruited and oligomerize onto the nascent RNA. Hence, several factors probably regulate the switch between a replicative polymerase and a transcribing polymerase.

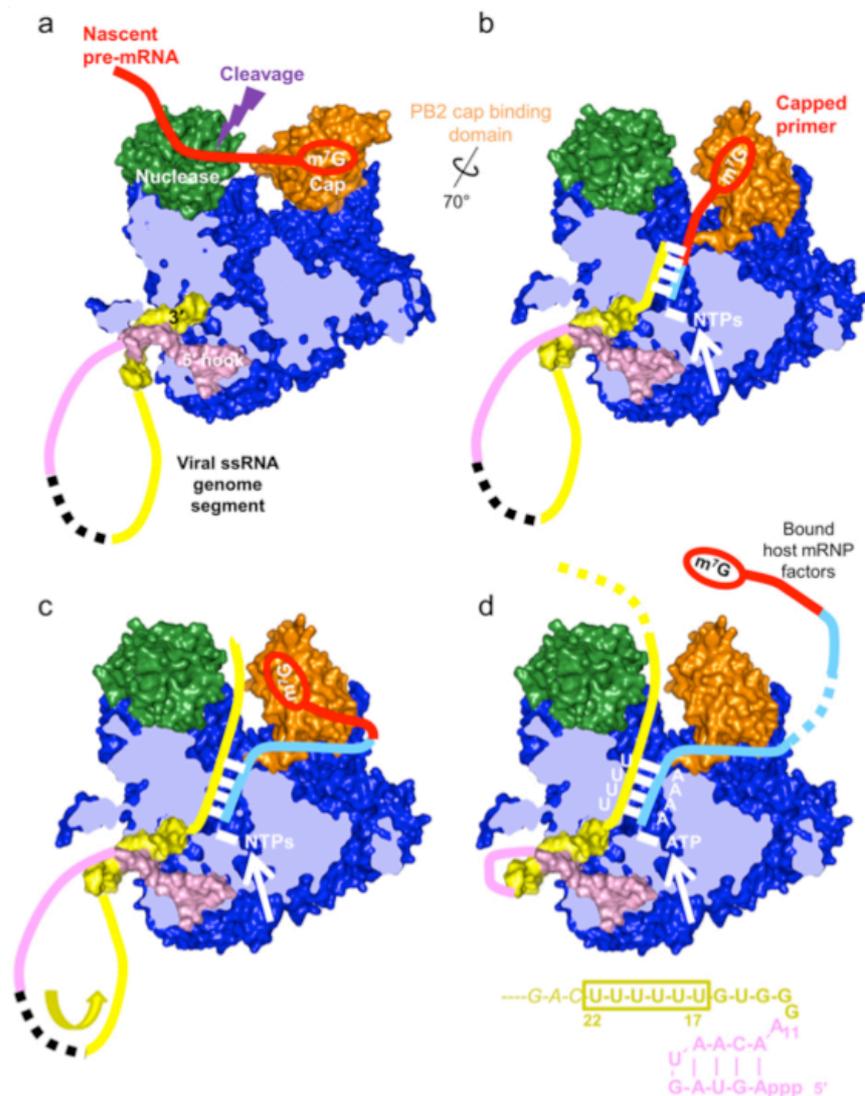


Figure 16: “Cap-snatching mechanism” modelled with the new polymerase structures. (A) The RdRp bound to vRNA catches a nascent mRNA with the PB2 cap binding domain in orange and cleaves it 12-13nt downstream with the PA endonuclease in green. (B) The cap binding domain then rotates 70° down to present the cleaved mRNA primer within the polymerase active site in blue where it forms a partial complementarity with the 3' end. (C) Elongation phase where both template and product are released as separate strands. (D) Poly A generated through a stuttering mechanism on the poly U motifs on the 5' represented in pink. (Adapted from Reich et al., 2014)

These factors may include cellular factors such as a host cell interactants like the RNA Pol II (Engelhardt et al., 2005) or the MCM replicative helicase (Kawaguchi and Nagata, 2007). However, strong evidence supports newly synthesised free NP as one of the key switches from a transcriptive to a replicative vRNP complex; this could be the effect of a polymerase/NP interaction (Biswas et al., 1998; Mena et al., 1999; Poole et al., 2004) or an interaction of NP with the vRNP template RNA (Hsu et al., 1987; Fodor et al., 1994; Klumpp et al., 1997). NEP also has been shown to downregulate *de novo* synthesis and increase transcription (Bullido et al., 2001;

Robb et al., 2009). Another important regulatory element could be the initiation mechanism. In both *de novo* synthesis steps, an ApG dinucleotide primer is synthesised by the polymerase. However, in the case of cRNA the ApG primes the +4 position requiring the product to realign with the +1 position (Deng et al., 2006a; Zhang et al., 2010).

The first stage consists in the synthesis of cRNA from vRNA during the early phases of infection and for a short period of time. The amplification process generating more vRNA seems to happen in later stages of infection (Hay et al., 1977). This observation also correlates with later studies which used thermosensitive mutants towards vRNA synthesis yet showed normal levels of cRNA accumulation, suggesting that cRNA synthesis is a restricted process mostly performed on the incoming vRNAs (Wolstenholme et al., 1980; Thierry and Danos, 1982; Falcón et al., 2004)

In a later stage, the polymerase performs a second *de novo* synthesis on the cRNA to generate vRNA. For this second step, a new model has emerged from a trans-complementation experiment showing that cRNP would be used as a template by a second “trans” non-resident polymerase (Jorba et al., 2009; Resa-Infante et al., 2011). This “trans” model (figure 17A) can be rationalised if the 3’ vRNA promoter can be swapped and initiates with the incoming polymerase, which then proceeds into an elongation phase. The newly synthesised 5’ vRNA would then recruit another non-resident polymerase and free NP would oligomerise on the rest of the vRNA. This model also allows a chain of polymerase to perform replication on the same cRNA leading to an amplification of the vRNA.

Another “trans” model has also recently been proposed where the initial non-resident polymerase would directly bind to the nascent 5’ vRNA (Moeller et al., 2012). Termination in any case would also imply the release of the 5’ vRNA from the resident polymerase in order for a full read-through to occur (Figure 17B). In these trans models, the 5’ vRNA sequence would play a pivotal role as it would actually regulate the resident polymerase into only performing “cis” transcription. The bound 5’ vRNA would impose a steric block, inducing poly-A synthesis during termination. Observations of a specific rigid hook structure of the 5’ vRNA when bound to the polymerase (Pflug et al., 2014; Reich et al., 2014) and the specific biochemical recognition of the 5’ towards the polymerase (Tiley et al., 1994; González and Ortín, 1999) seem to support such models.

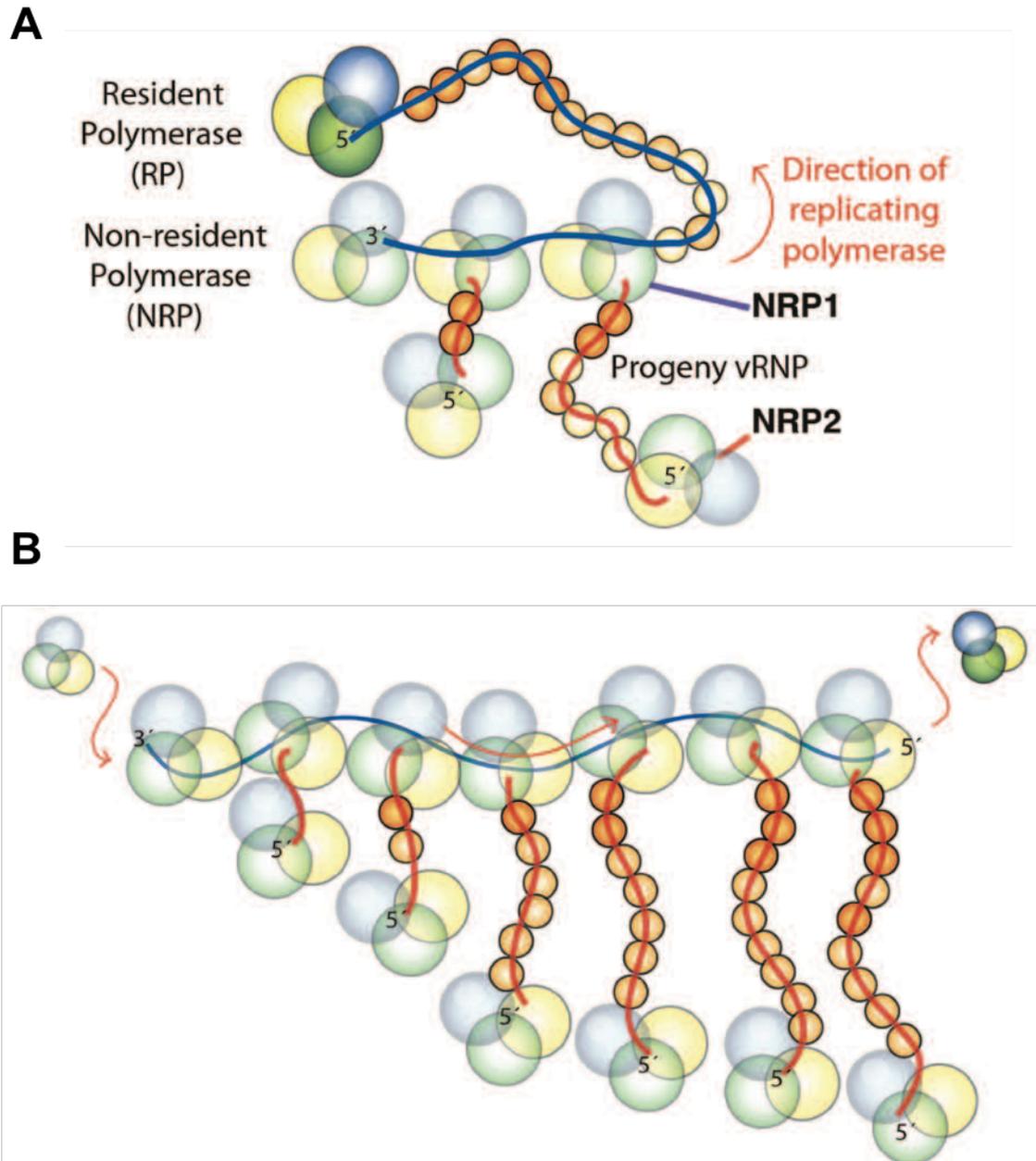


Figure 17: "Trans" model for vRNP synthesis. (A) Initiation and elongation phase of a non-resident polymerase on a released 3' vRNA promoter. The nascent 5' is bound by a second incoming non-resident polymerase. (B) Termination step where the resident polymerase is forced to dissociate from the 5' vRNA end. (Adapted from Resa-Infante et al., 2011)

1.4.5 ASSEMBLY PATHWAY OF THE POLYMERASE

As previously described, the RdRp is an intricate multi-subunit protein performing a plethora of activities and interacting with a number of host cell and viral proteins. The detailed structural fold of all subunits implies a complex assembly process requiring specific refolding kinetics of all subunits in order to generate functional complexes. As previously mentioned, the karyopherin- α and β class of

proteins play a major role in this process. They transport the different subunits from the cytosol, where they are expressed, to the nucleus where they will assemble into fully formed complexes to then perform transcription and replication. When looking at the RNP as a whole, import and export mechanisms are required during the initial and final stages of infection (as previously described in chapter 1.2.3.B and 1.2.3.D)

Importantly, for the polymerase complex assembly within the nucleus, this import process takes place with two distinct nuclear import routes (Figure 18A).

PB2 on one side is imported by recruiting the importin- α isoforms (Figure 18), giving access to the classic nuclear import pathway (Görllich and Kutay, 1999; Marfori et al., 2011). This import pathway is a multi-step process in which importin- α will bind to the bipartite NLS of PB2 (Mukaigawa and Nayak, 1991; Tarendeau et al., 2007), then recruit importin- β 1 for nuclear translocation. Upon entry within the nucleus, the importin- β 1 will release the importin- α when bound to RanGTP which acts as the association/dissociation regulator of nuclear import. This import pathway, although conventional, relies on a conformational state of PB2 rendering the NLS domain accessible, which has not been observed when the NLS domain is expressed with the C-terminal domain (Tarendeau et al., 2008; Kuzuhara et al., 2009). These data suggest a required conformational change in PB2 for importin- α binding. The interaction with importin- α isoforms is also a crucial element in host adaptation and cellular tropism (Resa-Infante et al., 2008; Boivin et al., 2010; Resa-Infante et al., 2014). Also, during the release from importin- α , other cellular factors such as Nup50 from the NPC may play a role in maintaining PB2 from re-binding importin- α (Pumroy et al., 2012). Additionally, the chaperone protein Hsp90 seems to play a role in stabilising PB2 during this process and is imported together with PB2 in the nucleus during infection (Naito et al., 2007; Chase et al., 2008).

PA and PB1 seem to associate upon expression within the cytoplasm possibly through the recruitment of additional cellular factors; also, they seem strictly interdependent on each other for nuclear accumulation (Nieto et al., 1992; Fodor and Smith, 2004; Loucaides et al., 2009; Huet et al., 2010). Both proteins contain NLS signals. PA has two NLS motifs, one (res 124-139) which is on the endonuclease domain. The second (res 186-247) is located on the PA hinge region linking the endonuclease to the C-terminal domain on the influenza A structure (Pflug et al., 2014). Most importantly, PB1 has a bipartite NLS domain (res 187–190 and 207–211) located on the extremity of the long and highly exposed “ β -ribbon” motif on the full polymerase structure (Pflug et al., 2014; Reich et al., 2014). It has been shown to recruit RanBP5 (Deng et al., 2006b; Hutchinson et al., 2011). Also called β -importin Ran Binding protein 5, importin 5 or Karyopherin β -3, this β -class importin is the only

described importin for the import of PA-PB1 and, contrary to PB2, does not seem to be subject to host restriction (Hutchinson et al., 2011). With this regard, the bipartite NLS of PB1 sequence is found strongly conserved among many influenza A subtypes (Figure 18B) and the inhibition of interaction with PA-PB1 through mutagenesis affects viral proliferation (Hutchinson et al., 2011; Hutchinson and Fodor, 2012).

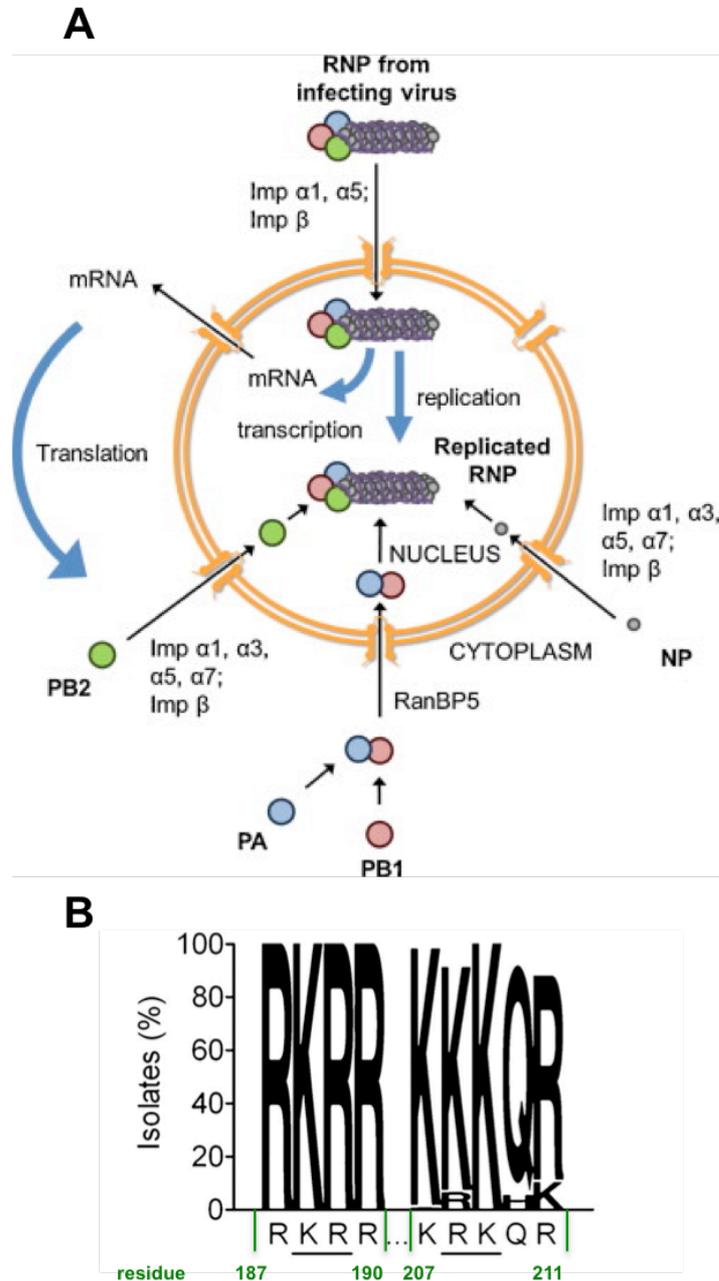


Figure 18: Nuclear import routes of full RNPs and single constituents through the NPC (PA, PB1, PB2 and NP). (B) Residue propensity on the PB1 bipartite NLS after a multiple sequence alignment on human and avian influenza A strains. (Adapted from Hutchinson et al., 2012 and Hutchinson et al., 2011)

RanBP5 is a non-classical import factor binding the cargo protein directly without an importin- α intermediate (Jäkel and Görlich, 1998). It is however regulated, like other conventional β -importins through the interaction with RanGTP. Other than the influenza polymerase, RanBP5 has been shown to enable the import of other proteins involved in nucleic acid binding such as ribosomal proteins (Jäkel and Görlich, 1998), core histones (Baake et al., 2001) and other core cellular proteins (Ross et al., 2003; Fu et al., 2006; Chung et al., 2008). Moreover, it is also involved in the interaction and/or import of a multitude of other viral proteins such as the rev protein from HIV (Arnold et al., 2006), several of the Human Papilloma virus core proteins, such as L1-L2-L2 capsid proteins (Nelson et al., 2003; Darshan et al., 2004; Klucsevsek et al., 2006) and the E5 oncoprotein (Krawczyk et al., 2008). More than a nuclear import factor, RanBP5 also plays a chaperone role by interacting with highly basic motifs on RNA binding proteins (Jäkel et al., 2002), a role which it is also believed to play with the PA-PB1 polymerase dimer (Hutchinson and Fodor, 2012; Hutchinson et al., 2011).

Although structurally un-characterised, the overall fold of the protein will be comparable with that of other Karyopherin- β proteins, several of which have been crystallised in free, NLS bound or RanGTP bound forms (Cook et al., 2007; Xu et al., 2010). However, to date no model exists describing an importin- β bound to a complete cargo protein. The closest protein sequence homologue to be structurally characterised was the yeast Kap121p (Kobayashi and Matsuura, 2013) which shares 30% of sequence identity with RanBP5. Kap121p is organised as a right hand super-helix constituted of 24 HEAT repeats which are two A and B antiparallel α -helices connected by loops of varying length (Figure 19A). As seen on the Figure 19A, these helices are exposed in a consistent fashion throughout the whole structure. A- α -helices form the outer convex surface and will interact with the NPC during nuclear import whereas B- α -helices form the inner surface which mediates both NLS and RanGTP binding in the N-terminal half of the molecule. The N-terminal and C-terminal first and last heat repeats are tilted, compared to the other heat repeats. This has the effect of capping the structure, which also packs with itself through an unfolded loop domain insertion (H15 or H18). Strong conformational changes occur upon binding of either a short NLS motif or RanGTP. In the case of Ran GTP, the binding event is shown to push out the bound NLS which is on an adjacent binding site (Figure 19B), the overall structure additionally undergoes conformational changes in the C-terminal half of the molecule.

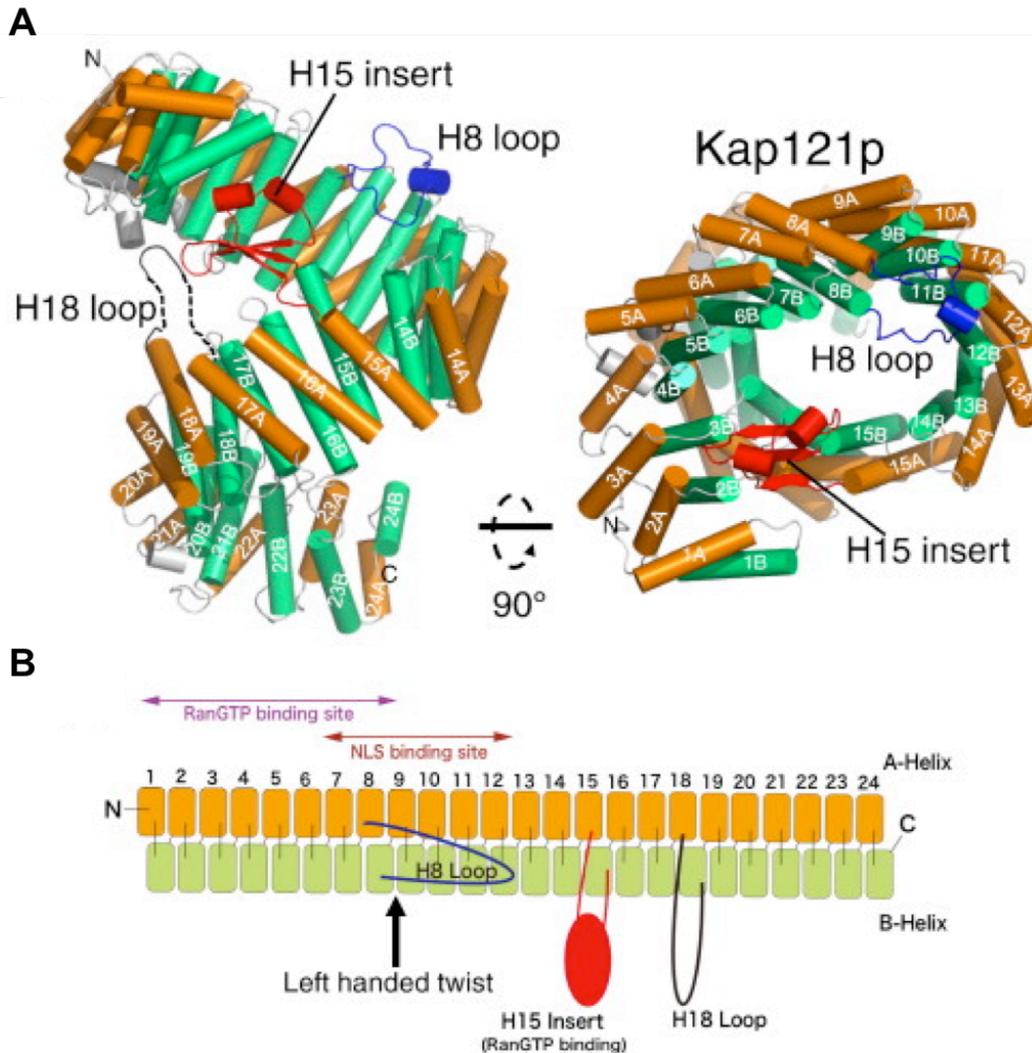


Figure 19: Kap121p structural architecture. (A) Cartoon depiction of Kap121p with heat repeats coloured in orange for A α -helices and green for B α -helices. Unstructured loops H8 and H18 are coloured respectively in blue and black, the H15 insert is coloured in red. (B) Schematic view with the same colour code, RanGTP and NLS binding sites are detailed in purple and red. (Adapted from Kobayashi et al., 2013)

For RanBP5, a comparable structural behaviour can be expected. There is however a multitude of binding strategies for β -importins with conventional bipartite NLS motifs. Structural information on this binding process, combined with the understanding of the potential chaperone role played by RanBP5 on PA-PB1 could open a new pathway for potential inhibition of the polymerase assembly process. This in turn could lead to a new druggable target in the fight against influenza.

1.5 THESIS OBJECTIVES

In the initial phase of the project, the only structural data describing the influenza A virus (IAV) RNA-polymerase was either fragmented or of low resolution. Most of crystal structures of soluble domains on the polymerase had already been described. As such, the biggest unknown feature was PB1 and parts of PB2. The global fold of the polymerase was also unknown although several low-resolution EM maps existed.

Our group set out to express the full polymerase using recombinant polyprotein baculovirus technology. It was quickly found that full polymerase expression for the H3N2 and H5N1 strains was very difficult, especially in the quantities required for structural biology studies. Early crucial work established a truncated trimer construct which could express PA-PB1-PB2(1-116). This truncated polymerase was both replicase and endonuclease active. Historically, this construct was most probably one of the first to enable recombinant expression of PB1 in the world. It however never crystallised for reasons that will be discussed in this thesis. Within this context, my thesis project was both to find a minimal crystallising domain for the polymerase but also investigate the factors limiting full polymerase expression and propose eventual solutions to these problems. As such, multiple truncated polymerase constructs were designed with the intent to produce crystallising polymerase core domains which all included PB1. Some of these constructs, including the polymerase heterodimeric assembly PA-PB1(1-686) could be purified without glycerol and to very high concentrations. These new properties enabled for Small Angle X-Ray Scattering (SAXS) to be performed and made the protein easier to use in affinity binding measurements.

In parallel to working on the polymerase, my project also focused on expressing the cellular factor RanBP5 for crystallographic studies. In the later stages of my thesis, RanBP5 was inserted within the polymerase polyprotein expression construct with the intent to reconstitute the import complex. This thesis therefore initially focused on two very different protein complexes, however towards the end of the thesis, both projects came together in the form of the PA-PB1(1-686)-RanBP5 complex.

Finally, the production of an important construct library enabled us to set up and validate small scale high-throughput insect cell experiments with the intent of establishing a robust protocol for polymerase complex screening. This methodology will also enable wider screening of any multi-protein complex expressed in insect cells.

CHAPTER 2: EXPERIMENTAL PROCEDURE

Résumé du chapitre 2 en Français

Ce projet a fait intervenir plusieurs techniques et méthodes expérimentales ayant conduit aux résultats présentés dans ce manuscrit.

Une stratégie d'expression de la polymérase tronquée a été mise en place en cellules d'insecte afin de produire une polyprotéine clivée à la TEV. Pour ce faire, des techniques de biologie moléculaire ont permis le clonage de gènes synthétiques de polymérase et de RanBP5 dans des vecteurs de transfert. Ces derniers ont été utilisés pour générer des baculovirus recombinants.

Les méthodes de purification des différents complexes protéiques par lyse cellulaire, résine d'affinité au nickel et chromatographie d'exclusion de taille sont aussi détaillées dans ce chapitre.

Une série de méthodes permettant la caractérisation fonctionnelle de PA-PB1 est décrite avec notamment les méthodes de détermination de constante de dissociation (K_d) par anisotropie de fluorescence et rétention sur membrane ou « Filter Binding Assay ». Le thermofluor est présenté comme méthode qualitative afin de déterminer la température de fusion d'une protéine. Deux méthodes d'analyse des protéines purifiées sont décrites. Tout d'abord, la chromatographie d'exclusion de taille couplée à un système MALLS-refractomètre (Multiple Laser Light Scattering) permettant de déterminer précisément la masse moléculaire d'un complexe ainsi que sa monodispersité. Ensuite, la microscopie électronique à transmission qui permet d'estimer la pureté de l'échantillon analysé.

Deux techniques de caractérisation structurales sont aussi détaillées, dont la diffusion des rayons X aux petits angles (SAXS) permettant la détermination d'une enveloppe basse résolution du complexe à étudier. Enfin, une portion de ce chapitre décrit le principe et la mise en place de techniques de cristallisation et diffraction de protéines aux rayons X.

2.1 CLONING OF THE RdRp POLYPROTEIN GENES AND BACULOVIRUS GENERATION

All polymerase fusion constructs detailed within this thesis were cloned within the same “cassette” cloning vector called K7-pMARq which carries a resistance gene to ampicillin. This cloning vector is composed of an adaptor sequence flanked by BstEII and RsrII cleavage sites (Figure 20). The cassette was designed to directly clone up to three different proteins, each within the same reading frame and separated by a TEV cleavage site linker GSGSG(ENLYFQG)AGSGSGSG. An 8-histidine tag is present and is coded within the N-terminus of the polyprotein. For polymerase constructs, PA, PB1 and PB2 genes from A/Viet-Nam/1203/2004(H5N1) and A/Victoria/3/1975(H3N2) strains (UniProtKB access number Q6DNN3 and H9XIJ5 respectively) were synthesised by GeneArt (Life Technologies) and separately cloned within the A, B and C sites respectively (Figure 20).

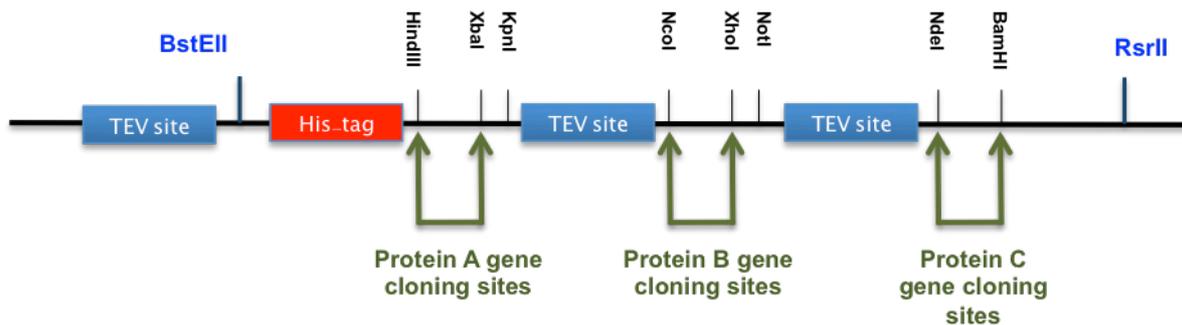


Figure 20: Cassette cloning sites within the K7-pMARq vector. Genes of interest are separately cloned within the A, B and C gene cloning sites located between TEV cleavage sites.

To proceed with the bacmid generation stage, the polyprotein fusion gene was sub-cloned within the pBAC vector (Imre Berger, EMBL Grenoble) using the BstEII and RsrII enzymes. Positive clones are selected by the gentamycin resistance present on the pBAC. Once cloned within pBAC, the original fusion protein is integrated within a larger polyprotein ORF which now additionally has TEV expressed in the N-terminus and CFP (Cyan Fluorescent Protein, $\lambda_{\text{excitation}} = 430\text{nm}$ and $\lambda_{\text{emission}} = 450\text{-}550\text{nm}$) in its C-terminus.

Bacmid generation is undertaken using the established protocols (Berger et al., 2013; Trowitzsch et al., 2010). First, DH10EMBacY cells (Imre Berger) are transformed with purified pBAC vector and left overnight at 37°C in LB (Leman Broth) media. This overnight incubation enables the pBAC plasmid containing the

fusion gene to integrate the EMBacY baculovirus genome through a Tn7 transposition mechanism. Positive clones are selected on Luria Broth (LB, Sigma Aldrich) agar plates containing tetracyclin, gentamicin, kanamycin and distinguished from non re-integrated pBAC/EMBacY by their white phenotype in the presence of BluOgal (Sigma Aldrich) and IPTG (Sigma Aldrich). The Tn7 transposition site is indeed positioned within the LacZ gene, positive transpositions therefore disrupting the positive β -galactosidase phenotype which otherwise produces blue colonies. After a second isolation step on another Tet/Gent/Kan/IPTG/BluOgal LB agar dish, positive clones can inoculate a 2 ml LB culture bacmid prep. Bacmid DNA is purified with a standard plasmid miniprep kit (Qiagen) which relies on an alkaline lysis of the bacteria and precipitation of genomic DNA. After retrieving the supernatant, which contains soluble bacmid, a DNA precipitation using isopropanol/ethanol is used. The DNA pellet is then dried of ethanol and resuspended under the sterile hood and used immediately in the transfection stage.

Transfection is performed on adherent Sf21 (*Spodoptera frugiperda*) insect cells within 6 well plate dishes and performed in duplicates for every bacmid. Chemical transfection is performed using X-tremeGENE HP DNA Transfection Reagent (Roche). After 60 h the well growth media is harvested, it constitutes the first baculovirus amplification (V0). The V0 is then used to infect a small (25 ml) suspension culture of sf21 cells at 0,5 cells/ml. After, infection cells are daily counted and maintained at 0,5-1,0 $\cdot 10^6$ cells/ml until they stop dividing which occurs generally 24 to 48 h after infection. Following the day of proliferation arrest stage (dpa), YFP (Yellow Fluorescent Protein, $\lambda_{\text{excitation}} = 488$ nm and $\lambda_{\text{emission}} = 515 - 550$ nm) fluorescence is monitored using a Tecan Infinite M200 Pro fluorometer until a fluorescence intensity signal is observed. At this moment, the cultures are centrifuged at 100 g for 5 min and the supernatant is harvested and constitutes the second virus amplification (V1). The remaining cell pellet is resuspended within an equal volume of fresh media and monitored until the YFP and CFP signals reach a plateau. After reaching a plateau, the cells are pelleted (10 min at 400 g) and frozen in order to perform small-scale purification tests.

YFP and CFP serve as two very important fusion genes within this system. YFP is a transcriptional fusion to the polyhydriin promoter and reports baculovirus proliferation. CFP on the other hand is a translational fusion gene, which is the last translated protein within the polyprotein system and therefore directly reports on the efficacy of expression of the full polyprotein. (Note: A more visual representation of the method can be viewed in the figure 1 of **Chapter 3**).

2.2 HETEROTRIMERIC AND HETERODIMERIC POLYMERASE COMPLEXES PURIFICATION

Large scale suspension culture expressing polymerase fusion constructs were prepared using High Five insect cells grown in Express Five media (Life Technologies) at $0.5 \cdot 10^6$ cells/ml infected at 0.2 % (V/V) with the baculovirus mother solution. Cultures were maintained at 0.5 to $1 \cdot 10^6$ cells/ml until proliferation arrest (24-48h after infection). Following the proliferation arrest, YFP and CFP measurements were performed on $1 \cdot 10^6$ cells which were spun down and sonicated in 500 μ L of PBS buffer. YFP/CFP measurements were performed every 12 h until a fluorescence plateau was reached (72-96h after infection). Cultures were then spun down at 400 g for 20 min and cell pellets were stored at -80°C .

During purification, cell pellets were resuspended in 50 ml of lysis buffer (50 mM Tris pH 8.5, 300 mM NaCl and 2 mM beta-mercaptoethanol (BME)) for $5 \cdot 10^8$ cells in the presence of EDTA-free anti-protease cocktail (complete from Roche). Lysis was performed by two cycles of freezing (-180°C) / thawing (26°C) after which 10% glycerol was added to the lysate before centrifugation (45 min, 40 000 g, 4°C). After retrieval of the clarified lysate, 30 mM of imidazole pH 8 was added before loading on Ni-NTa superpose resin (Quiagen). After flowing through the lysate, the resin was washed with 10 column volumes (cv) of buffer A (lysis buffer with 30 mM imidazole), 10 cv of wash buffer B (50 mM tris pH 8.5, 1M NaCl, 10% glycerol and 2 mM BME) and 10 cv of wash buffer A. Elution of the bound complex was performed with 300 mM Imidazole. Elution fractions containing polymerase were then pooled together and directly injected on a 5 ml Hitrap Heparin resin (GE Healthcare) which had previously been equilibrated with 4 cv of buffer A. After binding to the resin, a 5 cv wash was performed with buffer A before eluting with a 40 ml salt gradient on an äkta prime system (GE Healthcare). The elution peak corresponding to stoichiometric polymerase assemblies was then pooled and injected on an S200 (GE Healthcare) size exclusion chromatography column equilibrated in 50 mM Tris pH: 8.5, 300 mM NaCl, 5 mM BME and run on an äkta purifier system. Peak fractions were finally pooled and concentrated from 10 to 20 mg/ml using a 100 kDa cutoff concentrator (Amicon). For long-term storage, 20% glycerol was added to the protein solution which was then stored at -80°C .

2.3 CLONING OF RANBP5

The RanBP5 synthetic gene was ordered at Lifetechnology-Genart and subcloned within the pFast-BAC (Life Technologies) vector using NcoI and XhoI DNA cleavage sites. Positive clones were grown on ampicillin LB agar plates from which a midprep DNA purification was prepared. For the baculovirus generation step, DH10EMBacY cells were transformed with 2 µg of DNA. The following steps towards baculovirus generation are identical as those described in **chapter 2.1**. Importantly, during the expression of this protein in insect cells only YFP fluorescence is monitored in order to follow cell growth parameters.

2.4 PURIFICATION OF RANBP5

Large scale suspension culture expressing RanBP5 were prepared using High Five insect cells grown in Express Five media (Life Technologies) at $0.5 \cdot 10^6$ cells/ml infected at 0.2% (V/V) with the baculovirus mother solution. Cultures were maintained at 0.5 to $1 \cdot 10^6$ cells/ml until proliferation arrest (24-48h after infection). Following the proliferation arrest, YFP measurements were performed on $1 \cdot 10^6$ cells that were spun down and sonicated in 500 µL of Phosphate Buffered Saline (PBS) buffer. YFP measurements were performed every 12h until a fluorescence plateau was reached (72-96h after infection). Cultures were then spun down at 400 g for 20 min and cell pellets were stored at -80°C.

For purification, cell pellets were re-suspended in 50 ml of lysis buffer (50 mM Hepes pH 7.5, 300 mM NaCl and 2mM BME) per $500 \cdot 10^6$ cells in the presence of EDTA free anti-protease cocktail (complete from Roche). Lysis was performed by two cycles of freezing (-180 °C) / thawing (26 °C) after which 5% of glycerol were added to the lysate before centrifugation (45 min, 40 000 g, 15°C). All further purification steps were then performed at room temperature (20 °C). After retrieval of the clarified lysate, 30 mM of imidazole pH 8 was added before loading on Ni-NTa superpose resin (Quiagen). After flowing through the lysate, the resin was washed with 10 column volumes (cv) of lysis buffer containing 30 mM of Imidazole. Elution was then performed by increasing the Imidazole content to 300 mM. Following elution, the protein fractions were pooled, supplemented with TEV protease at a 100/1 (w/w) ratio and dialysed overnight in 50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM BME. After dialysis, the protein sample was flowed through 300 µL of Ni-Nta resin. The final purification stage consists in a size exclusion chromatography (SEC): the sample is injected on an S200 (GE healthcare S200 300/GL, 125 ml) column

using an äkta purifier system which monitors sample elution at 260 and 280 nm. The peak fractions are concentrated to the desired concentration using a 50 kDa Amicon concentrator.

2.5 THERMOFLUOR

2.5.1 PRINCIPLE

Thermofluor is an indirect method belonging to the thermal shift assay methods (Semisotnov et al., 1991) used to study the melting behaviour of a protein in a given condition. The method relies on the use of a hydrophobic fluorescent dye (Sypro) which is strongly quenched by water. As a protein is melted with an incremental increase in temperature, it will progressively expose the hydrophobic inner core. Sypro will bind to these hydrophobic patches, which indirectly results in an increase of Sypro fluorescence. Upon reaching a maximum of fluorescence, which corresponds to a fully unfolded state, fluorescence will then decrease as sypro dissociates from the protein in the higher temperatures. The T_m of a given protein can be determined by estimating the temperature at the midway point between the minimal and maximal fluorescence.

Thermofluor has the advantage of being very easy and quick to setup and can inform on the stability of a protein with regards to a buffer, additives or ligand conditions.

2.5.2 EXPERIMENTAL PROCEDURE

Thermofluor experiments were performed following the same established protocols (Pantoliano et al., 2001). Samples were diluted at 10 μ M in a buffer consisting of 20 mM HEPES pH: 7,5, 150 mM NaCl, 2 mM $MgCl_2$ and 5X Sypro Orange (Invitrogen) at a final volume of 45 μ L within 96 well iCycler iQ plates (Biorad). The thermal stability experiment was measured using a real time q-PCR machine (Stratagene Mx3005P) performing a temperature gradient of 25 $^{\circ}C$ to 95 $^{\circ}C$ with a 1 $^{\circ}C$ / min increment. At every increment, sypro fluorescence was measured with an excitation wavelength of 492 nm and an emission wavelength of 512 nm. The relative fluorescence emission was then plotted against its corresponding temperature to produce the thermal shift profile curve. The T_m was estimated from the curve as the midpoint temperature between the baseline and maximum value of the curve.

2.6 AFFINITY BINDING TECHNIQUES

2.6.1 FLUORESCENCE ANISOTROPY

Fluorescence anisotropy or fluorescence polarisation relies on the intrinsic properties of fluorophores to absorb and re-emit polarised light along different axes of polarisation.

Fluorescence is derived from the ability of a molecule to absorb photons which induce electrons to migrate to higher energy states becoming “excited”. Upon excitation, the transition back to the ground state will re-emit energy under the form of both light and heat; the partial loss of radiation energy will therefore generate an emitted beam with a higher wavelength. When exciting a fluorophore with a polarised plane beam, fluorophores will be excited only when their transition moment is parallel to the electric vector of the incident beam. As well as absorption, emission of polarised light is also controlled by the transition moment orientation but also by the fluorescence lifetime, which differs between fluorophores (usually between 1 to 20 nanoseconds). In solution, where fluorophores can freely rotate, the degree of polarisation of the emitted light is reduced. Anisotropy, which describes a fluorophores’ ability to maintain this polarised plane, can be described by the following formula:

$$r = \frac{r_0}{1 + \tau/\phi}$$

where r is the observed anisotropy, r_0 the intrinsic anisotropy of a given fluorophore, τ the fluorescence lifetime and ϕ is the rotational time constant. For a hydrodynamic sphere in solution, ϕ can also be expressed with the following formula:

$$\phi = \frac{\eta V}{RT}$$

where R is the gas constant, T the temperature, η the solution viscosity constant and V the molecular volume. As such, when performing a titration of fluorescently labelled RNA at constant temperature and viscosity, the only factor changing upon binding a protein is V , hence increasing ϕ in the general anisotropy formula, therefore increasing anisotropy (Figure 21).

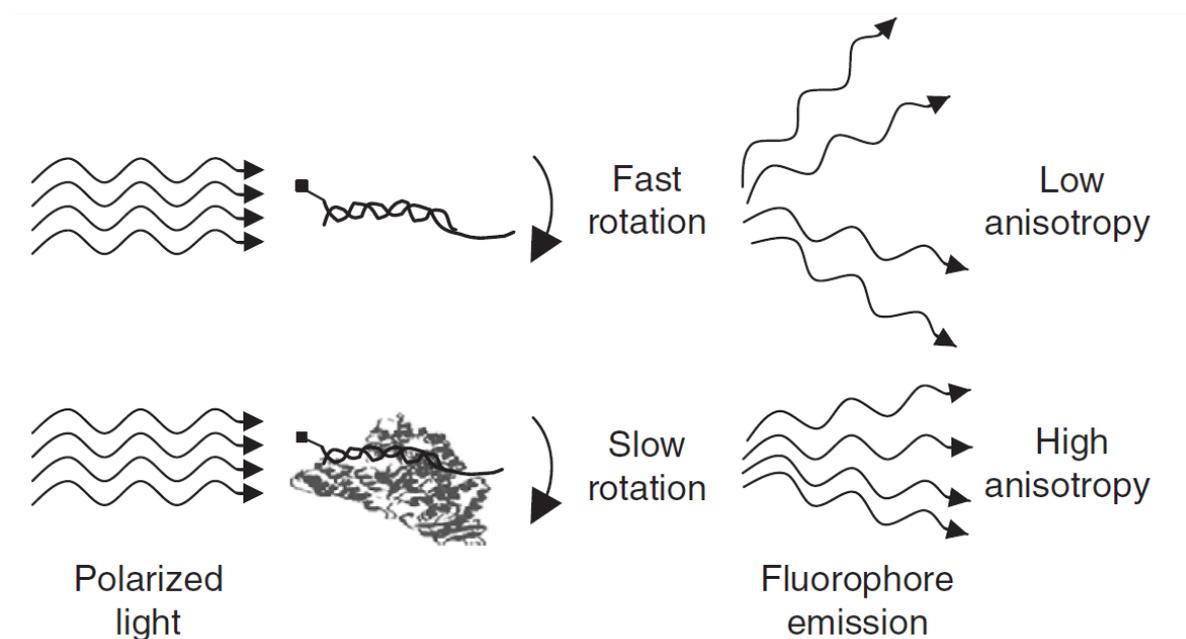


Figure 21: Schematic representation of the anisotropy of a fluorescent DNA molecule. Anisotropy is increased upon binding a larger partner which slows the tumbling rate of the DNA molecule. (Adapted from LiCata and Wowor, 2008)

In practice, fluorescence anisotropy titrations require the use of a high sensitivity fluorometer possessing two electro mechanical polarisers, one positioned at the excitation beam and the other on the emission path. For each titration point, two fluorescence measurements are required with polarisers either aligned (\parallel) or perpendicular to each other (\perp). Through these experimental values, anisotropy can be expressed with the following equation:

$$r = \frac{(I_{\parallel} - I_{\perp})}{(I_{\parallel} + I_{\perp})}$$

where I_{\parallel} and I_{\perp} are the fluorescence intensity values for the aligned and perpendicular planes respectively.

Anisotropy affinity titrations are usually performed using low nanomolar concentrations of RNA. However, due to inherent sensitivity limitations, sub nanomolar concentrations cannot be used. Another important factor to consider is the effect of the fluorophores moiety on the binding kinetics. This effect can usually be verified using a competition assay where labelled RNAs are outcompeted with non-labelled RNA. Upon data acquisition, the change in anisotropy (Δr) or subtracted anisotropy can be plotted against the protein concentration to produce equilibrium binding curves.

2.6.2 FILTER BINDING ASSAY

Filter binding assay (FBA) experiments require the use of radiolabelled RNAs; as such, a 5' P³² labelling is undertaken using the T4 polynucleotide kinase on every RNA to be tested. The use of P³² radiolabelled RNAs greatly increases the detection sensitivity when compared to fluorescence based methods. However, the short beta decay of P³² (14 days) does not enable prolonged conservation of labelled RNAs upon labelling. When recently labelled, RNA concentrations remain detectable well into the picomolar range enabling low nanomolar and sub-nanomolar affinity determination. FBA uses membrane immobilisation of the free or bound labelled RNA. In a double filter binding experiment (Wong and Lohman, 1993), affinity titration points are filtered onto two membranes. The first membrane (protran or PVDF, Amersham Systems) generally retains proteins; the second (Hybond-N+, Amersham Systems) will bind nucleic acid. Membranes are then “counted” by using a scintillation counter or revealed through the use of a phosphorescent screen. Upon revelation, spot intensity is integrated by imagery software. For each titration point, a bound RNA fraction is calculated using the following formula:

$$\text{Bound fraction} = \frac{I_{\text{protein}} - I_{\text{RNA}}}{(I_{\text{protein}} - I_{\text{RNA}}) + I_{\text{RNA}}}$$

Where I_{protein} and I_{RNA} are the relative intensities of the Protran membrane filter and Hynbond-N+ filter respectively. Upon calculation, the RNA bound fraction ratio can be plotted against titrant concentration in order to produce an equilibrium titration curve.

2.6.3 EXPERIMENTAL METHODS

Experimental procedures for affinity binding titrations and data analysis are detailed in the **chapter 3** methods part.

2.7 SIZE EXCLUSION CHROMATOGRAPHY COUPLED WITH MALLS

(MALLS: Multi Angle Laser Light Scattering)

2.7.1 PRINCIPLE OF MALLS

SEC-MALLS-RI is an useful analysis to undertake prior to any high cost biophysical characterisation or crystal screening. Indeed, success or failure of these different methods depends largely on the dispersity and oligomeric state of a given protein sample. During the last stage of purification, SEC combined with an SDS PAGE analysis can already give an estimate of the oligomeric state of a protein through the use of calibrated columns. However, SEC alone does not give an estimate of dispersity (monodisperse/polydisperse) within a single elution peak that can sometimes be polydisperse. Also, some proteins can be attracted or repulsed by the SEC (sephadex) sugar matrix, which can bias the correlation between protein mass and SEC elution volume.

SEC-MALLS-RI can accurately analyse protein samples through the combined use of Multi-Angle Laser Light Scattering (MALLS) and Refractive Index (RI) (Wyatt, 1991, 1993). MALLS uses a polarised and highly collimated laser beam focusing on the sample cell that is surrounded by several detectors measuring each a single angle of diffusion. Every angle intensity measurement is integrated and enables for the estimation of molecular weight through a correlation plot called the Debye Plot that requires a precise measurement of concentration. This measurement is possible through the combined used of a refractometer, which gives a concentration calculation through the measurement of a refractive index (Figure 22A). The final curve usually plots the excess of refractive index (or UV absorbance) and molecular weight as a function of volume. Analysis of this curve enables an estimation of dispersity through the observation of a stable molecular weight estimate within a single peak (Figure 22B).

2.7.2 EXPERIMENTAL PROCEDURE FOR MALLS

All MALLS runs were performed using a 25 ml S200 SEC column (GE healthcare) connected to a Optilab T-rEX refractometer (Wyatt technologies) and a DAWN HELEOS-II multi angle light scattering detector measuring 18 scattered angles (Wyatt technologies). Prior to injection, columns and systems were

equilibrated in 3 to 5 column volumes. 50 μL injections were performed using protein samples concentrated at a minimum of 2 mg/ml.

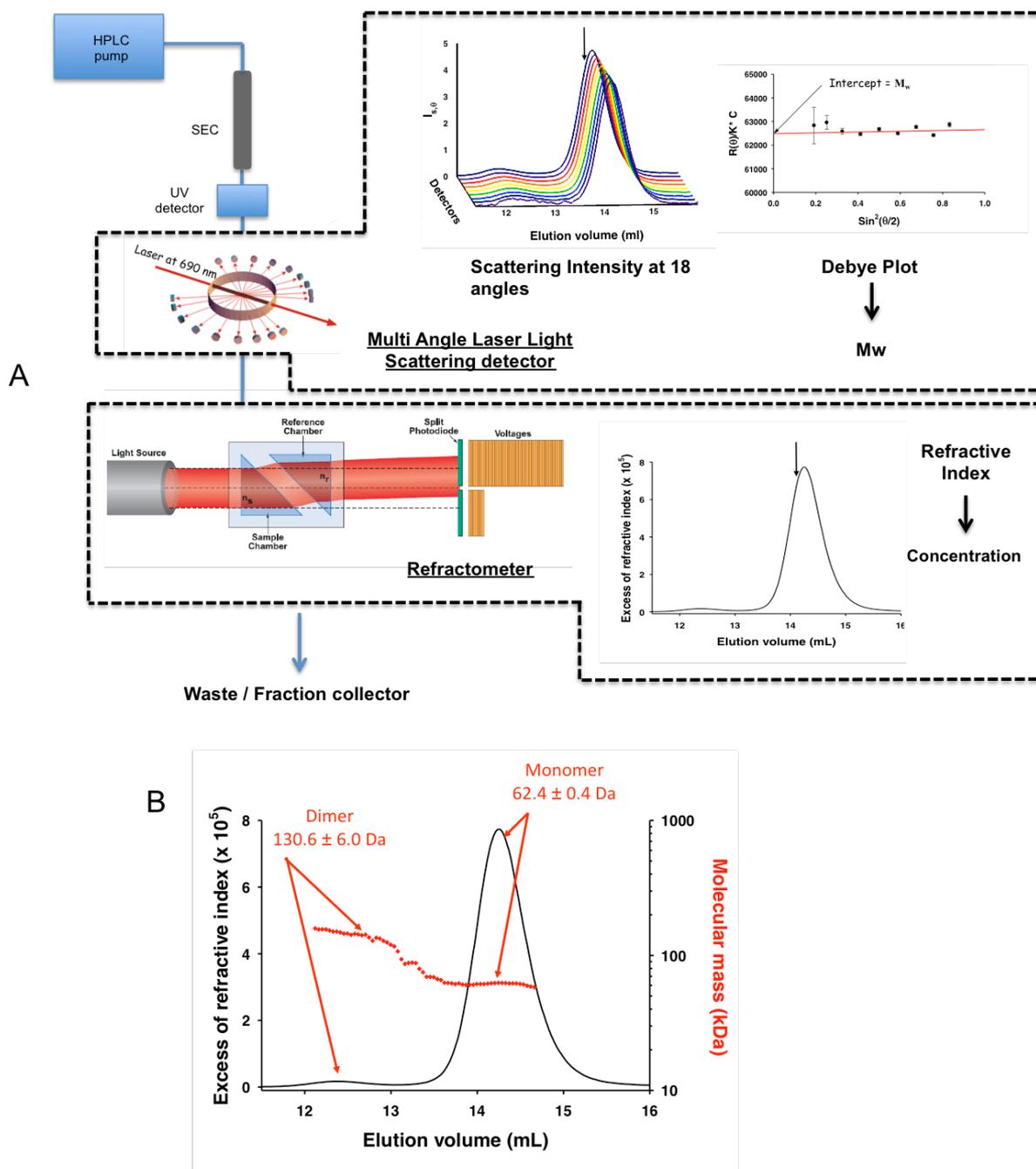


Figure 22: Schematic view of the SEC-MALLS-RI system and data analysis. (A) Measurements performed both by the MALLS detector and refractometer with their individual measurement output. (B) Combined data produced by a SEC-MALLS-RI run using pure serum albumin. The black curve corresponds to the excess of refractive index and the red spots correspond to molecular estimates at a given elution volume. (Images adapted from Marc Jamin, UVHCl Grenoble)

2.8 SMALL ANGLE X-RAY SCATTERING (SAXS)

2.8.1 SAXS PRINCIPLE

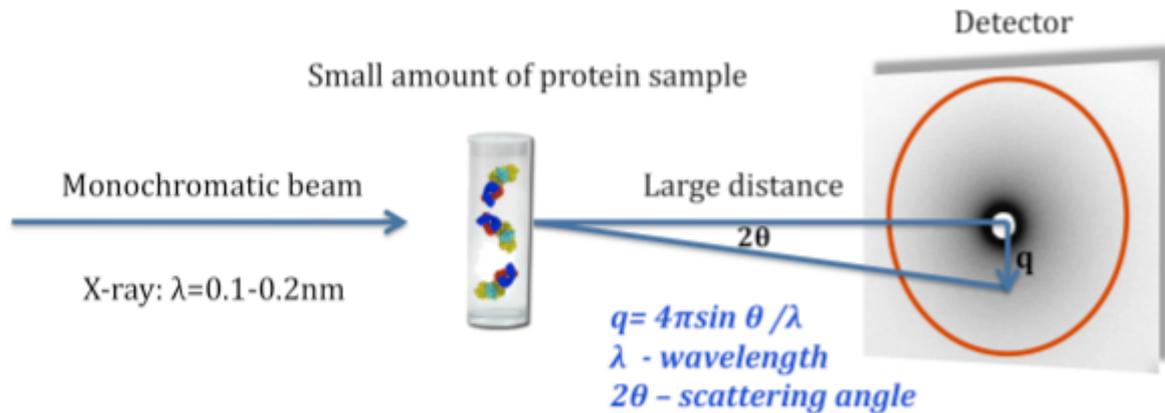


Figure 23: SAXS schematic principle and experimental setup. (Adapted from <http://www.cgl.ucsf.edu/Research/CandS/sali/>)

SAXS is a biophysical method which relies on the elastic scattering of X-rays at small angles by a given protein sample. Unlike X-ray crystallography, this method does not require the sample to crystallise, which makes it easy to apply on any purified monodisperse protein. However, the lack of a periodic organisation of the protein generates a signal averaging of all orientations of the protein. Therefore this method enables a size estimation of the protein by yielding a Radius of Gyration (**R_g**) and a maximal interatomic distance (**D_{max}**). Also this method enables a shape approximation by providing a pair-distribution function **P(r)**, which translates as the sum of all interatomic distances. In turn, the P(r) and experimental curve can generate low-resolution models (resolution between 20-100 Å) through computational modelling.

2.8.2 DATA AQUISITION

In practice, SAXS consists in the exposure of a given protein solution to an intense monochromatic beam of X-rays (λ within the range of 0.1 nm) and the measurement of scattered intensities at small angles ($0.1^\circ \leq 2\theta \leq 5^\circ$) which requires an important detector distance from the scattering sample (Figure 23). In order to limit X-ray air absorption, the entire distance between the sample and the detector is within a vacuum tube.

First integrated by the detector as an image, computational analysis generates a one dimensional SAXS curve $\ln I(q)$ plotting the averaged intensity with its calculated error as a function of q which is the distance from the incident beam centre and directly correlates with the 2θ through the formula:

$$q = 4\pi \sin\theta / \lambda$$

where θ is in degrees and λ in nm or Å. SAXS curves obtained for macromolecule samples at a given concentration must then be subtracted with the buffer $\ln I(q)$ curve in order to assess exclusively the contribution of the protein to the SAXS curve (Figure 24).

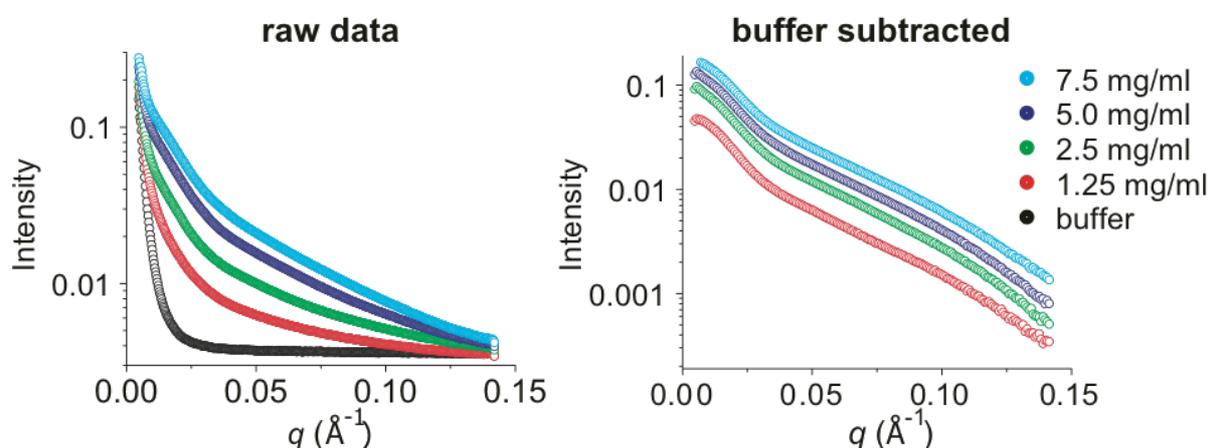


Figure 24: Example of a buffer subtraction on $I(q)$ raw data curves for four different concentrations (1,25 to 7,5 mg/ml). (Adapted from Putman et al., 2007)

2.8.3 DATA ANALYSIS

When analysing the data, the first performed analyses is the Guinier transform as it can indicate the presence of potential aggregation which would make any further analysis irrelevant. The Guinier plot, named after André Guinier (1911-2000), plots $\ln(I)$ as a function of q^2 at very small angles (Guinier region). In the presence of a monodisperse sample the curve is linear and remains so for all concentrations. However, when a protein has a tendency towards self-attraction or aggregation, a loss of linearity is observed and increases with concentration (Figure 25). The same loss of linearity can also be observed with proteins that strongly repulse each other. Data that does not respect a relative linearity within the Guinier region cannot be further interpreted. In the case where the Guinier region is linear, a curve can be fitted with the following equation:

$$\ln[I(q)] = \ln[I(0)] - \frac{q^2 R_g^2}{3}$$

From this equation, two important constants can be deduced: $I(\mathbf{0})$, which is the theoretical intensity at the origin where $q^2=0$ and depends both on protein mass and volume; R_g (nm or Å), which defines the quadratic mean of distances to the centre of mass.

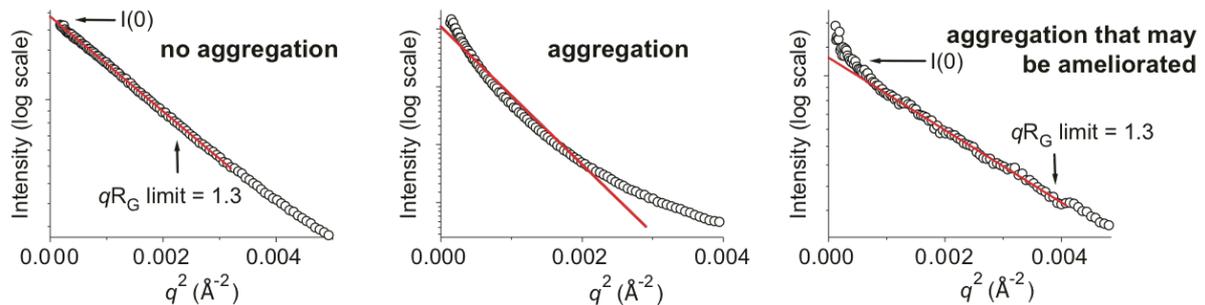


Figure 25: Guinier transform for a monodisperse protein (Left), strongly aggregated (centre) and partially aggregated (Right) protein. (Adapted from Putman et al., 2007)

The second analysis step is the $P(r)$ function that describes all the inter-atomic distances within the protein through the Fourier transform of the experimental curve $\ln I(q)$. This mathematical function is analogous to the Patterson function used in X-ray crystallography with the exception that the $P(r)$ is radially averaged and lacks vectors corresponding to the distance between the scattering particles. When using the GNOM program (Semenyuk and Svergun, 1991), indirect Fourier transforms of $P(r)$ functions will be performed iteratively in order to fit the original experimental curve. The produced curve is bell-shaped and $P(r)$ is equal to 0 at $r = 0$ and $r = D_{max}$. D_{max} given in nm or Å represents the largest inter-atomic distance. The shape of the curve can already indicate the presence of separate domains and describes the overall shape of the protein (Figure 26). The R_g can also be deduced from the $P(r)$ plot as it corresponds to the average of all distances over the whole range and with the R_g predicted with the Guinier approximation. This R_g calculation takes into account the entire experimental curve. Consistency in the calculated R_g between the Guinier plot and the pair-distribution function should always be observed.

From the predicted $P(r)$ and experimental curve, several ab initio envelope prediction programs such as SASHA, DAMMIN and GASBOR (Svergun et al., 1996; Svergun, 1999; Svergun et al., 2001) can compute envelopes. DAMMIN (Svergun, 1999) is the program exclusively used in this thesis project to compute polymerase envelopes. It uses iterative simulated annealing of a randomly generated bead model in order to generate an envelope consistent with the experimental data. Computational constraints such as connectivity, smoothness or internal symmetry can be tuned in order to reduce the solution space. This process is repeated several

times in order to generate an ensemble of models, which can then be aligned and averaged using the DAMAVER (Volkov and Svergun, 2003) program. For further biological interpretation, SAXS can be compared to pdb structures using the Crystol algorithm (Svergun et al., 1995) which produces a theoretical SAXS diffusion curve ($\ln I(q)$) for a given pdb model. Further reading of SAXS principles and applications can be provided by the following extensive and complete review: Putman et al. (2007).

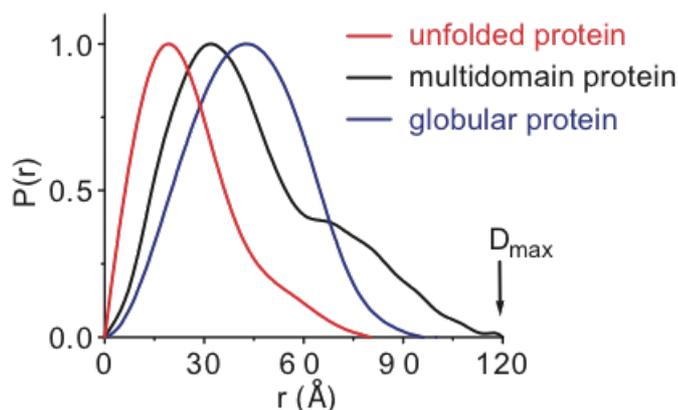


Figure 26: $P(r)$ function representations for different types of protein. (Adapted from Putman et al., 2007)

2.8.4 EXPERIMENTAL PROCEDURES FOR SAXS

All datasets were collected on BM29 at the ESRF (Grenoble). The initial trials using direct sample measurement showed a slight concentration dependency of $I(0)$ for polymerase constructs, which is indicative of partial oligomerisation. Online SEC-SAXS was therefore favoured for all other experiments in order to limit the contribution of aggregated, oligomeric or dissociated protein to the experimental curve. The experimental setup therefore consists of a High Pressure Liquid Chromatography (HPLC) system that injects the protein sample on a small S200 column (S200 5/150 GL, GE healthcare) and monitors its elution by UV absorbance (280 and 260 nm) before injecting it on the SAXS measurement capillary. SAXS measurements were performed every second with a Pilatus 1M detector at a distance of 2,87 m allowing a q range of 0,03 to 4,5 nm with a wavelength of 0.01 nm.

Following the data collection, experimental curves were subtracted and analysed using Primus (ATSAS) or Scatter programs packages. To estimate the molecular mass, the V_c correlation approach was used (Rambo and Tainer, 2013). R_g predictions using the Guinier extrapolation were plotted against the elution volume to select the most monodisperse part of the protein elution peak. SAXS

datasets within this zone were then scaled and averaged to produce one unique $\ln I(q)$ curve. After which GNOM was used to produce a $P(r)$ function that was subsequently exploited by DAMMIF to generate 10 to 20 ab initio models. DAMAVER (Volkov and Svergun, 2003) was then used to generate an average model. DAMMIN was used to assess consistency of the averaged models with the original experimental curve. Model representation was done using chimera (Pettersen et al., 2004) and curve representations using Graphpad (Prism).

2.9 X-RAY CRYSTALLOGRAPHY

2.9.1 PRINCIPLE

As SAXS, X-ray crystallography is a method that uses the elastic dispersion of X-rays by the electrons contained within the atomic structure of proteins. Differently from SAXS though, this method maintains orientational information, enabling the reconstruction of an electron density map, which in turn enables the building of a structure model at atomic resolution.

To reach this final result, three major limiting steps can be distinguished. First is crystallogenesi s, which is the formation of protein crystals and depends mostly on the biochemistry of a given protein (purity, stability, inherent flexibility) in order to be successful. Second is the data acquisition/integration step that is limited by the diffracting quality/intensity of the crystals. Third is solving the phase problem, which is done either by molecular replacement using homologous proteins and/or by experimental phasing. Final success is only ensured through the validation of these three stages. Any small flaws in the early stages can greatly impact the latter stages.

2.9.2 CRYSTALLOGENESIS

Crystallogenesi s is the artificially induced transition of a protein from a soluble state into a solid and highly organised state. By definition, a protein crystal is grown through a periodic arrangement of the same unit cell or “building block” into superposed lattices eventually forming a macroscopic structure.

In the early days of crystallography, crystal growth was considered more of an “art” than hard science due to the multiple parameters influencing its outcome and the sometimes empirical nature of the phenomenon. Even today, with a better understanding of physiochemical parameters controlling this process and the access

to high-throughput technologies, this process can be very difficult to attain and reproduce. Whatever the target protein, four steps are required for crystallogenesis: pre-nucleation, nucleation, growth, and cessation of growth that can all be visualised within the phase diagram (Figure 27).

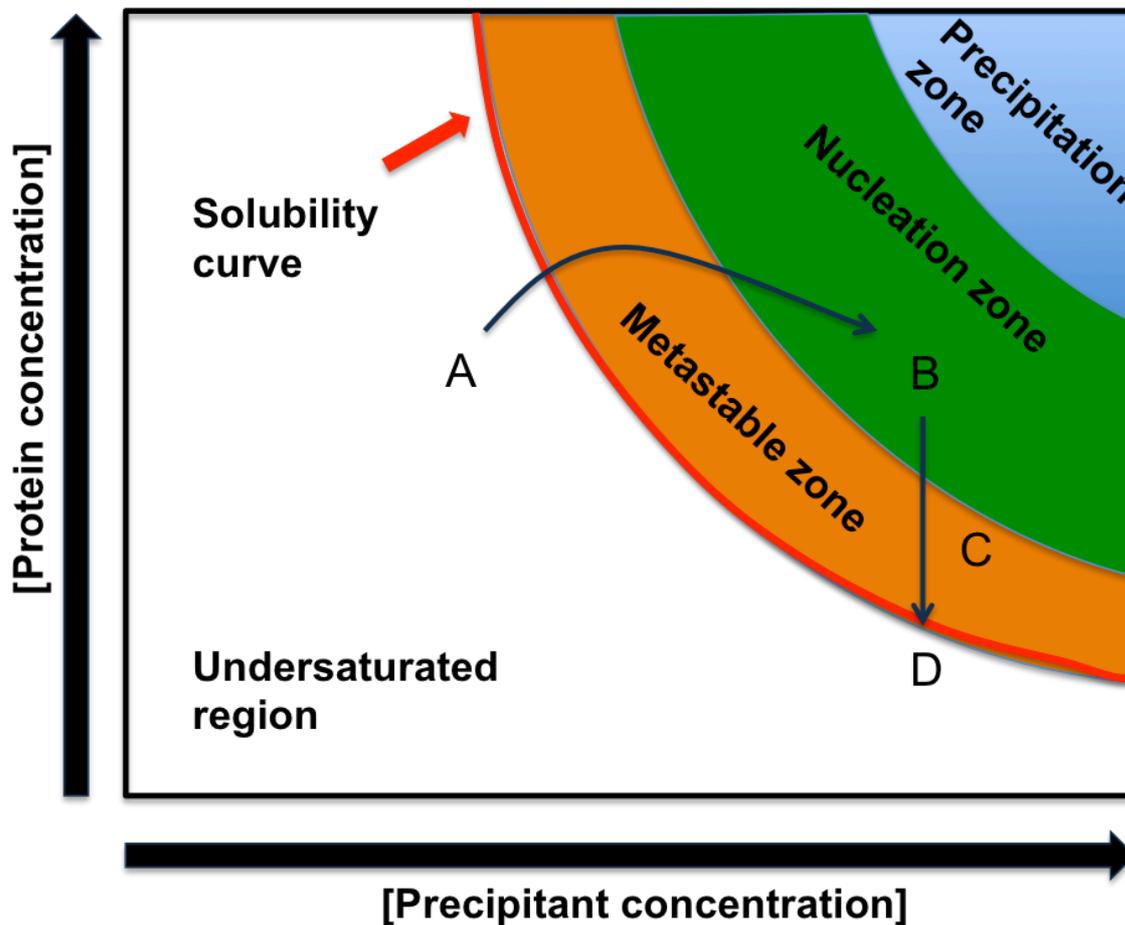


Figure 27: Crystallogenesis phase diagram representing the four stages of crystal growth. (A) Generation of a supersaturated state that can lead to crystal nucleation (B) followed by crystal growth (C) within the metastable zone. Growth arrest (D) is reached when the system reaches equilibrium at the solubility curve which separates the soluble region from the supersaturated state.

To crystallise proteins, several methods can be used to induce these phase transitions: vapour diffusion techniques (sitting or hanging drop) and dialysis methods. The first method is the most commonly used and easy to set up in high throughput; it is also the only crystallisation method used within this thesis project. Its principle is to set up a closed system containing a precipitant well, whose volume far exceeds that of the protein sample. Within that same closed environment, the protein sample is mixed with some precipitant to form a drop and is either sitting on an adjacent surface or hanging from the top of the well. Over time, vapour diffusion will equalise the precipitant concentrations between the precipitant well and the protein/precipitant drop, which in effect concentrates both protein and precipitant within the drop. This has the effect of pushing the protein into a supersaturated state,

which can either lead to nucleation or precipitation. Commonly used precipitants include salts and organic precipitants such as polyethylenglycol (PEG), 2-Methyl-2,4-pentanediol (MPD) or isopropanol.

Temperature, pH and initial additives within the protein mix also play an important role in the potential outcome. Crystal growth can take from 1 day to several weeks depending on the protein. Once an initial crystallisation condition is identified, fine-tuning of these different parameters is undertaken in order to produce the biggest and nicest looking crystals possible. Once grown, crystals are fished out of their precipitant liquor using a cryo-loop and flash frozen in liquid nitrogen for storage before x-ray exposure and analysis. Appropriate cryo-protecting solutions are usually used to guarantee crystal integrity upon freezing.

2.9.3 ACQUISITION AND INTEGRATION

When matter is exposed to an incident beam of X-rays, the electrons constituting this matter absorb in part those X-rays and become excited. In order to return to their fundamental energy state those electrons re-emit X-rays in all directions. Within a crystal, proteins are periodically organised in lattice planes and has the effect of only allowing certain diffraction angles to produce constructive interference that produces diffraction spots. This property is controlled by Braggs law (Figure 28):

$$2d \sin\theta = n \lambda$$

where d is the distance between two reticular planes, λ the wavelength of the incoming beam, n is an integer or whole number and θ the angle between the incoming beam and the lattice plane.

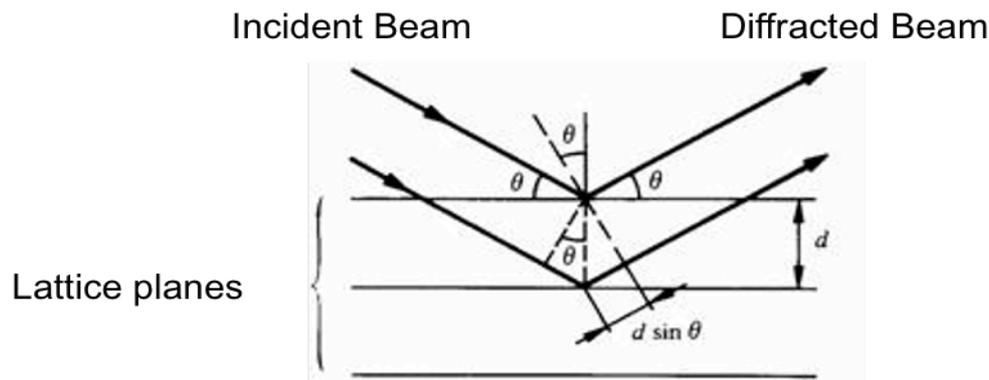


Figure 28: Bragg's Law represented with two incident beams of X-rays (black arrow), constructive interference is obtained only when the path difference $d \sin \theta$ is a multiple of the beam wavelength

It is important to note that in the case of a crystal, diffraction spots or “reflexions” are the product of constructive interference from all the electrons within the crystal lattice. In other words, each reflection contains information on all the atoms in the structure and each atom contributes to the intensity of each reflection.

On a diffraction pattern, the reflexion spots correspond to coordinates within reciprocal space of the crystal lattice. One of the first stages of computer analysis during data collection is indexing. This step can be performed on two or more images ideally separated by a 90° angle and produces an estimation of the unit cell dimensions which in turn gives a prediction of crystal symmetry or space group. Space group determination then enables a collection strategy to be developed accordingly in order to limit X-ray exposure and therefore radiation damage. Indeed, a higher order of symmetry space group (P4, 90° oscillation) will need a shorter oscillation angle than a lower symmetry space group (P2, 180° oscillation). Once established, a full data collection can then be undertaken in order to measure all reflexions in reciprocal space, usually aiming for a completeness $>98\%$. Modern beamlines will collect very thin sliced oscillations (0.2°) resulting in the production of more than 1000 images for a full data collection. All these images are then re-indexed with refined unit cell parameters that attribute a relative coordinate to all the recorded spots (Miller Indices, h, k, l). Following indexing is integration where all the images are converted into a single data file containing the h, k, l indexes of a spot and its measured intensity and error. Finally the last stages of data preparation include scaling and merging steps that combine the data and correct eventual discrepancies during the data collection (loop absorbance, beamline parameter variations and radiation damage).

2.9.4 PHASE DETERMINATION

In order to determine an electronic density map of the scattering lattice or structure factors one must combine both the amplitude and phase of all the reflections. Amplitudes can be derived from the measured intensity of the reflexions. However, phases cannot be directly measured and are lost during the process. Several methods can be used to find the phases.

The first, **molecular replacement** (MR), is purely computational and relies on the use of previously determined structures (Rossmann and Blow, 1962). In practice, molecular replacement programs (Phaser or Molrep) compare Patterson maps of the experimental data with predicted maps from the search model (homologous protein or domain). A Patterson map is an interatomic vector map generated by squaring all the structure factor amplitudes and setting the phases to zero. The Patterson map is self-centred so the MR search is divided into two stages, a rotational search (self-rotation function) and a translational search. In a successful MR search, both stages will yield preferably one or few solutions. The correct orientations can then be used by the MR program to predict a phasing solution using the homologous model. Factors controlling the success rate of MR include structural homology (40% at least) between the search model and actual structure within the crystal. MR can also yield a solution when trying to solve the structure of a large complex and searching with a known structure of one of the domains constituting a minimum of 30% of the complex. Finally, the quality and resolution of the collected dataset plays an important role in the potentially successful outcome.

The other method is through **experimental phasing**, which houses a large number of different sub-techniques; these all involve adding some other element within the native crystal. The first historically used method is Single Isomorphous Replacement (SIR), which requires soaking of the crystals in heavy atom solutions, the consequence of which does not modify the unit cell parameters. Diffraction spot amplitudes at the same wavelength of both native and soaked crystals will differ greatly (15-25 % difference) due to the contributions of the heavy atoms to the overall diffractions. Computer calculations can then predict the positions of the heavy atoms within the unit cell using Patterson functions; this in turn can provide a phase solution for all the other atoms. Using one isomorphous replacement can still produce a solution ambiguity, as two phase solutions are possible per structure factor. Combining derivative datasets using Multiple Isomorphous Replacement (MIR) can solve this eventual solution ambiguity.

Nowadays however, the most commonly used method for experimental phasing is to use Single Anomalous Dispersion (SAD). This method is analogous in

some aspects to SIR as it requires the introduction of another component within an otherwise native crystal (Hendrickson et al., 1990). Selenomethionine is commonly used for this purpose and replaces methionine during expression of the protein, which can sometimes be a problem with unstable or difficult to express proteins. Also, in order for the experiment to be successful, a sufficient number of methionines are required within the protein amino acid sequence, at minimum 1.5% of the total number of residues.

During diffraction and data processing, SAD relies on differences in Friedel Pair diffraction spots or anomalous dispersion that is maximised at a specific wavelength for selenium. Anomalous signal can be quantified through computer calculations and attributed to selenium atoms within the crystal structure. Like SIR, SAD can produce ambiguous phasing solutions, which can be reduced using Multiple Anomalous Dispersion (MAD), which takes two datasets: one maximising the anomalous dispersion like SAD and another one with poor anomalous dispersion signal. Due to the very weak amplitude differences in anomalous signal (1-5 %), strong data completeness is required.

Direct experimental phasing methods can also be combined together in order to produce a more accurate phasing model. Such is the case with Single/Multiple Isomorphous Replacement with Anomalous Dispersion (SIRAS, MIRAS). Finally, combination of molecular replacement and experimental phasing methods can be used to reduce phase ambiguity.

Once a phase solution is found, different programs can build a “raw” atomic model of the given protein taking into account the peptide sequence in order to explain the electronic density within the unit cell. Following this first build, manual tweaking and other programs can refine the model and also correct experimental data measurements in order to improve data to model statistics. Two major correlation factors are mainly used to evaluate model to data reciprocity: First, the R-factor that calculates how the calculated amplitudes differ from the measured amplitudes at the previous refinement stage. Second is the free R-factors. In order to evaluate the final model, the free R-factor is used to compare amplitude differences between the refined amplitudes and 5% of the initial data that has not been submitted to refinement. In all, these statistics must remain close and improve together during the refinement process.

Further reading for experimental principles of x-ray crystallography can be found in various reviews (Giegé and Sauter, 2010; Ilari and Savino, 2008; Wlodawer et al., 2013).

2.9.5 EXPERIMENTAL PROCEDURES FOR X-RAY CRYSTALLOGRAPHY:

RanBP5 crystals were prepared using a protein mother liquor at 6-8 mg/ml within the size exclusion buffer (HEPES pH: 7.5, 300 mM NaCl, 5 mM β). Sitting drops were set up on intelliplates (96 wells, Art Robbins Instruments) by the High Throughput Crystallography facility (HTX-EMBL, Grenoble) using 100 nL of sample and 100 nL of precipitant. After initial hits identified with the Hampton and Qiagen screening kits, refined crystallisation was obtained with a precipitant composed of 12% PEG 6000, 100 mM Sodium tri-Sodium Citrate pH: 4.5, 100 mM Ammonium Acetate. RanBP5 crystals would usually appear in 24h and grow for the following two days. Following growth, crystals were cryo-protected by adding 1 μ L 40% PEG 3350, 100 mM Sodium tri-Sodium Citrate pH: 4.5, 100 mM Ammonium Acetate to the screen drops. Crystals were then mounted on cryo loops (Hampton research) and frozen in liquid nitrogen.

Diffraction experiments were undertaken on ID30.A, ID 29 and ID 23.1 at the European Synchrotron Radiation Facility (ESRF). Data was collected using a Marr225 or Dectris Pillatus 6M detector. Indexing and integration was performed using the XDS program package (Kabsch, 1993). Molecular replacement attempts on native datasets was undertaken using Molrep (Vagin and Teplyakov, 1997) and Phaser (McCoy et al., 2007) using the ccp4 (Winn et al., 2011). Derivative datasets, once integrated and scaled, were treated using SHARP.

2.10 TRANSMISSION ELECTRON MICROSCOPY USING NEGATIVE STAINING

2.10.1 PRINCIPLE

Transmission Electron microscopy (TEM) is a biophysical method which uses strongly focused high voltage electron beams instead of light to probe matter. By doing so, the magnification power is considerably increased compared to a light based microscope by a factor 50000x. However, both visible light and electron microscopes share common magnifying features. As such, a TEM microscope consists in a big vacuum tube which has a source of electrons on the top. The

electrons are produced by a tungsten cathode or electron gun and accelerated by an anode which has a voltage potential with regards to the cathode, usually at around 100 KeV when studying protein samples. Downstream from the source, a first series of electromagnetic/electrostatic lenses focus the electron beam on the sample. Once transmitted through the sample, a second series of lenses magnify the beam on a detector (phosphorescent screen or camera) enabling a direct computer integration of the images. On a homogeneous carbon surface, the presence of protein particles will modify through absorption and scattering the incoming X-rays. However, due to their weak atomic mass composition (carbon, hydrogen, nitrogen, oxygen), the signal contrast remains very weak.

In order to solve this problem, negative staining methods are used to study small biological particles. In practice heavy metal salts such as Uranium Acetate (UrAc) or Sodium Silico Tungstate (SST) are used to increase the contrast by covering the entire surface of the grid but surrounding the protein particles. This produces a negative stain as the background absorbs more electrons, therefore appearing darker than the protein particles which appear brighter on the images. It is however important to note that EM observes artefacts of complexes and particles as the staining process can modify the sample stability and homogeneity.

TEM, like X-ray crystallography and NMR can also be used as a tool for structural determination notably with the use of single particle analysis combined with cryo-EM (Zhou, 2008). Recently it has been increasingly used to produce high-resolution structures (4-20 Å range). However, like all other methods, EM reconstruction requires a number of conditions to be met. First of all, the size of the target protein; as single particle analysis requires the alignment and class averaging of numerous particles (within the order of 10^5 for a near atomic resolution). Viruses and big macromolecular complexes with a high degree of internal symmetry constitute the best targets for this process but in recent years several smaller particle structures have been solved with resolutions around 4 Å (Bai et al., 2015; Chang et al., 2015; Du et al., 2015). The sample must also show a high degree of purity and stability especially when using cryo-EM.

For further reading on the principle of negative staining EM can be found in various reviews (Boekema et al., 2009; Frank, 1996; Zhou, 2008)

2.10.2 DATA COLLECTION

For all polymerase and/or RanBP5 samples, negative staining TEM was used to assess the homogeneity of the sample. Samples (0,1 mg/ml) were analysed at the

Institut de Biologie Structurale (IBS) through a collaboration with the EM platform (Guy Schoen). The carbon flotation technique was used to prepare negative staining grids with UrAc or SST. Images were collected on a FEI Technai 12 LaB6 equipped with a Gatan Orius 1000 camera.

2.11 AUTOMATED MICRO BIO-REACTOR (AMBR)

2.11.1 BRIEF OVERVIEW

AMBR (Tap Biosystems) is an automated workstation capable of performing 24 to 48 micro-scale fermenter experiments simultaneously. Placed under a sterile hood (Figure 29A), the platform uses plastic single-use, 15 ml bioreactors (Figure 29B). For every reactor, AMBR can follow in real time both pH and cellular density. An automated robotic arm can also perform several functions such as sample retrieval (for cell counting and viability testing) and media supplementation. O₂, N₂ and CO₂ delivery per micro-scale reactor is individually controlled for optimal growth tuning. Homogenisation of the cultures is synchronised for all bioreactors and performed by the integrated plastic propeller within each bio-reactor.



Figure 29: AMBR platform overview. (A) Picture of the AMBR platform within the insect cell facility of EMBL. (B) Single use, plastic bio-reactor containing mammalian cell media (Adapted from Tap Biosystems website).

2.11.2 AUTOMATED INSECT CULTURES

In the first culture station, the 12 bioreactors were seeded with *Sf21* (*Spodoptera frugiperda*) insect cells in SF 900 II SFM medium (Life Technologies). 1 ml to 1.3 ml V0 baculovirus of each of the 11 constructs was added in the 11 first bioreactors respectively. The last bioreactor was not infected and served as a control. For all 12 reactors, the cellular density was of $0.6 \cdot 10^6$ viable cells/ml for a final volume of 13 ml. In the second culture station, the 12 bioreactors were seeded with *Hi5* (*Trichopulsia ni*) insect cell lines also in SF 900 II SFM medium. In order to keep the same MOI as for the SF21 infections, 1.4 to 1.6 ml Vo baculovirus of each of the 11 constructs was added in 11 bioreactors respectively. The last bioreactor was not infected and served as a control. For all 12 reactors, the cellular density was of $0.4 \cdot 10^6$ viable cells/ml with 13 ml of final volume. In the following days, liquid handling arm with sterile tips were used to sample out different volumes from the bioreactors. Samples were analysed for cell counts, viability and cell diameter with the Vi-cell XR (Beckman Coulter), for pH calibration using a micro pH probe (Crison Basic 20+ pH-meter) and for YFP/CFP fluorescence using the Tecan Infinite 200 Pro (Tecan).

Bioreactors containing Hi5 insect cells were harvested 96 h after infection. Bioreactors containing Sf21 insect cells were harvested later when cell viability was around 70%. After a 10 min, 1000 g centrifugation step at room temperature, the culture supernatant was discarded, cell pellets were flash frozen in liquid nitrogen and stored at -80°C .

2.11.3 SMALL-SCALE PURIFICATION

Upon reaching the CFP/YFP plateau phase, AMBR cultures were spun down and re-suspended in 1.6 ml of lysis buffer (50 mM Tris pH: 8.5, 300 mM NaCl, 2 mM BME, 1 complete tab/50 ml). Two steps of freeze/thaw in liquid nitrogen were performed for lysis. Soluble fractionation was then performed with a 1 h centrifugation at 16000 g. Upon extraction of the supernatant and supplementation with 40mM imidazole, the nickel resin pull-down was undertaken using 100 μL of resin (Qiagen) with a 15 min incubation at 10°C . Three 1 ml wash steps were then performed including a high salt wash (1 M NaCl in the lysis buffer). Two 200 μL elutions were then performed by increasing the imidazole content to 300 mM. After elution, the two eluted fractions were concentrated to 50 μl using an 30 kDa cut-off concentrator (Amicon) and injected on a 3 ml S200 (PC 3.2/30, GE Healthcare) connected to the micro-äakta running at 0.05 ml/min. Peak fractions were collected at 75 μL intervals within a 96 well Greiner plate.

CHAPTER 3: RNA BINDING AND ASSEMBLY OF HUMAN INFLUENZA A VIRUS POLYMERASES

Résumé du chapitre 3 en Français

Dans le but de comprendre les activités de transcription et de réplication de l'ARN polymérase du virus de la grippe, des études fonctionnelles ont été entreprises depuis plus de quarante ans. Il aura fallu attendre le développement de l'expression de polyprotéines recombinantes en cellules d'insecte pour obtenir des formes solubles de polymérases entières. Cette méthode a permis l'obtention de structures cristallographiques de polymérases d'influenza A issue d'une souche infectant les chauves souris et d'influenza B. Or, au moment de l'écriture de ce manuscrit, aucune structure de polymérase entière de souche de grippe A aviaire ou humaine n'a été publiée. Notre groupe a depuis la mise en place de cette méthode, recherché à exprimer des polymérases de souches aviaires (H5N1) et humaines (H3N2) mais s'est heurté à un problème d'expression de ces deux souches. Ce manuscrit d'article présente tout d'abord la méthode d'expression par polyprotéine de fusion qui a permis de produire et d'assembler de façon stoechiométrique PA, PB1 et PB2. Le rôle délétère de PB2 dans l'expression globale de la polyprotéine de fusion a pu être mis en évidence par le suivi des rapporteurs d'expression YFP/CFP et par western blot. L'obtention de polymérases tronquées ou complexées à RanBP5 a ensuite été détaillée par analyses SDS PAGE, SEC-MALLS et par microscopie électronique en coloration négative. La mesure d'affinité par anisotropie de fluorescence a permis de mesurer des affinités extrêmement fortes de la polymérase envers le promoteur 5' ARNv. Une mesure plus précise par rétention sur filtre d'ARN marqué au P₃₂ a confirmé ce résultat, mesurant un K_D de l'ordre du picomolaire. Le rôle crucial de PB2 dans la liaison de la polymérase au promoteur 3' ARN a été démontré par l'absence d'affinité de PA-PB1 pour cet ARNv à la différence de PA-PB1-PB2(1-116) qui le lie avec une affinité de l'ordre du nanomolaire.

Ce manuscrit présente des résultats tout à fait novateurs dans la caractérisation par SAXS du complexe PA-PB1-RanBP5 et son absence totale d'affinité pour l'ARNv 5'. Ces résultats démontrent un rôle crucial de RanBP5 dans le processus d'assemblage de la polymérase qui reste encore mal compris. Il y est proposé modèle d'assemblage séquentiel de la polymérase où PA-PB1 est maintenu inactif par RanBP5 jusqu'à la phase d'assemblage finale avec PB2 où la présence du promoteur 3' ARN peut jouer un rôle important. Cette découverte est importante, car elle ouvre de nouvelles voies d'inhibition du virus de la grippe notamment par le ciblage de l'interaction PA-PB1 et de RanBP5.

NOTE OF INTRODUCTION TO CHAPTER 3

The following chapter is the manuscript of an article covering most of the results obtained with truncated RdRp constructs. In this manuscript, we investigate the factors leading to a loss of expression of the polymerase in our polyprotein strategy. We also demonstrate the expression and high-purity purification of IAV RdRps as well as RdRp-complexes. Functional studies focus mostly on the vRNA promoter binding and are coupled with MALLS characterisations and SAXS envelope determination. Altogether, the combined data of different RdRps and RanBP5 constructs further advance the current models of polymerase assembly by notably describing for the first time, the PA-PB1-RanBP5 complex in solution. It will be submitted at the beginning of October 2015. The manuscript is inserted directly and has its own page numbering. The following chapters continue after at page 93.

CHAPTER 3 MANUSCRIPT

Structural characterization of recombinant IAV polymerase reveals a stable complex between viral PA-PB1 heterodimer and host RanBP5

Christopher SWALE^{1,2§}, Alexandre MONOD^{1,2§}, Laura TENGO^{1,2}, Alice LABARONNE^{1,2}, Frédéric GARZONI^{1,2,3}, Jean-Marie BOURHIS^{1,2}, Stephen CUSACK^{1,2,3}, Guy SCHOEHN^{1,2,4,5,6}, Imre BERGER^{1,2,3,7}, Rob WH RUIGROK^{1,2} & Thibaut CREPIN^{1,2*}

¹ Université Grenoble Alpes, Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France

² CNRS, Unit of Virus Host Cell Interactions, UMI 3265 UJF-EMBL-CNRS, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France

³ EMBL Grenoble, 71 avenue des Martyrs, CS 90181, F-38042 Grenoble Cedex 9, France

⁴ CNRS - Institut de Biologie Structurale, UMR 5075, 71 avenue des Martyrs, CS 10090, F-38044 Grenoble Cedex 9, France

⁵ CEA - Institut de Biologie Structurale, UMR 5075, 71 avenue des Martyrs, CS 10090, F-38044 Grenoble Cedex 9, France

⁶ Université Grenoble Alpes, Institut de Biologie Structurale, UMR 5075, 71 avenue des Martyrs, CS 10090, 38044 Grenoble Cedex 9, France

⁷ The School of Biochemistry, University of Bristol, Clifton BS8 1TD, United Kingdom

*** Corresponding author**

E-mail: crepin@uvhci.fr

§ These authors contributed equally to this work

Short title: RNA binding and assembly of human IAV polymerase

Keywords: Influenza virus, RNA polymerase, vRNA specificity, nuclear import, assembly

Abstract

The genome of influenza A virus (IAV) comprises eight RNA segments (vRNA) which are transcribed and replicated by the heterotrimeric IAV RNA-dependent RNA-polymerase (RdRp). RdRp consists of three subunits (PA, PB1 and PB2) and the heterotrimeric complex binds both the highly conserved 3'- and 5'-ends of the vRNA segment. The IAV RdRp is an important antiviral target, but its structural mechanism has remained largely elusive to date. By applying a polyprotein strategy, we produced RdRp complexes and define a minimal human IAV RdRp core complex. We show that PA-PB1 forms a stable heterodimeric submodule that can strongly interact with 5'-vRNA. In contrast, 3'-vRNA recognition critically depends on the PB2 N-terminal domain. Moreover, we demonstrate that PA-PB1 forms a stable and stoichiometric complex with host nuclear import factor RanBP5 that can be modelling by SAXS and we show that the PA-PB1-RanBP5 complex is no longer capable of 5'-vRNA binding. Our results provide further evidence for a step-wise assembly of IAV structural components from preformed submodules, regulated by nuclear transport mechanisms and host factor binding.

Author contributions

Experimental conception: SC, IB, RWHR, TC. Biochemistry: CS, AM, LT, FG, AL, TC. EM: AL, GS. SAXS: CS, JMB, TC, X-ray crystallography: CS, AM, SC, TC. Data analysis: CS, AM, AL, RWHR, TC. Paper writing: CS, AM, SC, GS, JMB, IB, RWHR, TC.

Introduction

The influenza virus is a negative sense RNA virus with a segmented genome belonging to the *Orthomyxoviridae* family. The viral RNA (vRNA) is divided into 8 segments that encode for a total of 10 core proteins and a few additional accessory proteins. Every vRNA segment is coated by numerous nucleoprotein (NP) units and by one RNA-dependent RNA-polymerase (RdRp) complex composed of PA (716 aa for influenza A virus (IAV) strains), PB1 (756 aa for IAV strains) and PB2 (757aa for IAV strains). The RdRp binds to the 5'- and 3'- terminal extremities of the vRNA which acts as a promoter region [1-3]. The macromolecular assembly between the vRNA, NP and the RdRp forms the ribonucleoprotein (RNP). The RdRp performs both replication and transcription of the vRNA genome in the nucleus of the cell. As such, the RdRp can produce either cRNA/vRNA through *de novo* replication or viral mRNA through "cap snatching" transcription [4]. The factors regulating the switch from a transcriptive to a replicative RdRp remain to be characterised.

Recently, the structures of influenza B [5] and bat influenza A [6] were published, providing tremendous insight into the complex architecture of the RdRp. All three subunits are tightly intertwined and form intricate quaternary structures at the vRNA promoter binding sites. Current models of RdRp assembly suggest that PB2 is imported via the importin- α pathway [7,8] whereas the PA-PB1 heterodimer is imported through direct interaction with RanBP5 [9-11].

Structural studies of RdRp from the human infecting influenza A (human-IAV) strains in contrast, have been limited to date, partly due to difficulties to produce recombinant H3N2 or H5N1 polymerases. Nonetheless, a cryo-EM reconstruction was recently obtained of a truncated influenza A/H5N1 polymerase [12]. In the present article, we identify PB2 as the main bottleneck hampering complete recombinant polymerase expression in our insect cells expression system. Moreover, we present the

1 characterisation of several constructs of human-IAV RdRp, including the biologically relevant PA-PB1
2 heterodimer in solution. The titration measurements against the 5'- and 3'-vRNA promoters show a
3 strong sub-nanomolar affinity of the PA-PB1 heterodimer towards the 5'-vRNA whereas the specific
4 binding of the 3'-vRNA requires the presence of PB2. By applying our co-expression strategy, we show
5 that the previously proposed PA-PB1-RanBP5 import complex can be purified, which we characterise
6 using small angle X-ray scattering (SAXS). Biochemical analysis of this complex reveals a role for
7 RanBP5 in hindering 5'-vRNA binding. Taken together, these data provide evidence for a cellular RdRp
8 assembly pathway following a sequential and conditional process of assembly.

9

Results

PB2 is a key limiting factor for recombinant expression of the heterotrimer

Polyproteins are naturally used by certain viruses to structure their proteome [13,14]. Recombinant polyproteins mimicking viral polyproteins have recently emerged as a powerful tool to express protein complexes for functional characterisation as well as structural determination (reviewed in [15]). Using this strategy [16], soluble and active heterotrimeric RNA-dependent RNA-polymerases of influenza B virus (IBV) and bat influenza A (bat-IAV) virus could be produced, crystallized and the structure determined at high resolution [5,6]. We have applied a similar strategy to the RdRp of two human-IAV strains, A/Victoria/3/1975(H3N2) and the highly pathogenic A/Viet-Nam/1203/2004(H5N1). The three genes of each heterotrimeric complex were combined with Tobacco Etch Virus (TEV) protease and Cyan Fluorescent Protein (CFP) in a single large open reading frame (ORF). Each gene was separated by a DNA sequence encoding for a peptide segment comprising a short serine/glycine linker and a TEV protease cleavage site (Figure 1.A). The constructs were optimized for recombinant expression in insect cells using the MultiBac system [17,18]. During expression, the TEV protease cleaves the polyprotein co-translationally, resulting in a stoichiometric assembly of the RdRp that can then be selectively purified using an affinity nickel resin. The production of the polyprotein is monitored using the fluorescence of two reporter proteins: CFP encoded within the polyprotein reports directly on recombinant protein yield, while YFP, integrated in the baculovirus genome, monitors virus performance. The ratio of YFP/CFP in our system is a highly useful criterion to determine recombinant polyprotein translation efficacy of the different constructs.

Table 1 summarizes the results obtained with all constructs tested. Several attempts were made to express the H3N2 polymerase heterotrimer (construct number 12, Table 1). Although YFP and CFP can

1 be detected and quantified, CFP values were extremely low compared to YFP with a YFP/CFP ratio of
2 50, the highest observed for any polyprotein constructs that we expressed. No clear bands were
3 observed on western blot using specific PA, PB1 and PB2 antibodies. Similar results were obtained
4 with the H5N1 RdRp polyprotein, indicating failure of these recombinant expression experiments for
5 these full-length IAV heterotrimeric polymerases.

6 In order to understand the difference between the expression of the human-IAV and IBV
7 polymerases, we designed new constructs with a C-terminally truncated PB2 based on available
8 structural information. These were PB2(1-36) [19], PB2(1-116) [20] and based on the cap-binding
9 domain [21], PB2(322-483). Expression and purification experiments were undertaken on constructs
10 15, 17 and 18, all of which all have full-length PA, PB1 and truncations of PB2 extending to residues
11 116, 320 and 483 respectively (Figure 1.B). As the PB2 extension increases beyond residue 116, we
12 observed a loss of expression of the polyprotein, thus identifying a critical region in PB2 that limits
13 yield. This can first of all be observed through the monitoring of YFP/CFP reporter genes during
14 expression (Figure 1.C). YFP fluorescence values reach a similar plateau for all constructs whereas CFP
15 values are reduced significantly for constructs 17 and 18. Moreover, after nickel resin pull-down on
16 lysates from an equal amount of cells, we observed by western blot (Figure 1.D) decreased intensity
17 bands of not only PB2 but also PB1, indicative of an overall expression loss. Polyprotein size increase
18 alone could not explain the reduced expression because other control constructs (19 and 20) with
19 additional protein sequences beyond PB2 residue 116 conserve lower YFP/CFP ratios. Furthermore,
20 replacing the PB2(1-116) by another large protein such as RanBP5 (residues 1-1115) generates a
21 construct with a lower YFP/CFP ratio than construct 15. The analysis of the amino acids sequences of
22 PB2 does not provide more indication on the real nature of this phenomenon (supplementary Figure
23 1).

1 To determine whether this “loss-of-expression” phenomenon results from our polyprotein expression
2 strategy, which combines PA-PB1 and PB2 in one ORF, we generated a MultiBac baculovirus co-
3 expressing a PA-PB1 fusion and PB2 from an independent expression cassette [22]. This strategy also
4 resulted in high YFP/CFP ratios (≥ 35), even though the CFP gene was only fused to PA-PB1. These
5 results identify PB2, especially when extended beyond the residue 116, as the bottleneck for
6 expression of complete human-IAV RdRp.

7 **Dissecting the heterotrimer**

8 In order to identify a minimal active core of the human-IAV polymerase, we have systematically
9 dissected the heterotrimeric complex for further characterisation using the polyprotein strategy. The
10 expression of PA-PB1 increases significantly when PB2 is totally removed (compare constructs 1 and
11 13/15; Table 1). The same observation is made after removing the endonuclease domain (i.e. PA-Nter;
12 compare constructs 14 and 13/15). After purification, PA-PB1 forms homogeneous monomeric and
13 stable particles (Figure 2.B) whereas the complexes with PB2 (constructs 15 and 16) give rise to
14 dimers in solution (Figure 2.A). The dimers are stable enough to withstand salt concentrations up to 2
15 M NaCl during purification, but the SEC elution peaks are broad suggesting that the oligomerization
16 process is quite dynamic. By adding vRNA promoter-like molecules (e.g. the IAV panhandle; [23]), we
17 found that the dimers dissociated into monomeric RNA:protein complexes. We were able to show
18 that the truncated PA-PB1 (construct 1) and the PA-PB1-PB2 (construct 14) exhibit expected
19 polymerase activities. The endonuclease activity is similar to that of the isolated PA-Nter domain
20 (supplementary Figure 2), can be inhibited by the same point mutation (i.e. PA-E80A) or compound
21 (i.e. DPBA) and shows comparable dependency on manganese [24,25]. The constructs are also
22 functional in RNA synthesis. Upon addition of a mixture of IAV panhandle, NTPs and [α - 32 P]-UTP, both

1 truncated PA-PB1 and PA-PB1-PB2 are able to generate an 80-nucleotide long product plus other
2 minor products in absence or in presence of ApG (supplementary Figure 3).

3 We have also shown that further N-terminally truncating PA and/or C-terminally truncating PB1 has
4 no impact on the expression of the corresponding polyproteins (constructs 1 to 11). The heterodimer
5 PA-PB1 can be purified with or without the N-terminal domain of PA (i.e. PA-Nter; compare constructs
6 1 and 2). However, as soon as the PA-hinge (residues 200-260) is removed, PB1 becomes insoluble.
7 This implies that the PA-hinge linking the Nter- and Cter-domains of PA is crucial for the stability of
8 PB1. However, it is not sufficient in itself since the C-terminal region of PA interacting with the N-
9 terminal part of PB1 [26,27], cannot be deleted either (compare constructs 2 and 5 to 11). PB1 can be
10 shortened on its C-terminus but when purifying the construct 4 (i.e. PB1 until amino acid 560), only
11 PA-Cter remains soluble. Thus, the minimal construct required to obtain soluble PB1 is construct 3
12 with PA from 197 to 716 and PB1 from 1 to 660. All the PA-PB1 constructs without the N-terminal
13 domain of PA form dimers in solution (supplementary Figure 4) and adding vRNA-like molecules has
14 no incidence on the dimerization.

15 Over all the constructs tested, crystals have only been obtained using construct 2 (PA from 197 to 716
16 and PB1 from 1 to 686). Crystals grew in a few days in low PEG content solutions. They diffract poorly,
17 with diffraction limited to typically ≈ 10 Å. Adding vRNA-like molecules had no effect on the
18 diffraction. One dataset has been collected at 7.95 Å (space group $P6_12$ with cells parameters $a = b =$
19 349.2 Å, $c = 166.4$ Å). A molecular replacement solution has been found using bat-IAV PA-PB1
20 structure [6] confirming that there is a dimer of two molecules of PA(197-716)-PB1(1-686) in the
21 asymmetric unit (solvent content = 88 %). Because of the poor resolution, no model improvement
22 was possible but there were no obvious significant differences compared to the input molecular
23 replacement search model (supplementary Figure 5).

1 **PA-PB1 dimer exhibits high-affinity interaction with 5'-vRNA but requires PB2 for binding 3'-vRNA**

2 All influenza RNA segments have the same organization, a central coding region flanked by 2 un-
3 translated regions containing the highly conserved and complementary 5'- and 3'-ends [28-30]. The
4 viral polymerase specifically interacts with both the 5'- and 3'-ends, as recently visualized in crystal
5 structures [5,6], and uses them as a promoter [1-3]. During their biochemical characterisation, we
6 have seen significant effects of vRNA-like RNA molecules on the stability and/or the oligomeric state
7 of most of the constructs. The initial RNA molecule used was the 80-nt panhandle [23] from which we
8 made the 5'-vRNAp corresponding to the 5'-end (5'-AGUAGAAACAAGGGUA) and its 3' equivalent
9 called 3'-vRNAp (AUACCCUGCUUUUGCU-3'). By thermal shift assay experiments, a 10°C stabilisation
10 was observed when 5'-vRNAp was added to all the constructs, whereas when 3'-vRNAp was added,
11 the effect was less significant and dependent on the construct. Similar data had already been
12 published [31].

13 We used fluorescence anisotropy to measure the interaction between our truncated polymerase
14 constructs and the conserved vRNA ends by following the fluorescence polarization increase of a
15 fluorescently labelled RNA when it binds the polymerase. For this purpose, we used 5'-vRNAp and 3'-
16 vRNAp, both labelled with fluorescein amidite (FAM) at the opposite extremity of the putative
17 interaction, i.e. at the 3'-end for the 5'-vRNAp and *vice versa* (Figure 3 and Table 2). At 300 mM NaCl,
18 the Kd for the 5'-vRNAp are similar (sub-nanomolar) for both constructs 1 and 14, whereas for the 3'-
19 vRNAp, a Kd of 36 nM for the construct with PB2 is obtained (Figure 3.A) and no binding was observed
20 for the construct without PB2 (Figure 3.C). As the estimated Kd for the 5'-end was in a range below
21 the working concentration and beyond the sensitivity limit of the fluorescence anisotropy
22 measurement, another method was required to determine the 5'-vRNAp affinity. Filter binding assays
23 (FBA) were then therefore used for precise Kd determination (Figure 3.B). With the corresponding

1 radioactive probes, we found values of 0.2-0.4 nM for the 5'-end. For the 3'-end, the K_d was 100-fold
2 higher and critically dependent on the presence of PB2. Salt concentration is also an important
3 parameter. We show different binding characteristics at 150 and 300 mM NaCl of the PA-PB1 towards
4 different RNAs. The K_d value of PA-PB1 for the 5'-end remains sub-nanomolar in the range of 150 to
5 300 mM NaCl. In contrast, at 150 mM NaCl, 3'-vRNAP binds to PA-PB1 with the same affinity than for
6 a poly-UC (Figure 3.C), and so is non-specific. All relevant titrations were performed at 300 mM NaCl
7 to prevent non-specific affinity measurements. Recently published affinity data [32] working at 500
8 mM NaCl likewise consistent with very high affinity of the 5'-vRNAP towards IAV polymerase but this
9 study failed to measure 3'-vRNAP binding within the nanomolar range, indicating that 3'-vRNAP
10 binding is also tightly influenced by salt concentration.

11 **RanBP5 interacts tightly with PA-PB1(1-686) when co-expressed**

12 During the viral cycle, the assembly of influenza RdRp follows a multi-factorial pathway involving
13 many host partners proteins. After transcription in the nucleus, viral mRNAs are exported to the
14 cytoplasm to be translated by the cellular machinery. The components of the replication machinery
15 (i.e. PA, PB1, PB2 and NP) then must be imported into the nucleus. Whereas PB2 and NP use
16 importins- α [8,33-36], PA and PB1 are conjointly imported as a preformed heterodimeric submodule
17 by the importin- β RanBP5 [9,11,37]. Having confirmed the existence of a stable PA-PB1 complex we
18 sought to demonstrate its interaction with RanPB5. The human *IPO5* gene which encodes for RanBP5,
19 was expressed in insect cells. After purification, RanBP5 forms a homogeneous monomeric sample
20 (Figure 2.C) that has been used in mixing experiments with freshly purified PA-PB1(1-686) followed by
21 SEC-MALLS-RI experiments. The two samples were mixed in stoichiometric ratio and incubated before
22 size exclusion chromatography (SEC). Both RanBP5 and PA-PB1(1-686) were eluted in the same single
23 peak, but MALLS-RI indicated a molecular weight corresponding of a mixture rather than a stable

1 ternary complex (supplementary Figure 6). We then attempted to produce a PA-PB1(1-686)-RanBP5
2 heterotrimer by the self-processing polyprotein strategy. A long ORF encoding for the trimeric
3 complex was created by inserting the *IPO5* gene downstream of the PB1(1-186) coding sequence
4 (Table 1, construct 21). Remarkably, from this polyprotein construct, PA-PB1 co-purifies with RanBP5,
5 forming a homogeneous and stable heterotrimer (Figure 2.D).

6 Online SEC-SAXS was used for the characterisation of PA-PB1(1-686), RanBP5 and PA-PB1(1-686)-
7 RanBP5, respectively in solution. Using the Vc determination method [38] on the diffusion data of PA-
8 PB1(1-686) shows a Mw estimate of 146 kDa (Table 3 and supplementary Figure 7). The calculated
9 Mw is 166 kDa. The Guinier transform measures a hydrodynamic radius (Rg) of 37.9 Å and GNOM
10 produces the pair distribution function fit with a Dmax of 128 Å. Further analysis of the SAXS curve
11 (Figure 4.A) shows a visually adequate correlation with the CRY SOL curve which uses the bat-IAV
12 polymerase structure (PDB id: 4WSB) as a model. The high χ^2 values are observed for this dataset,
13 which are probably the result of a noisy SAXS curve. A slight deviation of the fit is observed above a q
14 range of 0.8 nm⁻¹, suggesting conformational differences between the scattering curve and the crystal
15 structure coordinates. *Ab-initio* modelling was performed using 15 DAMMIF models which were
16 averaged by DAMAVER. Averaged model correlation with the diffusion curve was then checked using
17 the damstart file as a starting envelope for DAMMIN. The polymerase DAMAVER envelope (Figure
18 4.A) appears as a pear shaped structure in which the homologous model can be fitted. Additional
19 envelope volume is visible close to the endonuclease domain, implying that it adopts multiple
20 conformations in solution.

21 Using the same methodology, we determine RanBP5 to be a monomer in solution with a measured
22 Mw of 144 kDa close to the theoretical Mw of 126 kDa (Table 3 and supplementary Figure 7). RanBP5
23 displays an Rg of 38.8 Å and a Dmax of 136 Å. With a Mw lower than that of PA-PB1(1-686), RanBP5

1 displays a larger Dmax and Rg indicating an extended structure. CRY SOL curve fitting using the closest
2 sequence homologue Kap121 (PDB id: 3W3T) against the diffusion curve reveals important
3 conformational differences between the two structures (Figure 4.B). Envelope modelling confirms an
4 elongated structure of RanBP5 in solution.

5 Analysis of PA-PB1(1-686)-RanBP5 describes a significantly larger complex in solution with an Rg of
6 51.8 Å and a Dmax of 181 Å. The Mw estimate of 323 KDa is close to the expected Mw of 292 KDa,
7 confirming the presence of a stoichiometric 1:1:1 PA-PB1(1-686)-RanBP5 complex in solution. MONSA
8 was used to perform the *ab-initio* modelling of the complex PA-PB1(1-686)-RanBP5. It separately uses
9 the PA-PB1(1-686) and RanBP5 diffusion curves (Figure 4.A and 4.B) in combination with that of PA-
10 PB1(1-686)-RanBP5 to propose a consensus dual envelope (Figure 4.C). Model fitting was performed
11 against the three experimental curves and matches all 3 curves with similar quality. Homologous
12 structures were manually fitted in order to compare envelope and crystal structure sizes. They do not
13 constitute a valid docking solution and again only serve as dimension references. With this *ab-initio*
14 modelling we also find a comparable envelope for PA-PB1(1-686) like the one produced with
15 DAMMIF/DAMAVAR. RanBP5 on the other hand displays a different envelope suggesting
16 conformational differences between the unbound and bound forms of RanBP5. In this envelope, the
17 Kap121 coordinates can no longer be totally fitted underscoring the conformational difference
18 between the envelope and the homologous X-ray structure. The model suggests a close interaction
19 between the two proteins, RanBP5 (violet envelope) tightly interacts well over half of its length with
20 PA-PB1(1-686) (grey envelope), suggesting molecular contacts well beyond those of the proposed PB1
21 NLS containing domain, which the crystal structure shows is a mobile and solvent exposed β-ribbon
22 [6]. This model further supports the hypothesis that RanBP5 may play a chaperone function for PA-
23 PB1 prior to assembly with PB2 in the nucleus to complete the polymerase.

1 **RanBP5 regulates the binding of the 5'-vRNAP**

2 The PA-PB1 complex not only exists as a stable submodule but can moreover recognise the 5'-vRNAP
3 with a high affinity and act as a replicase and endonuclease on its own, without PB2 present. The
4 current model predicts the NLS of PA-PB1 to be present within PB1 (residues 187-211) [11], at the end
5 of a beta-strand located parallel to the vRNAP binding sites, indicating a potential link between the
6 binding to RanBP5 and/or the binding to vRNAP.

7 Subsequently to measuring the sub-nanomolar K_d of PA-PB1(1-686) and PA-PB1-PB2(1-116) to the 5'-
8 vRNAP, an identical experiment was performed with the PA-PB1(1-686)-RanBP5 complex (Figure 3.D).
9 The association curve clearly shows a complete loss of binding affinity towards the 5'-vRNAP when
10 RanBP5 is present, both at 300 mM and 150 mM NaCl. Using SEC-MALLS-RI experiments, we have
11 shown that an excess of 5'-vRNAP does not induce any complex dissociation of the PA-PB1(1-686)-
12 RanBP5 trimer (supplementary Figure 8). Taken together, these experiments suggest that RanBP5
13 binding obscures the 5'-vRNAP binding site, thus blocking specific 5'-vRNAP binding.

14

Discussion

Expression of full-length IBV and bat-IAV polymerases has been successfully achieved using a polyprotein approach, resulting in atomic resolution structures by X-ray crystallography [5,6]. In marked contrast and unexpectedly, efficient expression of active full-length human or avian IAV polymerase at yields high enough for structural studies has not been reported to date, irrespective of the method used. Our studies identify PB2, and notably residues 120 to 250 as responsible for this poor expression. Interestingly, attempts by a different research group to reconstitute a full IAV polymerase (e.g. full IAV polymerase expression/purification for structural studies) likewise failed beyond the N-terminal PB1 interacting domain of PB2 [12]. The reason why PB2 hinders polyprotein expression is not clear at the moment. We speculate that the extended fold of the PB2 region 1-250 is unstable unless bound to one side of PB1 [5,6] and thus may require chaperones to maintain its stability prior to assembly onto PA-PB1. Of note, Hsp90 has been shown to stimulate influenza virus RNA synthesis [39] and the nuclear import of the subunits [34]. Moreover, Hsp90 inhibitors reduce influenza virus replication in cell culture [40]. We have shown previously that provision of co-factors in *trans* can ameliorate the quality and quantity of target proteins expressed in the baculovirus/insect cell system [41,42]. Provision of Hsp90 and the other chaperones, within the polyprotein or in *trans* on the baculovirus backbone, may provide a powerful handle to overcome the bottleneck of human-AIV polymerase production.

PA-PB1 can be produced as a stable submodule and forms a discrete heterodimer in solution. Dimerization of PA-PB1 heterodimers can appear when either the endonuclease is removed or the N-terminal of PB2 is added which may indicate that the superstructure formation could be artificially induced by truncations of the full-length polymerase. PA-PB1 can specifically bind to the 5'-vRNAP with sub-nanomolar affinity in solution. In contrast, 3'-vRNAP binding requires the presence of PB2

1 and is within the low nanomolar range, indicating a much stronger preference of the polymerase
2 towards the 5'-vRNAP. These results are consistent with the high-resolution crystal structures in
3 which the 5'-vRNAP binding site is located between the PA and PB1 subunits [5,6]. The 3'-vRNA, on
4 the other hand, interacts with all three subunits suggesting a sequential vRNAP binding mechanism.

5 Our results corroborate key steps of influenza polymerase assembly [9,37,43]. Purification of
6 stoichiometric complexes using the polyprotein strategy confirm that PA-PB1 can exist as a preformed
7 and stable submodule, independently of additional host factors, with PA acting as a putative
8 "chaperone" for PB1. Our data also shows that PA-PB1 is functional in RNA synthesis, which has
9 interesting physiological implications. Moreover, we demonstrate here that the PA-PB1(1-686)
10 heterodimer can assemble with RanBP5 into a stable, heterotrimeric complex that can be purified *in*
11 *vitro*. This likely corresponds to the import complex for these two subunits, although other
12 chaperones such as HSP90 may be co-imported [34]. In contrast to PA-PB1 with RanBP5, we failed to
13 reconstitute a binary PB2-importin- α complex using the polyprotein strategy even though functional
14 and structural data show that PB2 nuclear import depends on the importin- α pathway, through a
15 direct interaction between the PB2 C-terminal domain that contains the NLS [7,8]. This is probably
16 partly due to an inherent instability of isolated IAV PB2 in insect cells but could also reflect the lack of
17 sufficient chaperones, co-chaperones or other unidentified factors that stabilise PB2. Furthermore,
18 cytoplasmic expression of PB2 appears to destabilise many cellular processes [44,45]. The lack of high
19 expression levels of fully assembled PA-PB1-PB2 heterotrimers unfortunately currently limits our
20 efforts to elucidate the structure of the complete IAV polymerase at atomic resolution. Nevertheless,
21 docking of bat-IAV and IBV crystal structures into the PA-PB1(1-686)-RanBP5 SAXS envelope suggests
22 that assembly of PB2 onto the PA-PB1-RanBP5 complex might be possible, based on steric
23 considerations. Indeed RanBP5 could even facilitate trimer assembly perhaps concomitantly with Ran-

1 dependent dissociation from PA-PB1. In contrast, RanBP5 profoundly affects the ability of PA-PB1 to
2 bind vRNA. We show that whereas the 5'-vRNA is specifically recognized by PA-PB1 with high affinity,
3 this is abolished in the ternary complex with RanBP5. This again suggests that RanBP5 disassociation
4 may be coupled to assembly of the polymerase with the promoter RNA. In any case, only when PB2
5 associates is the polymerase able to associate with the 3'-vRNA end, consistent with the crystal
6 structure [5,6]. But surprisingly, the affinity of the heterotrimer for 3'-vRNA is in the same order of
7 magnitude than the affinity of NP for RNA substrates (Table 2), revealing a potential competition
8 between the complete polymerase heterotrimer and the nucleoprotein.

9 We demonstrate here for the first time that the polyprotein expression strategy can be used not only
10 to co-express subunits that form a stable complex, but also to successfully reconstitute interactions
11 with downstream partner proteins *in vivo* in the expression host for direct purification and functional
12 characterisation. Functional studies on IAV RdRp reveal that vRNAp promoter binding is completely
13 abolished when PA-PB1(1-686) is bound to RanBP5, suggesting a possible role for RanBP5 in
14 maintaining PA-PB1 in an inactive form during nuclear import. This discovery holds a promise for
15 influenza research as it provides a novel point of therapeutic intervention for targeting and inhibiting
16 influenza virus during assembly of its active components. We anticipate that compounds which block
17 polymerase/RanBP5 binding or release may prove to be highly efficient, broad spectrum antiviral
18 inhibitors in the treatment of influenza, in particular for patient groups where classical vaccination
19 strategies fail.

20 **Methods**

21 **Molecular biology**

22 The DNA coding sequences of A/Victoria/3/1975(H3N2), the highly pathogenic A/Viet-
23 Nam/1203/2004(H5N1) polymerase subunits and human *IPO5* have been ordered to GeneArt

1 (ThermoFisher Scientific), optimized for the expression in insect cells. Cloning has been achieved
2 following the supplier procedures (New England Biolabs). The pPBAC plasmid was used for the
3 polyprotein constructs [15,17]. Few constructs have been expressed using pFastBac-HTB (Life
4 Technologies). vRNA-like molecules (i.e. IAV panhandle; [23]) have been produced and purified using
5 classical *in vitro* transcription protocols [46,47].

6 **Expression and purification**

7 Large scale suspension cultures expressing polymerase fusion constructs were prepared using High
8 Five insect cells grown in Express Five media (Life Technologies) at 0.5×10^6 cells/mL infected at 0.2%
9 (V/V) with the baculovirus mother solution. Cultures were maintained at $0.5-1 \times 10^6$ cells/mL until
10 proliferation arrest (24-48h after infection). Following the proliferation arrest YFP and CFP
11 measurements were performed every 12h until a fluorescence plateau was reached (72-96h after
12 infection). Cultures were then spun down at 800 g for 10 min and cell pellets were stored at -80°C .
13 Cell pellets were resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 8.5, 300 mM NaCl and 2mM
14 β -mercaptoethanol) per 500×10^6 cells in the presence of EDTA-free anti-protease cocktail (complete
15 from Roche). Lysis was performed with two cycles of freezing (-180°C) / thawing (26°C) after which
16 10% of glycerol were added to the lysate before centrifugation (45 min, 40 000 g, 4°C). After retrieval
17 of the clarified lysate, 30 mM of imidazole pH 8.0 were added before loading on Ni-NTa superpose
18 resin (Quiagen). After flowing the lysate through the resin, three wash steps with 10 column volumes
19 (CV) of buffer A (lysis buffer with 30 mM imidazole), 10 CV of wash buffer B (50 mM Tris-HCl pH 8.5, 1
20 M NaCl, 10% glycerol and 2 mM β -mercaptoethanol) and 10 CV of wash buffer A were performed.
21 Elution of the bound complex was performed with 300 mM imidazole. Elution fractions containing
22 polymerase were then pooled together and directly injected on a 5 mL Hitrap heparin resin (GE
23 healthcare) which had previously been equilibrated with 5 CV of buffer A. After binding to the resin, a

1 5 CV wash was performed with buffer A before eluting with a 40 mL salt gradient on a FPLC system.
2 The elution peak corresponding to stoichiometric polymerase assemblies was then pooled and
3 injected on an S200 (GE healthcare) size exclusion chromatography equilibrated in 50 mM Tris-HCl pH
4 8.5, 300 mM NaCl, 5 mM β -mercaptoethanol. Peak fractions were pooled and concentrated to the
5 desired concentration using a 100 kDa concentrator. Once concentrated the protein prep could be
6 stored by addition of 20% glycerol and flash freezing in liquid nitrogen.
7 NP was purified using the protocol described in [48,49].

8 **SEC-MALLS-RI analysis**

9 All MALLS runs were performed using a S200 increase SEC column (10/300 GL, GE Healthcare). Sample
10 injection and buffer flow was controlled by a Hitachi L2130 pump, following the SEC column was a L-
11 2400 UV detector (Hitachi), Optilab T-rEX refractometer (Wyatt technologies) and a DAWN HELEOS-II
12 multi angle light scattering detector (Wyatt technologies). Prior to injection, columns and systems
13 were equilibrated in 5 to 10 column volumes of running buffer. 50 μ L injections were performed using
14 protein samples concentrated at a minimum of 2 mg.mL⁻¹, constant flow rate of 0.5 mL.min⁻¹ was
15 used. Accurate MALLS mass prediction was performed with the Astra software (Wyatt Technologies).
16 Curves were represented with Graphpad (Prism).

17

18 **Electron microscopy**

19 Samples (0.1 mg.mL⁻¹) were applied to the clean side of carbon on mica. The carbon was then floated
20 on sodium silicotungstate (2% v/w) or uranyl acetate (2% v/w) and a grid placed on the top of it. After
21 air-drying, the samples were observed in a T12 FEI electron microscope and images were taken using
22 an Orius SC1000 CCD camera (Gtan Inc., Pleasanton, CA).

1 **SAXS analysis**

2 All datasets were collected on BM29 (ESRF). The initial trials using direct sample measurement
3 showed a slight concentration dependency of the Guinier estimated R_g . Online SEC-SAXS was
4 therefore used for all experiments in order to limit the contribution of partially aggregated, oligomeric
5 or dissociated protein to the experimental curve. The experimental setup consists of a High Pressure
6 Liquid Chromatography (HPLC) system connected to an analytical S200 increase column (5/150 GL, GE
7 Healthcare) followed downstream by the SAXS sample capillary. SAXS measurements were performed
8 every second with a Pilatus 1M detector at distance of 2.87 m allowing a q range of 0.03 to 4.5 nm
9 with a wavelength of 0.01 nm.

10 Following data collection, experimental curves were subtracted and analysed using Primus (ATSAS) or
11 Scatter (Bioisis) programs suits [50]. To verify the molecular mass, the Rambo and Tainer method was
12 used [38,51]. R_g predictions using the Guinier extrapolation were plotted against the elution volume
13 to select the most monodisperse part of the protein elution peak. SAXS datasets within this zone were
14 then scaled and averaged to produce one unique $I(q)$ curve. GNOM [52] was then used to produce a
15 $P(r)$ function which was subsequently used by DAMMIF [53] to generate 15 *ab-initio* models.
16 DAMAVER [54] was then used to generate an average model. Consistency of the DAMAVER averaging
17 with the original experimental curve was assessed by using DAMMIN [55] and using the damstart.pdb
18 envelope as a starting model. Homologue PDB structure comparison was assed using Crysol [56].
19 Multiple diffusion curve *ab-initio* modelling was performed using Bunch [57]. Homologue structure
20 fitting within the DAMAVER envelope was performed with Chimera [58] and curve representations
21 using Graphpad (Prism).

22 **Fluorescence anisotropy assay**

1 Equilibrium RNA binding experiments were performed with the following vRNA set synthesized by
2 Integrated DNA Technologies (IDT): 5'-vRNAp (5'-pAGUAGAAACAAGGGUA-FAM3'), 3'-vRNAp (5'-FAM-
3 AUACCCUGCUUUUGCU-3') and polyUC (5'-pUCUCUCUCUCUCUCUCUCUC-FAM3'). 5'-vRNA sequences
4 were synthesized with a 5'-mono phosphate and labeled with 6-FAM in 3'OH. 3' vRNA were labeled
5 with FAM in 5'. Polymerase constructs was titrated into fluorescently labeled RNA using the buffer (50
6 mM HEPES pH 7.5, 5 mM β -mercaptoethanol plus 150 mM or 300 mM NaCl) at room temperature.
7 Initial reactions were performed with 2-4 nM of RNA with an initial volume of 600 μ L to limit the
8 dilution effect of protein addition (<5%). Anisotropy measurements were undertaken using a PTI
9 fluorometer equipped with automated polarizers. The excitation and emission wavelengths were 494
10 and 521 nm, respectively. 50 anisotropy measurements with a 1 second integration time were
11 performed per titration point.

12 **Filter Binding Assay**

13 Equilibrium RNA binding experiments were undertaken with the same vRNA set synthesized by IDT.
14 vRNA were labeled with 32 P in 5' using the T4 polynucleotide kinase (New England Biolabs). Double
15 filter binding was used following the previously established protocol [59]. Protein concentration was
16 titrated against constant concentrations of vRNA ($\leq 0,01$ nM) within a final volume of 200 μ L using the
17 standard protein buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM β -mercaptoethanol). 180 μ L were
18 then filtered on two membranes using a 96 well Whatman Minifold dot-blot apparatus. The first
19 membrane was Protran BA 85 membrane (Whatman) to retain the polymerase and the second one is
20 a nylon Hybond-N+ membrane (Amersham Bioscience) to retain unbound nucleic acid. Both
21 membranes were pre-incubated 1h at room temperature within the protein buffer prior to use. After
22 blotting, another 180 μ L of protein buffer were run through the membranes to wash any unbound
23 RNA before letting the vacuum dry the dot-blot for 30 seconds. Both membranes were then

1 dismantled from the dot-blot apparatus and exposed to a storage phosphor screen (BAS storage
2 phosphor screen, GE Healthcare) overnight. Revelation of the phosphor screen was then performed
3 with a Typhoon Trio imaging system (GE Healthcare). Dot blot phosphorescence intensity was
4 integrated using the Image J software. A derived bound ratio was then calculated using the following
5 formula:

$$Br = \frac{I_{bound}}{I_{bound} + I_{free}}$$

7 Where I_{bound} is the integrated intensity of protein/RNA complex which is retained by the protran
8 membrane and I_{free} is the unbound RNA retained by the nylon Hybond-N+ membrane.

9 **Binding assay data analysis**

10 RNA binding curves were plotted with the subtracted anisotropy or Bound RNA ratio as a function of
11 the protein concentration. Sigmoidal binding curves were fitted to the data using GraphPad (Prism)
12 with the following two equations:

13 1) For the anisotropy titrations, $[vRNA] \approx K_d$. Assuming a 1/1 stoichiometry the following equation was
14 used to estimate the K_d

$$15 \quad \Delta A = \left\{ \frac{(R_T + E_T + K_d) - [(R_T + E_T + K_d)^2 - 4 \times R_T \times E_T]^{1/2}}{2 \times R_T} \right\} \times \Delta A_T$$

16 Where ΔA is the change in subtracted anisotropy, ΔA_T the total change of subtracted anisotropy. R_T is
17 the total vRNA concentration, E_T is the total polymerase concentration at each given point and K_d is
18 the dissociation constant.

19 2) In the case of filter binding assay, a single site isotherm was used as we consider that $[vRNA] \ll K_d$.
20 This equation therefor assumes an effective equality between the free polymerase concentration and
21 the total polymerase concentration.

1

$$\Delta Br = \left\{ \frac{\Delta Br_T (E_T / K_d)}{(1 + E_T / K_d)} \right\}$$

2

Where ΔBr is the change in binding ratio, ΔBr_T is the total change in binding ratio. E_T is the total

3

polymerase concentration at every point in titration and K_d is the dissociation constant.

4

Standard error calculations for the apparent K_d within a given condition were calculated through

5

triplicate data measurement.

6

7

8

Acknowledgements

We are very grateful to Daphna Fenel, Francine Gérard, Raphaëlle Germi, Laurence Grossi, Darren J Hart, Marc Jamin, Thomas Lunardi, H  l  ne Malet, Stefan Reich and Adam Round and for their help. We thank Kausiki Datta, Klaus Klumpp and Christoph Bieniossek for discussion. We thank Juan Ortin for the gift of the antibodies. CS was funded by the Hofmann-La Roche pRED external collaboration program, AM by the R  gion Rh  ne-Alpes (ARC 1 Sant  ) and AL through the Labex GRAL. This work was supported by the French agency for Research through the ANR Flunucleovir (ANR-2010-Blanc-1307-1301), the ANR RNAP-IAV (ANR-14-CE09-0017) and the European Commission projects Flupharm (contract number 259751) and ComplexINC (contract number 279039). This study used the platforms of the Grenoble Instruct centre (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).

Competing interest statement

The authors declare that they have no competing financial interests.

Figures legends

1

2 **Figure 1**

3 **Truncated human-IAV RdRp polyprotein expression.** (A) Logic of the TEV-PA-PB1-PB2-CFP
4 polyprotein expression in a schematic view. During expression, TEV will process the polyprotein
5 ensuring a stoichiometric assembly of PA, PB1 and PB2. YFP and CFP are produced during the process
6 and monitor respectively baculovirus proliferation and polyprotein translation. (B) Truncated RdRp
7 constructs where PB2 is incrementally extended until residues 116, 320 or 483. (C) YFP (left) and CFP
8 (right) fluorescence kinetics measured during Hi-5 insect cells culture of truncated RdRp constructs.
9 YFP ($\lambda_{\text{exi}} = 488 \text{ nm}$, $\lambda_{\text{emi}} = 525 \text{ nm}$) and CFP ($\lambda_{\text{exi}} = 430 \text{ nm}$, $\lambda_{\text{emi}} = 480 \text{ nm}$) measurements were
10 performed on cellular extracts prepared by sonicating 1×10^6 cells in PBS (500 μL) follow by
11 centrifugation. Fluorescence intensities are plotted against time after infection. (D) Small scale nickel
12 resin purification analysis by western blot. Purifications were performed on the 50 mL of Hi-5 insect
13 cells cultures used for the YFP and CFP fluorescence kinetics (C). Deposits feature total lysate after
14 freeze/thaw (T), supernatant after centrifugation (S), resin flow through (Ft) and the primary elution
15 fraction (E). After migration on a 12 % SDS-PAGE, proteins were transferred on PVDF membrane.
16 Primary antibodies targeting human-IAV PB2 (rabbit IgG) and human-IAV PB1 (mouse IgG) have been
17 used. Revelation was performed with secondary goat antibodies coupled with Alexa Fluor 532 ($\lambda_{\text{exi}} =$
18 632 nm , $\lambda_{\text{emi}} = 647 \text{ nm}$) and Alexa Fluor 633 ($\lambda_{\text{exi}} = 531 \text{ nm}$, $\lambda_{\text{emi}} = 554 \text{ nm}$) targeting mouse and rabbit
19 H + L domains respectively, using a Typhoon Trio imaging system (GE Healthcare). After integration of
20 the raw data, PB1 and PB2 revelation are visible in red and green respectively. Black and white signal
21 of PB2 is also shown (bottom) to highlight the PB2 truncations. The upper bands appearing in
22 green/yellow, correspond to unprocessed polyproteins.

1 **Figure 2**

2 **Homogeneous sample analysis of recombinant polymerases, RanBP5 and polymerase-RanBP5.**

3 Purified product analysis are horizontally grouped for (A) PA-PB1-PB2(1-116), (B) PA-PB1(1-686), (C)
4 RanBP5 and (D) PA-PB1(1-686)-RanBP5. On the left are the Coomassie blue colored SDS PAGE gel of
5 the purified sample with corresponding subunits bands indicated by colored arrows (PA in blue, PB1
6 in green and RanBP5 in purple). In the middle are the SEC-MALLS chromatograms with the UV signal
7 as a backtrace and molecular weight estimate below the peak. Estimated average molecular weight
8 for each sample is also detailed. SEC-MALLS-RI runs of PA-PB1(1-686), RanBP5 and PA-PB1(1-686)-
9 RanBP5 were performed in the same buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl) whereas the
10 SEC-MALLS-RI run of PA-PB1-PB2(1-116) was performed with 50 mM Tris-HCl pH 8.0, 300 mM NaCl
11 and 10% glycerol. (right) corresponding electron microscopy images.

12 **Figure 3**

13 **vRNA binding and specificity.** (A) Binding titration of the truncated trimer PA-PB1-PB2(1-116) towards
14 the 5'-vRNAp (blue triangle) and 3'-vRNAp (red square) sequences using fluorescence anisotropy at
15 300 mM NaCl. (B) Binding titration performed by filter binding assay against P³² labelled 5'-vRNAp
16 (blue triangle) and 3'-vRNAp (red square) using 300 mM NaCl. Bound RNA fraction is plotted as a
17 function of polymerase concentration. (C) Binding titration of the truncated dimer PA-PB1(1-686)
18 performed at 150 and 300 mM NaCl against the 5'-vRNAp (dark and light blue triangles), 3'-vRNAp
19 (orange and dark red squares) and polyUC RNA (light and dark green circles) by fluorescence
20 anisotropy. (D) Binding titration of different polymerases and polymerase-RanBP5 constructs against
21 the 5'-vRNAp at 300 mM NaCl by fluorescence anisotropy. PA-PB1-PB2(116) and PA-PB1(1-686) are
22 depicted by blue and orange triangles respectively, PA-PB1(1-686)-RanBP5 is depicted with purple

1 triangles. For all anisotropy titrations (**A**, **C** and **D**) subtracted anisotropy is plotted as a function of
2 protein concentration.

3 **Figure 4**

4 **Online SAXS analysis of PA-PB1, RanBP5 and PA-PB1-RanBP5 complex in solution.** Results are shown
5 for (**A**) PA-PB1(1-686), (**B**) RanBP5 and (**C**) PA-PB1-RanBP5. On the left are the experimental $\ln[I(q)]$
6 curves with the *ab-initio* DAMMIN curve fit (blue) using the Damstart (DAMAVER) as an initial
7 constraint and the CRY SOL fit (red) of the closest homologous X-ray structure. The MONSA fit is also
8 shown for the 3 curves in black. The χ^2 of the different curves are also shown. On the right are the
9 homologous PDB structure depicted as cartoon structures fitted into the DAMAVER *ab-initio* envelope
10 for (**A**) and (**B**) and into the MONSA modeling in (**C**). RanBP5 is depicted in purple, PA in blue, PB1 in
11 green and the NLS motif in deep-purple. All three structures are at the same scale. In (**C**) the PA-
12 PB1(1-686) and RanBP5 envelopes are colored in grey and purple respectively.

13

14

Supplementary information

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Supplementary Figure 1

Sequences alignment of PB2. The alignment has been made for the sequences of the 2 human-IAV strains used for this work plus the 2 strains for which the structures have been recently solved. bat-IAV corresponds to A/little yellow-shouldered bat/Guatemala/060/2010(H17N10) (UniProtKB access number: H6QM90), H3N2-IAV to A/Victoria/3/1975(H3N2) (H9XIJ5), H5N1-IAV to the highly pathogenic A/Viet-Nam/1203/2004(H5N1) (Q6DNN3) and IBV to B/Memphis/13/2003 (Q5V8X3). The secondary sequences elements shown over and below the sequences alignment correspond to bat-IAV (PDB access numbers 4WSB) and IBV (4WSA) respectively. The alignment has been done using Clustal X2 [60] and drawn with ESPript [61].

Supplementary Figure 2

Endonuclease activity. An unstructured 52 nt poly-UC RNA (10 μ M) [25] was incubated with the wild-type (wt) or E80A mutant of PA-PB1-PB2(1-116) (12 μ M) or the N-terminal domain of PA (PA-Nter) (12 μ M) for 2 h at 37°C in a final volume of 30 μ l. Reactions were done in the presence and/or in the absence of 0.5 mM MnCl₂ in the reaction buffer 50 mM Tris-HCl pH 8,5, 25 mM KCl, 2,5 mM NaCl, 10 mM β -mercaptoethanol as previously described [24,25]. Reactions were inhibited by addition of 2 mM 2,4-dioxo-4-phenylbutanoic acid (DPBA) in the same conditions. As controls, the RNA was incubated with 0.5 mM MnCl₂ or with 2 mM DPBA or alone (ctrl). Reactions were stopped by adding 20 mM EGTA. The reactions were then loaded on an 8 M urea polyacrylamide gel. After migration, the gel was revealed by methylene blue staining.

1 **Supplementary Figure 3**

2 **De novo synthesis activity assay.** (A) PA-PB1(1-686) and (B) PA-PB1-PB2(1-36) polymerases were
3 incubated at 37°C with the 80-mer panhandle RNA, NTPs and radiolabelled [α P³²] UTP. Reactions
4 were stopped at 30, 60, 90 and 120 minutes with the addition of EGTA before loading on a 15 % urea
5 PAGE. Revelation was performed using a phosphorus screen and Typhoon scanner (GE Healthcare).

6 **Supplementary Figure 4**

7 **Effect of PA-Nter depletion on PA-PB1 oligomerzation.** Individual SEC-MALLS chromatography runs
8 of (A) PA-PB1(1-686) and (B) PA(196-716)-PB1(1-686) have been performed in the same buffer (50
9 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol) using a S200increase (10/300 GL, GE
10 Healthcare) column. UV absorbance at 280 nm is plotted against the elution volume. MALLS-RI
11 molecular mass estimates are plotted below the curve.

12 **Supplementary Figure 5**

13 **Crystallization of PA(197-716)-PB1(1-686).** (A) Snapshot of a crystal mounted on the beamline. The
14 red cross corresponds to the position of the beam. (B) Statistics of the best data collected on a single
15 crystal. (C) Comparison of (left) PA(196-716)-PB1(1-686) obtained by molecular replacement and
16 (right) the already published crystal structure of bat-IAV RdRp (PDB access: 4WSB) [6]. Both models
17 use the same color code, i.e. marine for PA, green-forest for PB1. The additional parts of bat-IAV RdRp
18 are colored in light-blue and light-red for the endonuclease domain of PA (PA1-195) and PB2 (1-759)
19 domain respectively. The panel was prepared using PYMOL [62]. Note that PA(196-716)-PB1(1-686)
20 model is made by 2 molecules of PA(197-716)-PB1(1-686) in the asymmetric unit. On the figure, only
21 one is shown. The second occupies the position of the endonuclease domain in the bat-IAV RdRp
22 model.

1 **Supplementary Figure 6**

2 **RanBP5 and PA-PB1(1-686) does not form a stable complex when mixed.** PA-PB1(1-686) and RanBP5
3 were separately purified before being mixed in equimolar proportions. The mixture was incubated 1h
4 at 20°C before analysis. Individual SEC-MALLS chromatography runs of **(A)** PA-PB1(1-686), **(B)** RanBP5
5 and **(C)** PA-PB1(1-686) + RanBP5 were performed in the same buffer (50 mM Tris-HCl pH 8.0, 150 mM
6 NaCl, 5 mM β -mercaptoethanol) using a S200increase (10/300 GL, GE Healthcare) column. UV
7 absorbance at 280 nm is plotted against the elution volume. MALLS-RI molecular mass estimates are
8 plotted below the curve.

9 **Supplementary Figure 7**

10 **R_g determination by Guinier extrapolation.** Guinier plots were calculated on low q regions and are
11 linear for (A) PA-PB1(1-686), (B) PA-PB1(1-686)-RanBP5 and (C) RanBP5.

12 **Supplementary Figure 8**

13 **The 5'-vRNAp does not dissociate or bind to the PA-PB1(1-686)-RanBP5 complex.** Prior to injection
14 on SEC-MALLS, the RdRp-RanBP5 complex was incubated with a 3 fold molar concentration of 5'-
15 vRNAp for 1h at 20°C. SEC-MALLS chromatograms are shown for both RdRp-RanBP5 without **(A)** and
16 with **(B)** 5'-vRNAp. Both analyses were performed in the same buffer (50 mM Tris-HCl pH 8.0, 150 mM
17 NaCl, 5 mM β -mercaptoethanol).

18

References

1. Cianci C, Tiley L, Krystal M (1995) Differential activation of the influenza virus polymerase via template RNA binding. *J Virol* 69: 3995-3999.
2. Hagen M, Chung TD, Butcher JA, Krystal M (1994) Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J Virol* 68: 1509-1515.
3. Tiley LS, Hagen M, Matthews JT, Krystal M (1994) Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J Virol* 68: 5108-5116.
4. Plotch SJ, Bouloy M, Ulmanen I, Krug RM (1981) A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* 23: 847-858.
5. Reich S, Guilligay D, Pflug A, Malet H, Berger I, et al. (2014) Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 516: 361-366.
6. Pflug A, Guilligay D, Reich S, Cusack S (2014) Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516: 355-360.
7. Mukaigawa J, Nayak DP (1991) Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *J Virol* 65: 245-253.
8. Tarendeau F, Boudet J, Guilligay D, Mas P, Bougault C, et al. (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* 14: 229-233.
9. Deng T, Engelhardt OG, Thomas B, Akoulitchev AV, Brownlee GG, et al. (2006) Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J Virol* 80: 11911-11919.
10. Huet S, Avilov SV, Ferbitz L, Daigle N, Cusack S, et al. (2010) Nuclear import and assembly of influenza A virus RNA polymerase studied in live cells by fluorescence cross-correlation spectroscopy. *J Virol* 84: 1254-1264.
11. Hutchinson EC, Orr OE, Man Liu S, Engelhardt OG, Fodor E (2011) Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J Gen Virol* 92: 1859-1869.
12. Chang S, Sun D, Liang H, Wang J, Li J, et al. (2015) Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution. *Mol Cell* 57: 925-935.
13. Al-Tawfiq JA, Zumla A, Memish ZA (2014) Coronaviruses: severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus in travelers. *Curr Opin Infect Dis* 27: 411-417.
14. Lucas S, Nelson AM (2015) HIV and the spectrum of human disease. *J Pathol* 235: 229-241.
15. Crepin T, Swale C, Monod A, Garzoni F, Chaillet M, et al. (2015) Polyproteins in structural biology. *Curr Opin Struct Biol* 32: 139-146.
16. Nie Y, Bellon-Echeverria I, Trowitzsch S, Bieniossek C, Berger I (2014) Multiprotein complex production in insect cells by using polyproteins. *Methods Mol Biol* 1091: 131-141.
17. Bieniossek C, Imasaki T, Takagi Y, Berger I (2012) MultiBac: expanding the research toolbox for multiprotein complexes. *Trends Biochem Sci* 37: 49-57.
18. Trowitzsch S, Bieniossek C, Nie Y, Garzoni F, Berger I (2010) New baculovirus expression tools for recombinant protein complex production. *J Struct Biol* 172: 45-54.

- 1 19. Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JR, et al. (2009) Structural insight into the essential
2 PB1-PB2 subunit contact of the influenza virus RNA polymerase. *Embo J* 28: 1803-1811.
- 3 20. An Y, Meresse P, Mas PJ, Hart DJ (2011) CoESPRIT: a library-based construct screening method for
4 identification and expression of soluble protein complexes. *PLoS One* 6: e16261.
- 5 21. Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, et al. (2008) The structural basis for cap
6 binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* 15: 500-506.
- 7 22. Bieniossek C, Richmond TJ, Berger I (2008) MultiBac: multigene baculovirus-based eukaryotic protein
8 complex production. *Curr Protoc Protein Sci Chapter 5: Unit 5 20*.
- 9 23. Baudin F, Bach C, Cusack S, Ruigrok RW (1994) Structure of influenza virus RNP. I. Influenza virus
10 nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *Embo J*
11 13: 3158-3165.
- 12 24. Crepin T, Dias A, Palencia A, Swale C, Cusack S, et al. (2010) Mutational and metal binding analysis of the
13 endonuclease domain of the influenza virus polymerase PA subunit. *J Virol* 84: 9096-9104.
- 14 25. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, et al. (2009) The cap-snatching endonuclease of
15 influenza virus polymerase resides in the PA subunit. *Nature* 458: 914-918.
- 16 26. He X, Zhou J, Bartlam M, Zhang R, Ma J, et al. (2008) Crystal structure of the polymerase PA(C)-PB1(N)
17 complex from an avian influenza H5N1 virus. *Nature* 454: 1123-1126.
- 18 27. Obayashi E, Yoshida H, Kawai F, Shibayama N, Kawaguchi A, et al. (2008) The structural basis for an
19 essential subunit interaction in influenza virus RNA polymerase. *Nature* 454: 1127-1131.
- 20 28. Desselberger U, Racaniello VR, Zazra JJ, Palese P (1980) The 3' and 5'-terminal sequences of influenza A, B
21 and C virus RNA segments are highly conserved and show partial inverted complementarity. *Gene* 8: 315-
22 328.
- 23 29. Robertson JS (1979) 5' and 3' terminal nucleotide sequences of the RNA genome segments of influenza
24 virus. *Nucleic Acids Res* 6: 3745-3757.
- 25 30. Skehel JJ, Hay AJ (1978) Nucleotide sequences at the 5' termini of influenza virus RNAs and their transcripts.
26 *Nucleic Acids Res* 5: 1207-1219.
- 27 31. Brownlee GG, Sharps JL (2002) The RNA polymerase of influenza A virus is stabilized by interaction with its
28 viral RNA promoter. *J Virol* 76: 7103-7113.
- 29 32. Tomescu AI, Robb NC, Hengrung N, Fodor E, Kapanidis AN (2014) Single-molecule FRET reveals a corkscrew
30 RNA structure for the polymerase-bound influenza virus promoter. *Proc Natl Acad Sci U S A* 111: E3335-
31 3342.
- 32 33. Melen K, Fagerlund R, Franke J, Kohler M, Kinnunen L, et al. (2003) Importin alpha nuclear localization
33 signal binding sites for STAT1, STAT2, and influenza A virus nucleoprotein. *J Biol Chem* 278: 28193-28200.
- 34 34. Naito T, Momose F, Kawaguchi A, Nagata K (2007) Involvement of Hsp90 in assembly and nuclear import of
35 influenza virus RNA polymerase subunits. *J Virol* 81: 1339-1349.
- 36 35. Portela A, Digard P (2002) The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal
37 to virus replication. *J Gen Virol* 83: 723-734.
- 38 36. Weber F, Kochs G, Gruber S, Haller O (1998) A classical bipartite nuclear localization signal on Thogoto and
39 influenza A virus nucleoproteins. *Virology* 250: 9-18.
- 40 37. Fodor E, Smith M (2004) The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of
41 the influenza A virus RNA polymerase complex. *J Virol* 78: 9144-9153.

- 1 38. Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering.
2 Nature 496: 477-481.
- 3 39. Momose F, Naito T, Yano K, Sugimoto S, Morikawa Y, et al. (2002) Identification of Hsp90 as a stimulatory
4 host factor involved in influenza virus RNA synthesis. J Biol Chem 277: 45306-45314.
- 5 40. Chase G, Deng T, Fodor E, Leung BW, Mayer D, et al. (2008) Hsp90 inhibitors reduce influenza virus
6 replication in cell culture. Virology 377: 431-439.
- 7 41. Fitzgerald DJ, Schaffitzel C, Berger P, Wellinger R, Bieniossek C, et al. (2007) Multiprotein expression
8 strategy for structural biology of eukaryotic complexes. Structure 15: 275-279.
- 9 42. Palmberger D, Rendic D (2015) SweetBac: Applying MultiBac Technology Towards Flexible Modification of
10 Insect Cell Glycosylation. Methods Mol Biol 1321: 153-169.
- 11 43. Deng T, Sharps J, Fodor E, Brownlee GG (2005) *In vitro* assembly of PB2 with a PB1-PA dimer supports a new
12 model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. J Virol 79:
13 8669-8674.
- 14 44. Graef KM, Vreede FT, Lau YF, McCall AW, Carr SM, et al. (2010) The PB2 subunit of the influenza virus RNA
15 polymerase affects virulence by interacting with the mitochondrial antiviral signaling protein and inhibiting
16 expression of beta interferon. J Virol 84: 8433-8445.
- 17 45. Patel D, Schultz LW, Umland TC (2013) Influenza A polymerase subunit PB2 possesses overlapping binding
18 sites for polymerase subunit PB1 and human MAVS proteins. Virus Res 172: 75-80.
- 19 46. Degut C, Monod A, Brachet F, Crepin T, Tisne C (2016) *In vitro/in vivo* production of tRNA for X-ray studies.
20 Methods Mol Biol 1320: 37-57.
- 21 47. Price SR, Ito N, Oubridge C, Avis JM, Nagai K (1995) Crystallization of RNA-protein complexes. I. Methods for
22 the large-scale preparation of RNA suitable for crystallographic studies. J Mol Biol 249: 398-408.
- 23 48. Boulo S, Akarsu H, Lotteau V, Muller CW, Ruigrok RW, et al. (2011) Human importin alpha and RNA do not
24 compete for binding to influenza A virus nucleoprotein. Virology 409: 84-90.
- 25 49. Chenavas S, Estrozi LF, Slama-Schwok A, Delmas B, Di Primo C, et al. (2013) Monomeric nucleoprotein of
26 influenza A virus. PLoS Pathog 9: e1003275.
- 27 50. Petoukhov MV, Konarev PV, Kikhney AG, Svergun DI (2007) ATSAS 2.1 – towards automated and web-
28 supported small-angle scattering data analysis. J Appl Crystallogr 40: s223–s228.
- 29 51. Rambo RP, Tainer JA (2013) Super-resolution in solution X-ray scattering and its applications to structural
30 systems biology. Annu Rev Biophys 42: 415-441.
- 31 52. Semenyuk AV, Svergun DI (1991) GNOM – a program package for small-angle scattering data processing. J
32 Appl Crystallogr 24: 537–540.
- 33 53. Franke D, Svergun DI (2009) DAMMIF , a program for rapid *ab-initio* shape determination in small-angle
34 scattering. J Appl Crystallogr 42: 342–346.
- 35 54. Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. J Appl
36 Crystallogr 36: 860–864.
- 37 55. Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering
38 using simulated annealing. Biophys J 76: 2879–2886.
- 39 56. Svergun D, Barberato C, Koch MHJ (1995) CRY SOL – a program to evaluate X-ray solution scattering of
40 biological macromolecules from atomic coordinates. J Appl Crystallogr 28: 768–773.
- 41 57. Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-
42 angle scattering data. Biophys J 89: 1237-1250.

1 58. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. (2004) UCSF Chimera--a visualization
2 system for exploratory research and analysis. *J Comput Chem* 25: 1605-1612.

3 59. Wong I, Lohman TM (1993) A double-filter method for nitrocellulose-filter binding: application to protein-
4 nucleic acid interactions. *Proc Natl Acad Sci U S A* 90: 5428-5432.

5 60. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X
6 version 2.0. *Bioinformatics* 23: 2947-2948.

7 61. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server.
8 *Nucleic Acids Res* 42: W320-324.

9 62. DeLano WL (2002) *The PyMOL Molecular Graphics System*. San Carlos, CA, USA: DeLano Scientific.

10
11
12

FIGURES AND SUPPLEMENTARY FIGURES

(Note : Figure captions were re-inserted next to the figures in order to help in the reading process)

Table 1: Details of human-IAV RdRp polyproteins. The expressions have been made on both A/Victoria/3/1975(H3N2) and A/Viet-Nam/1203/2004(H5N1) strains. The ratio YFP/CFP were calculated with the 2 maxima of the fluorescence spectra measured on the same sample (YFP: $\lambda_{\text{exi}} = 488 \text{ nm}$, $\lambda_{\text{emi}} = 525 \text{ nm}$; CFP: $\lambda_{\text{exi}} = 430 \text{ nm}$, $\lambda_{\text{emi}} = 480 \text{ nm}$) corresponding to 1×10^6 infected cells. This ratio is proportional to the level of expression of the corresponding constructs. The table contains also data on the solubility of each construct in the classical purification buffer (i.e.: 50 mM Tris-HCl pH 8.5, 300 mM NaCl, 2mM β -mercaptoethanol, 2-10 % glycerol). *This construct has been cloned in pFastBac-HTB without TEV nor CFP. The only fluorescent reporter protein was the YFP. Its signal was in accordance with the YFP values obtained for the other constructs.

	Construct number	PA	PB1	PB2	RanBP5	YFP/CFP	Solubility
heterodimers	1	1-716	1-686	-	-	6	yes
	2	197-716	1-686	-	-	4	yes
	3	197-716	1-660	-	-	4	yes
	4	197-716	1-560	-	-	4	no
	5	210-716	1-686	-	-	6	no
	6	222-716	1-686	-	-	6	no
	7	231-716	1-686	-	-	6	no
	8	240-716	1-686	-	-	6	no
	9	250-716	1-686	-	-	6	no
	10	263-716	1-686	-	-	6	no
	11	197-263	16-686	-	-	n.d.*	no
heterotrimers	12	1-716	1-757	1-759	-	50	n.d.
	13	1-716	1-757	1-36	-	15	yes
	14	197-716	1-757	1-116	-	7	yes
	15	1-716	1-757	1-116	-	17	yes
	16	1-716	1-757	1-250	-	28	n.d.
	17	1-716	1-757	1-320	-	30	n.d.
	18	1-716	1-757	1-483	-	41	n.d.
	19	1-716	1-757	1-116 - MBP	-	17	yes
	20	1-716	1-757	1-116 - 320-483	-	20	yes
	21	1-716	1-686	-	1-1115	12	yes

Table 2: Titration measurements against vRNA promoters. Values in parentheses correspond to the values obtained by filter-binding assay experiments. All the experiments have been made in triplicate.

Construct	NaCl (mM)	RNA	Kd (nM)	std error (nM)	R ²
construct 14 : PA-PB1- PB2(1-116)	300	5'-vRNAp	0.83 (0.19)	0.08 (0.02)	0.99 (0.97)
		3'-vRNAp	36 (17)	2 (2)	0.99 (0.95)
		polyUC	≥ 1000	n.d.	n.d.
construct 1 : PA-PB1(1-686)	300	5'-vRNAp	0.87 (0.38)	0.1 (0.05)	0.99 (0.99)
		3'-vRNAp	≥ 1000	n.d.	n.d.
		polyUC	≥ 1000	n.d.	n.d.
	150	5'-vRNAp	0.53	0.12	0.99
		3'-vRNAp	22.1	1.0	0.99
		polyUC	22.8	0.7	0.99
construct 21 : PA-PB1(1-686) RanBP5	300	5'-vRNAp	≥ 1000	n.d.	n.d.

Table 3: SAXS data-collection and scattering-derived parameters.

	Construct 1 : PA-PB1(1-686)	Construct 21 : PA-PB1(1-686)- RanBP5	Construct 22 : RanBP5
Data collection parameters			
Instrument		ESRF - BM29	
Beam size at sample (μm)		700 x 700	
Wavelength (\AA)		0.9919	
q range (\AA^{-1})		0.25 - 50	
Detector		Pilatus 1M	
Detector distance (m)		2.867	
Exposure (s per image)		1	
Column		S200inc 5/150 GL	
Flow rate ($\text{mL}\cdot\text{min}^{-1}$)	0.5	0.4	0.5
Injected sample concentrations ($\text{mg}\cdot\text{mL}^{-1}$)	3.4	4.3	8.5
Injection volume (μL)		50	
Temperature (K)		293	
Structural parameters			
R_g (\AA) [from $P(r)$]	37.9 ± 0.1	52.2 ± 0.1	39.4 ± 0.2
R_g (\AA) [from Guinier]	36.2 ± 0.4	51.8 ± 0.5	38.8 ± 0.8
D_{max} (\AA)	128	181	136
Porod volume estimate (\AA^3)	254 630	577 070	215 890
Molecular-mass determination			
Molecular mass M_r (Da) [from Rambo]	146 493	323 311	143 827
Calculated M_r (Da) from sequence	165 915	291 983	125 892
Software employed			
Primary data reduction		PRIMUS	
Data processing		PRIMUS	
<i>Ab initio</i> analysis		DAMMIF	
Validation and averaging		DAMAVER & DAMMIN / MONSA	
Computation of model intensities		CRY SOL	
3D graphics representations		CHIMERA	

Figure 1

Truncated human-IAV RdRp polyprotein expression. (A) Schematic theory of the TEV-PA-PB1-PB2-CFP polyprotein expression. During expression, TEV will process the polyprotein ensuring a stoichiometric assembly of PA, PB1 and PB2. YFP and CFP are produced during the process and monitor respectively baculovirus proliferation and polyprotein translation. (B) Truncated RdRp constructs where PB2 is incrementally extended until residues 116, 320 or 483. (C) YFP (left) and CFP (right) fluorescence kinetics measured during Hi-5 insect cells culture of truncated RdRp constructs. YFP ($\lambda_{\text{exi}} = 488 \text{ nm}$, $\lambda_{\text{emi}} = 525 \text{ nm}$) and CFP ($\lambda_{\text{exi}} = 430 \text{ nm}$, $\lambda_{\text{emi}} = 480 \text{ nm}$) measurements were performed on cellular extracts prepared by sonicating 1×10^6 cells in PBS (500 μL) follow by centrifugation. Fluorescence intensities are plotted against time after infection. (D) Small scale nickel resin purification analysis by western blot. Purifications were performed on the 50 mL of Hi-5 insect cells cultures used for the YFP and CFP fluorescence kinetics (C). Deposits feature total lysate after freeze/thaw (T), supernatant after centrifugation (S), resin flow through (Ft) and the primary elution fraction (E). After migration on a 12 % SDS-PAGE, proteins were transferred on PVDF membrane. Primary antibodies targeting human-IAV PB2 (rabbit IgG) and human-IAV PB1 (mouse IgG) have been used. Revelation was performed with secondary goat antibodies coupled with Alexa Fluor 532 ($\lambda_{\text{exi}} = 632 \text{ nm}$, $\lambda_{\text{emi}} = 647 \text{ nm}$) and Alexa Fluor 633 ($\lambda_{\text{exi}} = 531 \text{ nm}$, $\lambda_{\text{emi}} = 554 \text{ nm}$) targeting mouse and rabbit H + L domains respectively, using a Typhoon Trio imaging system (GE Healthcare). After integration of the raw data, PB1 and PB2 revelation are visible in red and green respectively. Black and white signal of PB2 is also shown (bottom) to highlight the PB2 truncations. The upper bands appearing in green/yellow, correspond to unprocessed polyproteins.

Figure 1

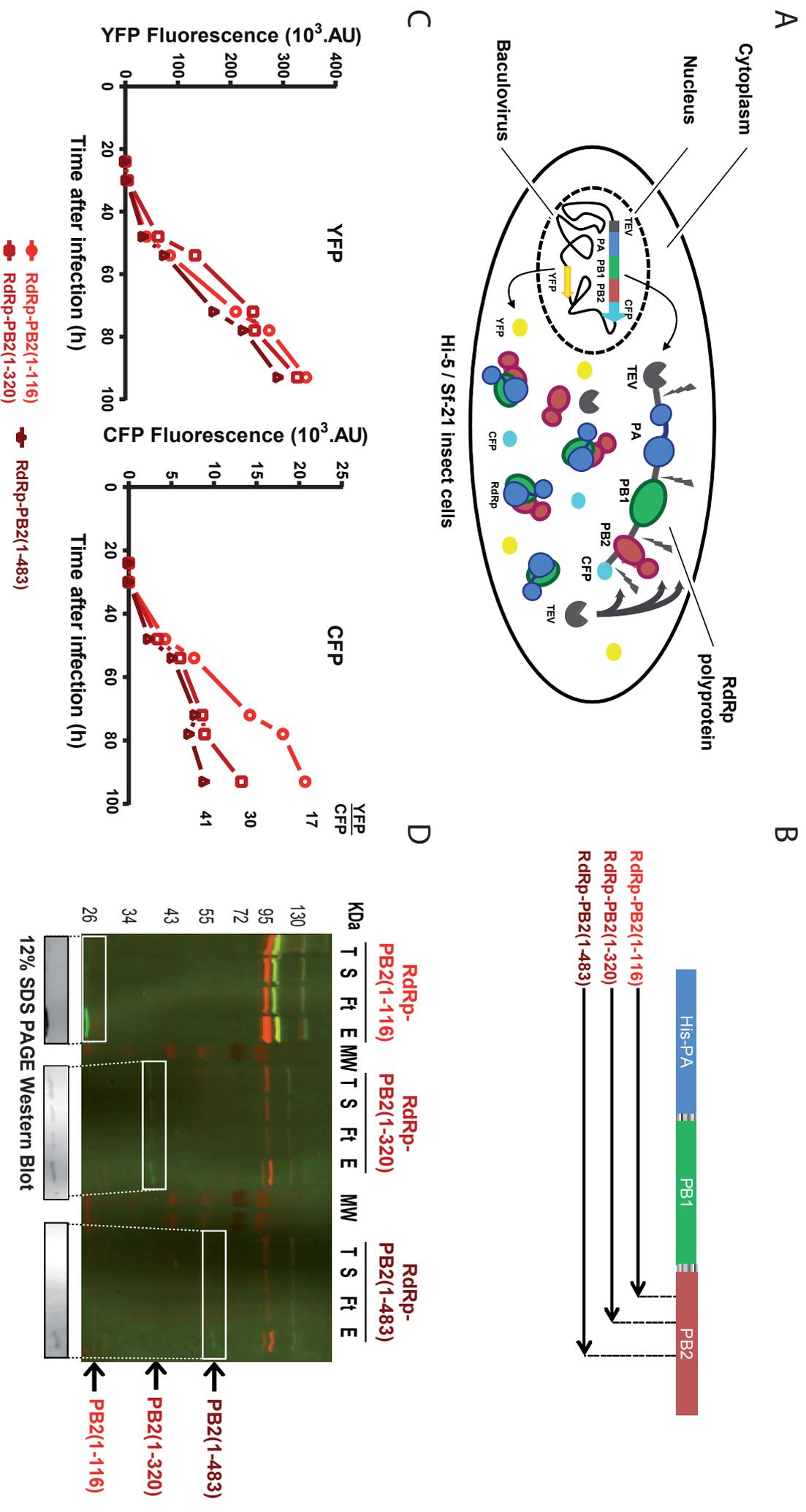
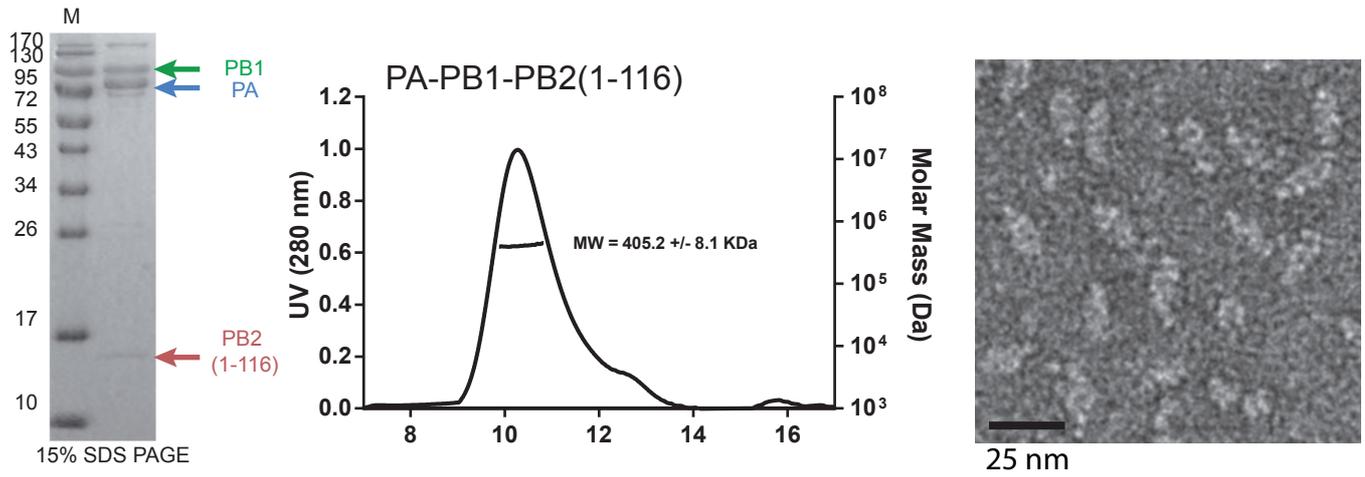


Figure 2

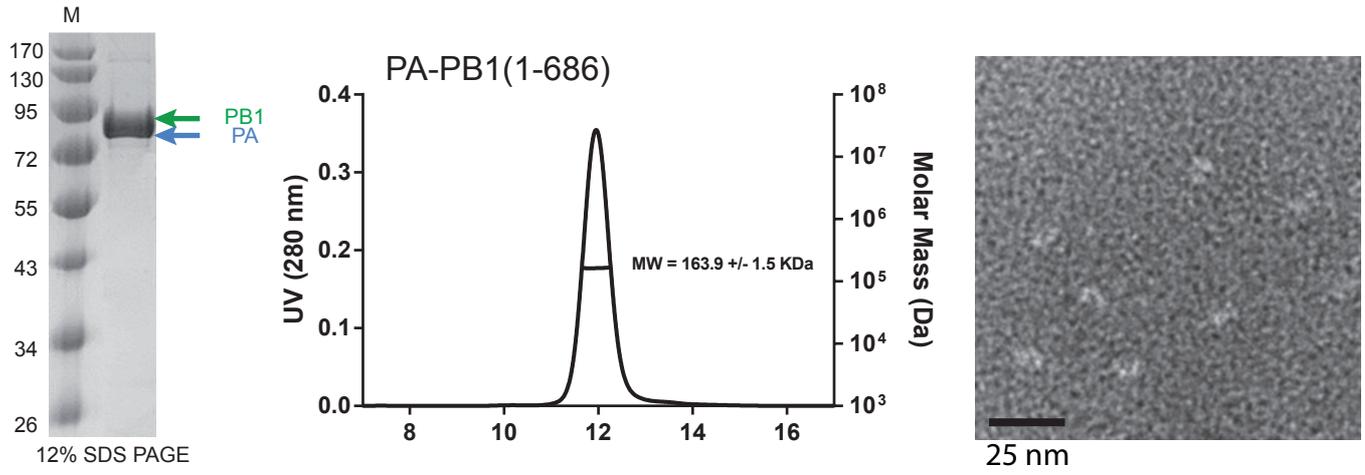
Homogeneous sample analysis of recombinant RdRp, RanBP5 and RdRp-RanBP5. Purified product analysis are horizontally grouped for **(A)** PA-PB1-PB2(1-116), **(B)** PA-PB1(1-686), **(C)** RanBP5 and **(D)** PA-PB1(1-686)-RanBP5. On the right are the coomassie blue colored SDS PAGE gel of the purified sample with corresponding subunits bands indicated by colored arrows (PA in blue, PB1 in green and RanBP5 in purple). In the middle are the SEC-MALLS chromatograms with the UV signal as a backtrace and molecular weight estimate below the peak. Estimated average molecular weight for each sample is also detailed. SEC-MALLS runs of PA-PB1(1-686), RanBP5 and PA-PB1(1-686)-RanBP5 were performed in the same buffer (50 mM Tris-HCl pH 8.0 and 150 mM NaCl) whereas the SEC-MALLS run of PA-PB1-PB2(1-116) was performed with 50 mM Tris-HCl pH 8.0, 300 mM NaCl and 10% glycerol. (Left) Negative stain transmission EM image realized on the preparation at 0.2 mg.mL^{-1} .

Figure 2

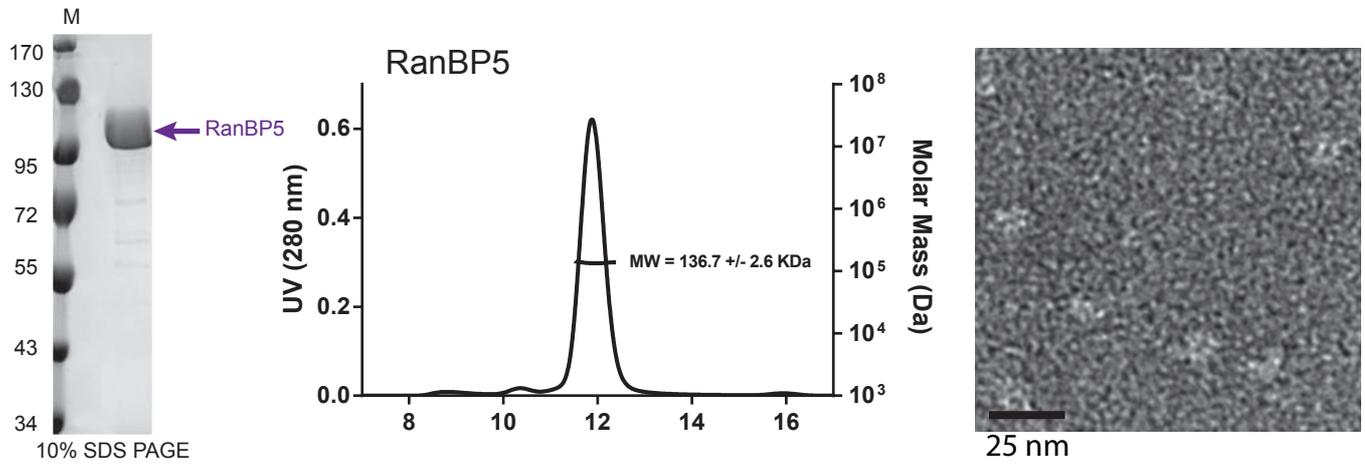
A



B



C



D

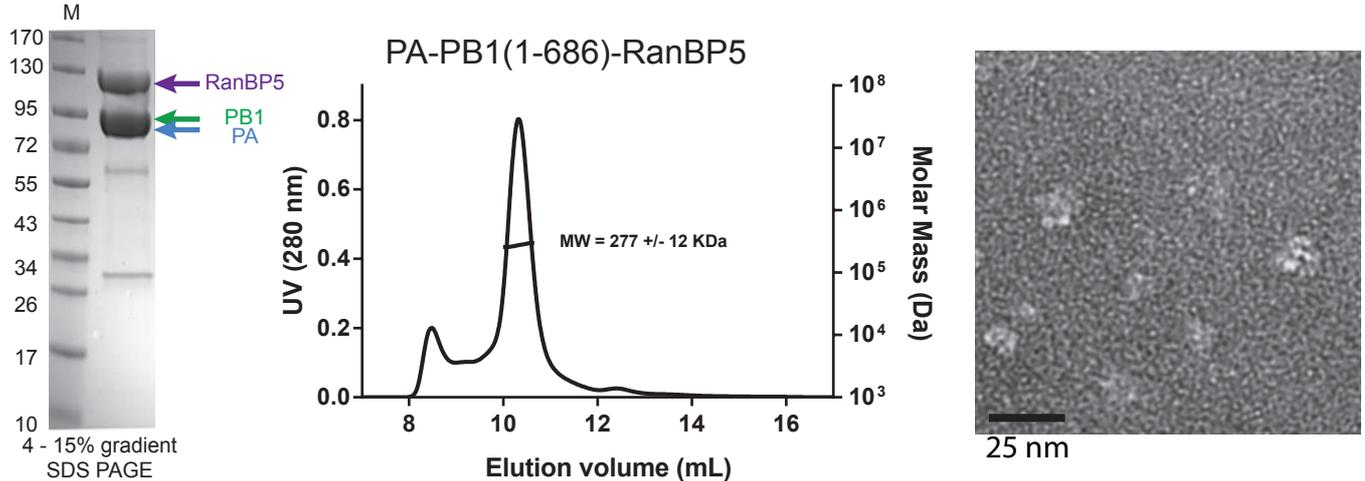


Figure 3

vRNA specificity. Binding affinity measurements of RdRp and RdRp-RanBP5 constructs towards 3' and 5' vRNA promoters. **(A)** Binding titration of the truncated trimer PA-PB1-PB2(1-116) towards the 5'-vRNAP (blue triangle) and 3'-vRNAP (red square) sequences using fluorescence anisotropy at 300 mM NaCl. **(B)** Binding titration performed by filter binding assay against P³² labelled 5'-vRNAP (blue triangle) and 3'-vRNAP (red square) using 300 mM NaCl. Bound RNA fraction is plotted as a function of polymerase concentration. **(C)** Binding titration of the truncated dimer PA-PB1(1-686) performed at 150 and 300 mM NaCl against the 5'-vRNAP (dark and light blue triangles), 3'-vRNAP (orange and dark red squares) and polyUC RNA (light and dark green circles) by fluorescence anisotropy. **(D)** Binding titration of different RdRp and RdRp-RanBP5 constructs against the 5'-vRNAP at 300 mM NaCl by fluorescence anisotropy. PA-PB1-PB2(116) and PA-PB1(1-686) are depicted by blue and orange triangles respectively, PA-PB1(1-686)-RanBP5 is depicted with purple triangles. For all anisotropy titrations **(A, C and D)** subtracted anisotropy is plotted as a function of protein concentration.

Figure 3

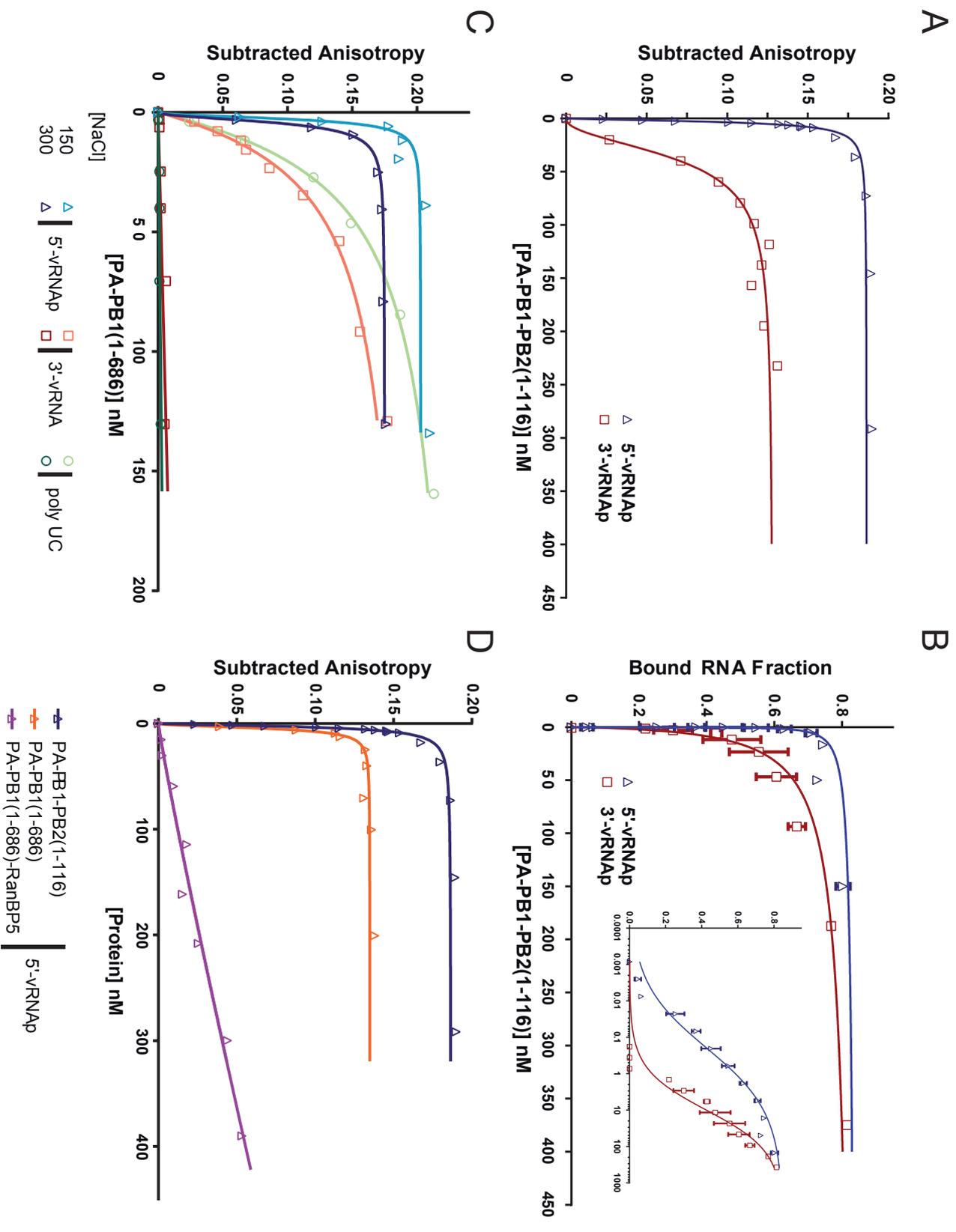
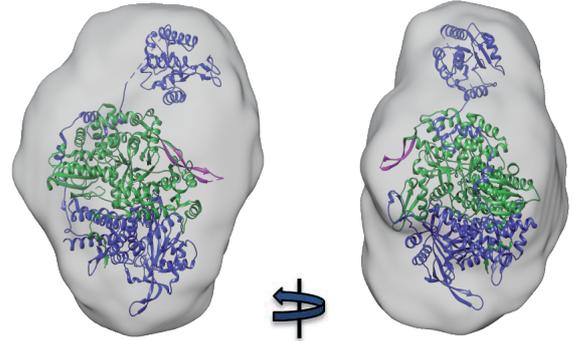
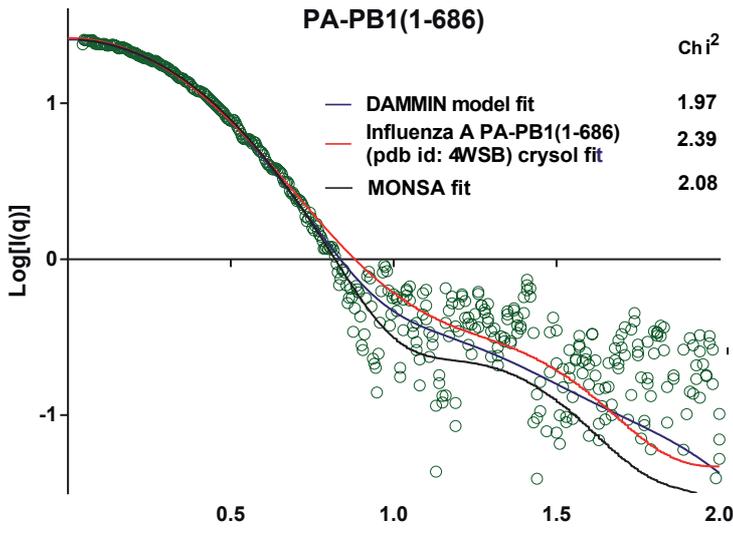


Figure 4

Online SAXS analysis of PA-PB1, RanBP5 and PA-PB1-RanBP5 complex in solution. Results are shown horizontally for (A) PA-PB1(1-686), (B) RanBP5 and (C) PA-PB1-RanBP5. On the left are the experimental $\ln[I(q)]$ curves with the *ab-initio* DAMMIN curve fit (blue) using the Damstart (DAMAVER) as an initial constraint and the CRY SOL fit (red) of the closest homologous X-Ray structure. The MONSA fit is also shown for the 3 curves in black. The χ^2 of the different curves are also shown. On the right are the homologous PDB structure depicted as cartoon structures fitted into the DAMAVER *ab-initio* envelope for (A) and (B) and into the MONSA modeling in (C). RanBP5 is depicted in purple, PA in blue, PB1 in green and the NLS motif in deep-purple. All three structures are at the same scale. In (C) the PA-PB1(1-686) and RanBP5 envelopes are colored in grey and purple respectively.

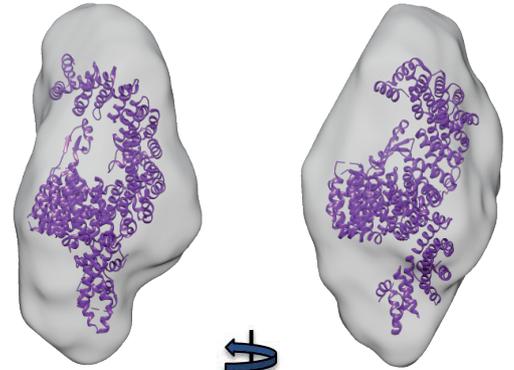
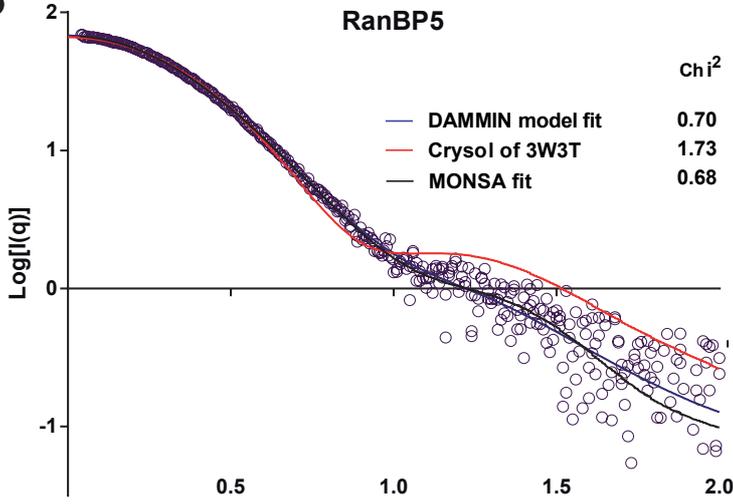
Figure 4

A



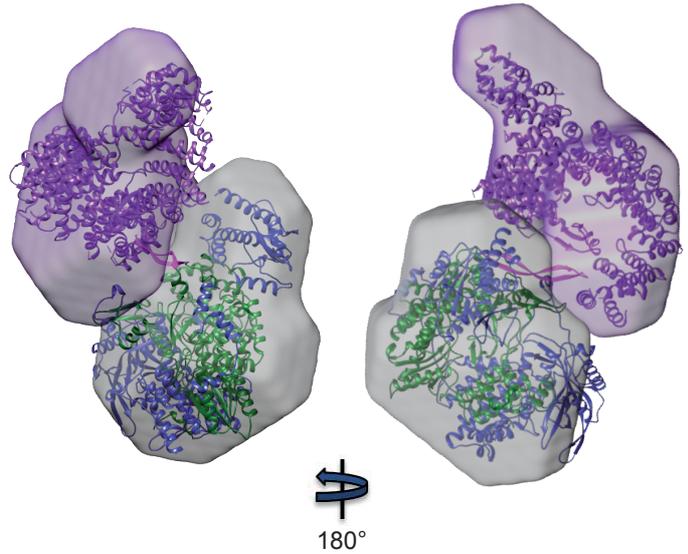
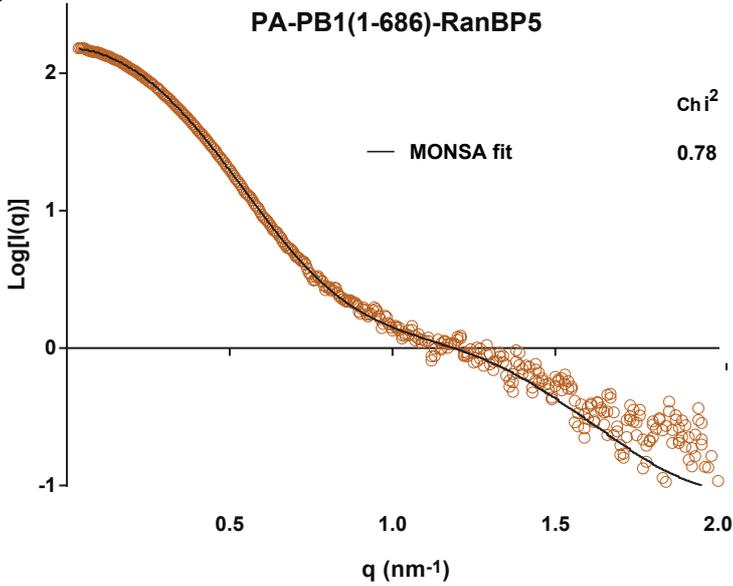
Rg (Å): 36.2
Dmax (Å): 128

B



Rg (Å): 38.8
Dmax (Å): 136

C



Rg (Å): 51.8
Dmax (Å): 181

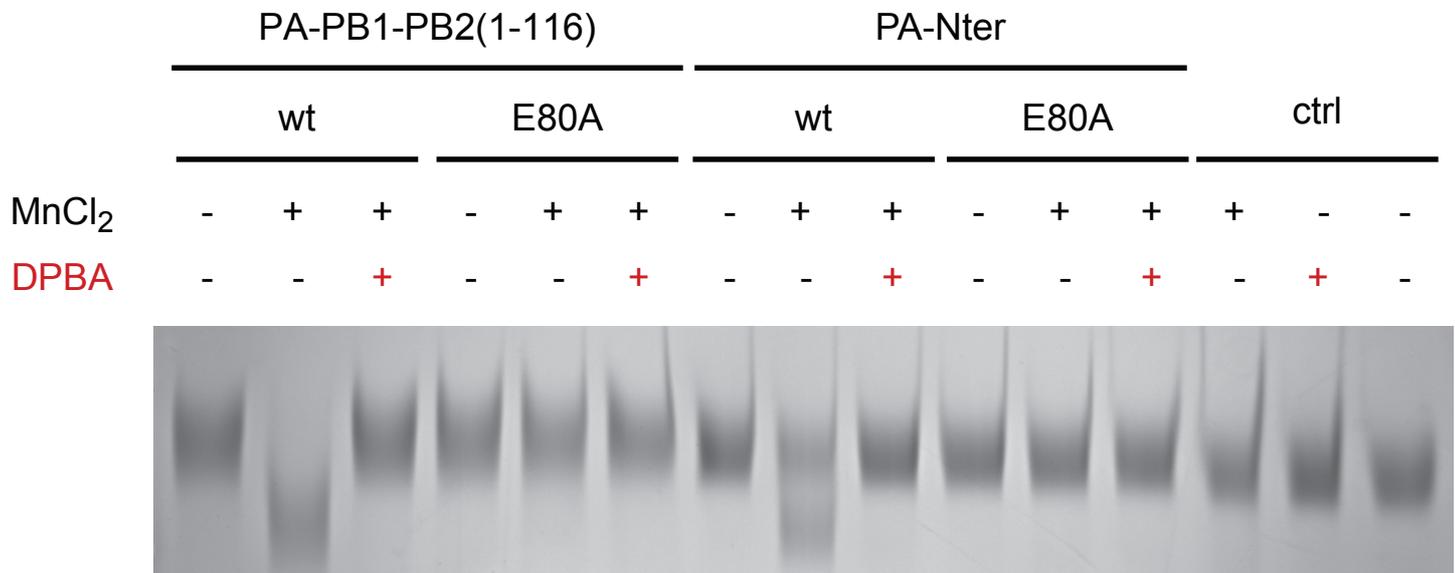
Supplementary Figure 1

Sequences alignment of PB2. The alignment has been made for the sequences of the 2 human-IAV strains used for this work plus the 2 strains for which the structures have been recently solved. bat-IAV corresponds to A/little yellow-shouldered bat/Guatemala/060/2010(H17N10) (UniProtKB access number: H6QM90), H3N2-IAV to A/Victoria/3/1975(H3N2) (H9XIJ5), H5N1-IAV to the highly pathogenic A/Viet-Nam/1203/2004(H5N1) (Q6DNN3) and IBV to B/Memphis/13/2003 (Q5V8X3). The secondary sequences elements shown over and below the sequences alignment correspond to bat-IAV (PDB access numbers 4WSB) and IBV (4WSA) respectively. The alignment has been done using Clustal X2 [53] and drawn with ESPript [54].

Supplementary Figure 2

Endonuclease activity. An unstructured 52 nt poly-UC RNA (10 μ M) [24] was incubated with the wild-type (wt) or E80A mutant of PA-PB1-PB2(1-116) (12 μ M) or the N-terminal domain of PA (PA-Nter) (12 μ M) for 2 h at 37°C in a final volume of 30 μ l. Reactions were done in the presence and/or in the absence of 0.5 mM $MnCl_2$ in the reaction buffer 50 mM Tris-HCl pH 8,5, 25 mM KCl, 2,5 mM NaCl, 10 mM β -mercaptoethanol as previously described [23,24]. Reactions were inhibited by addition of 2 mM 2,4-dioxo-4-phenylbutanoic acid (DPBA) in the same conditions. As controls, the RNA was incubated with 0.5 mM $MnCl_2$ or with 2 mM DPBA or alone (ctrl). Reactions were stopped by adding 20 mM EGTA. The reactions were then loaded on an 8 M urea polyacrylamide gel. After migration, the gel was revealed by methylene blue staining.

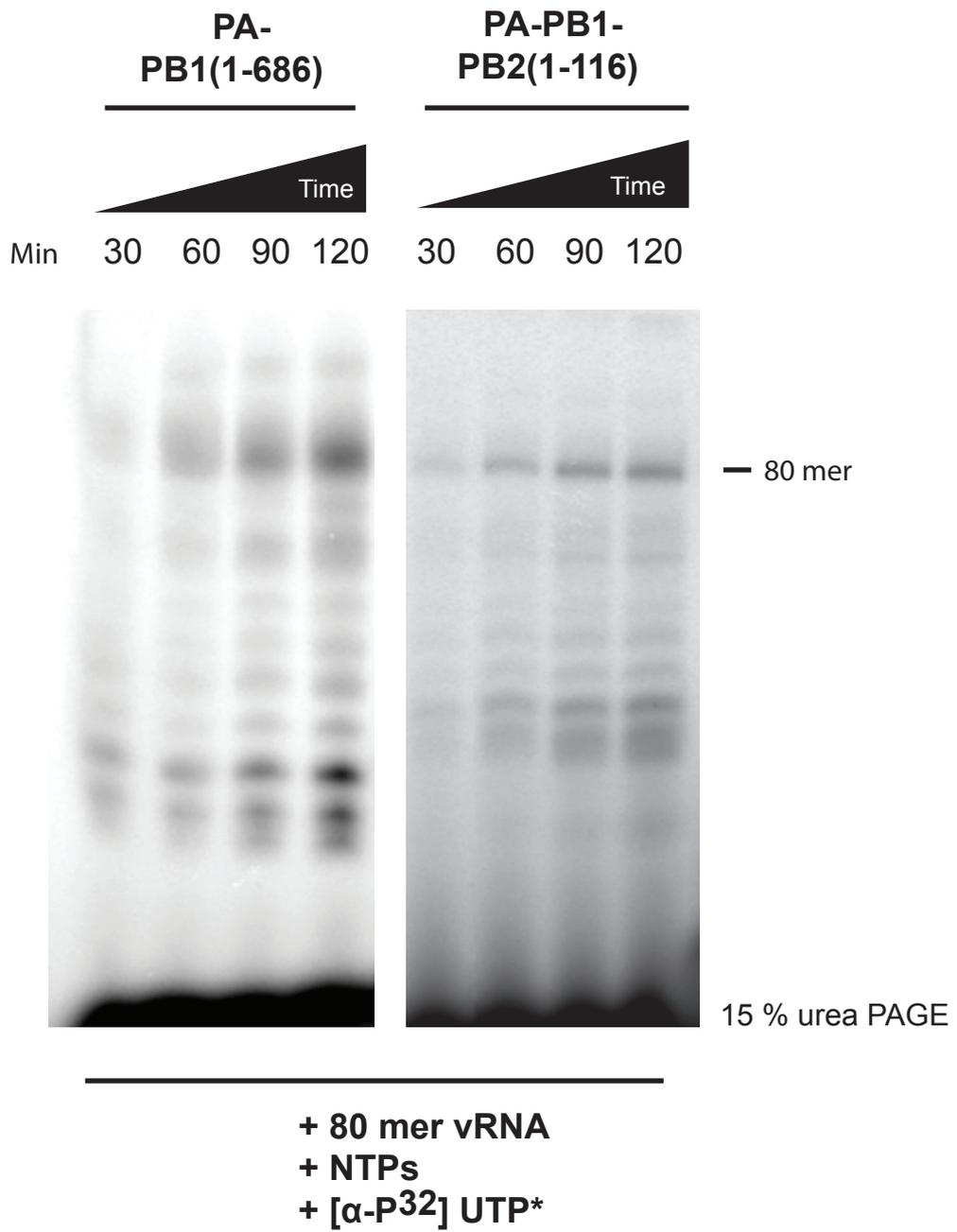
Supplementary Figure 2



Supplementary Figure 3

De novo synthesis activity assay. (A) PA-PB1(1-686) and (B) PA-PB1-PB2(1-36) polymerase preps were incubated at 37°C with the 80 mer panhandle RNA, NTPs and radiolabelled [α^{32}] UTP. Reactions were stopped at 30, 60, 90 and 120 minutes with the addition of EGTA before loading on a 15 % urea PAGE. Revelation was performed using a phosphorus screen and Typhoon scanner (GE Healthcare).

Supplementary Figure 3

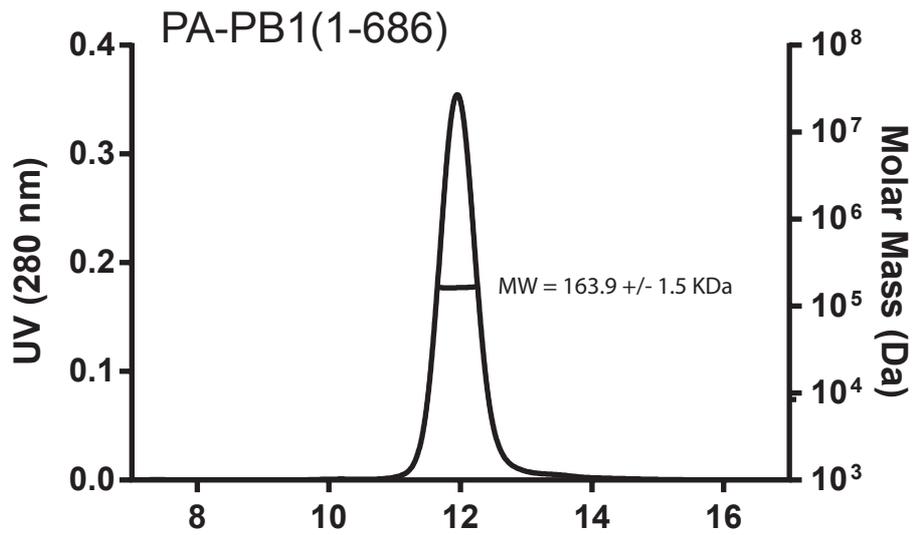


Supplementary Figure 4

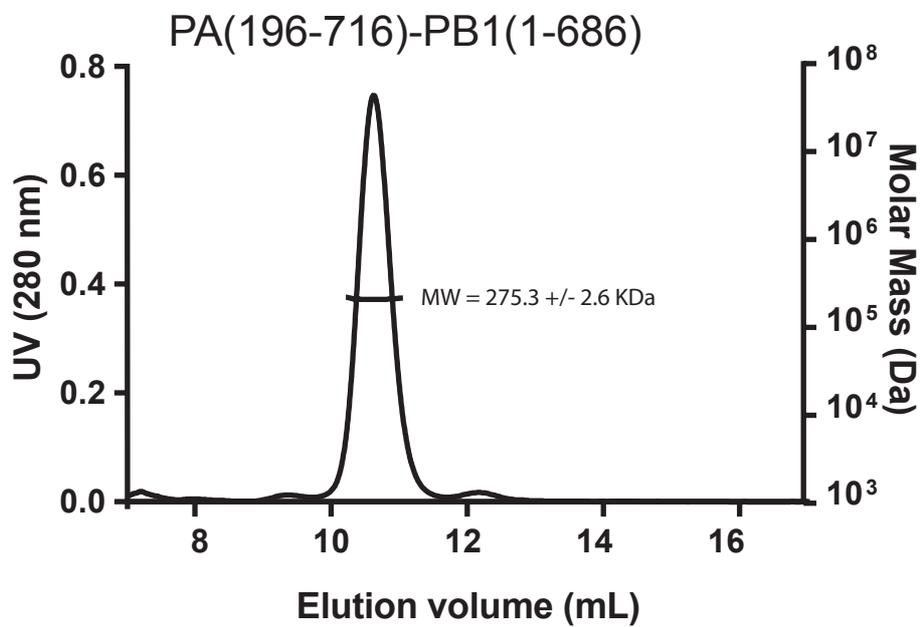
Effect of PA-Nter depletion on PA-PB1 oligomerization. Individual SEC-MALLS chromatography runs of (A) PA-PB1(1-686) and (B) PA(196-716)-PB1(1-686) have been performed in the same buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol) using a S200increase (10/300 GL, GE Healthcare) column. UV absorbance at 280 nm is plotted against the elution volume. MALLS molecular mass estimates are plotted below the curve.

Supplementary Figure 4

A



B

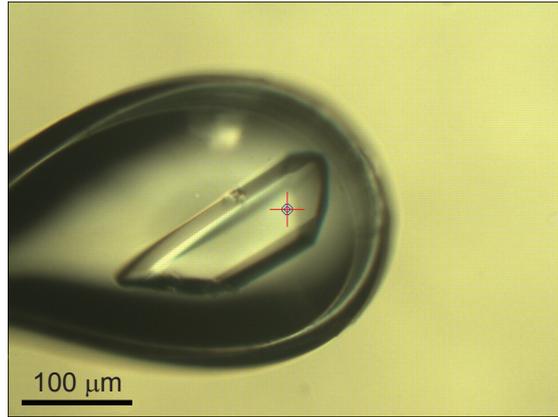


Supplementary Figure 5

Crystallization of PA(197-716)-PB1(1-686). (A) Snapshot of a crystal mounted on the beamline. The red cross corresponds to the position of the beam. (B) Statistics of the best data collected on a single crystal. (C) Comparison of (left) PA(196-716)-PB1(1-686) obtained by molecular replacement and (right) the already published crystal structure of bat-IAV RdRp (PDB access: 4WSB) [6]. Both models use the same color code, i.e. marine for PA, green-forest for PB1. The additional parts of bat-IAV RdRp are colored in light-blue and light-red for the endonuclease domain of PA (PA1-195) and PB2 (1-759) domain respectively. The panel was prepared using PYMOL [55]. Note that PA(196-716)-PB1(1-686) model is made by 2 molecules of PA(197-716)-PB1(1-686) in the asymmetric unit. On the figure, only one is shown. The second occupies the position of the endonuclease domain in the bat-IAV RdRp model.

Supplementary Figure 5

A



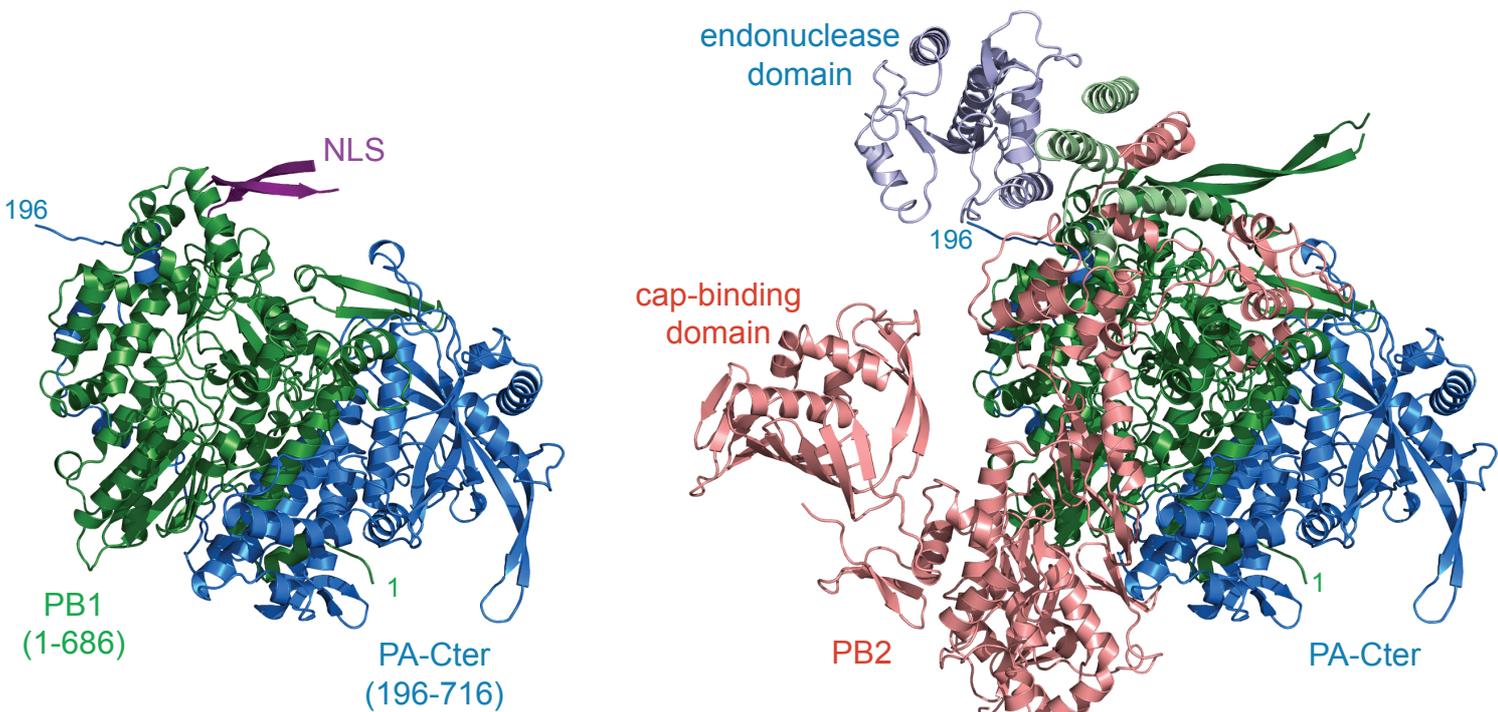
B

Crystal 1	
Beamline	ESRF ID23-eh2
λ (Å)	0.8726
Cell parameters	
a, b, c (Å)	349.2, 349.2, 166.4
Space group	P6 ₁ 2
Resolution (Å) ^a	50 - 7.95 (8.15 - 7.95)
Rmerge ^{a, b}	17.8 (89.7)
I/σ ^a	6 (1.4)
Completeness (%) ^a	97.6 (94.6)
multiplicity ^a	4.5 (5.2)

(a) Values in parentheses are for the highest-resolution shell.

(b) $R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ is the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations of reflection hkl .

C



A/Victoria/3/1975(H3N2)

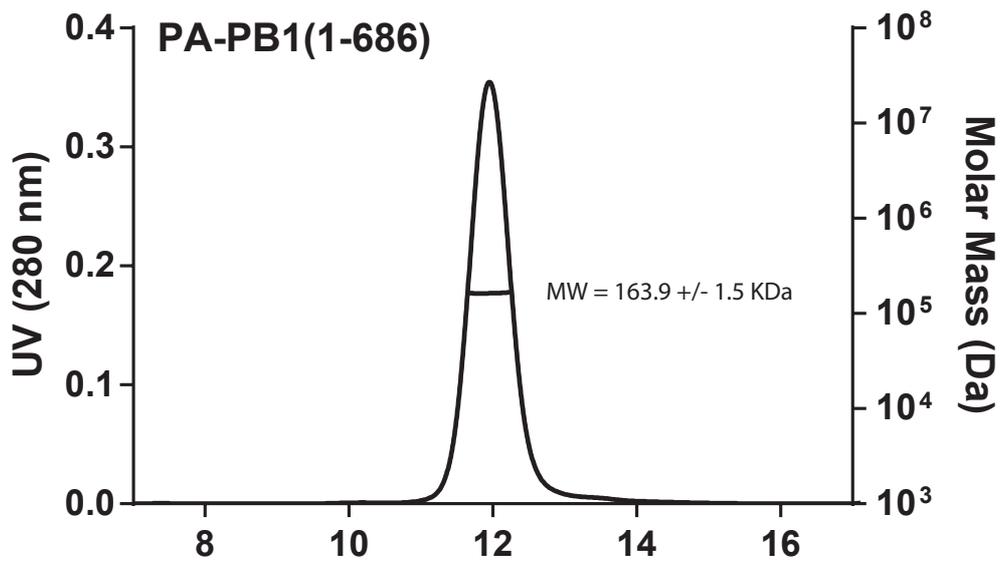
A/little yellow-shouldered bat
/Guatemala/060/2010(H17N10)

Supplementary Figure 6

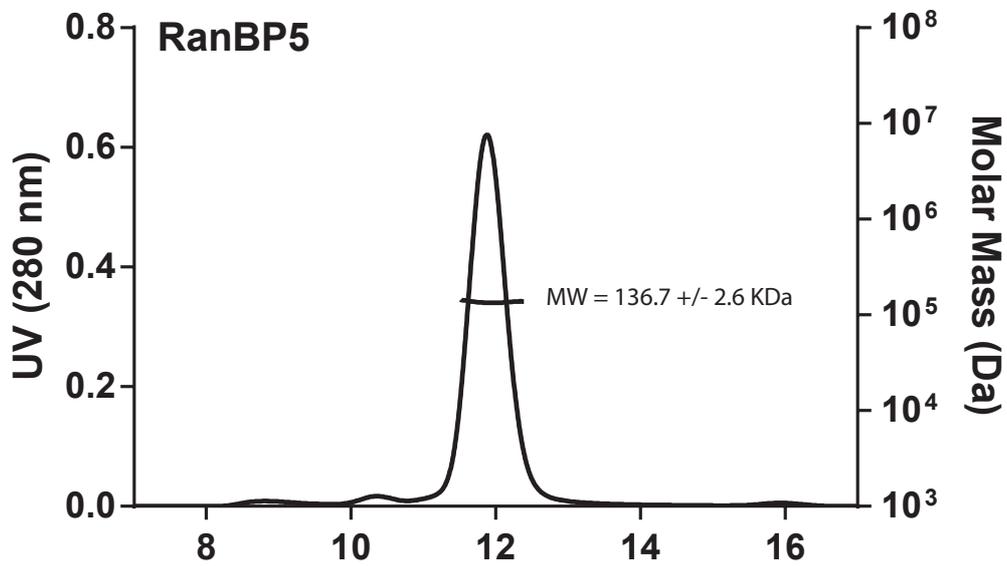
RanBP5 and PA-PB1(1-686) does not form a stable complex when mixed. PA-PB1(1-686) and RanBP5 were separately purified before being mixed in equimolar proportions. The mixture was incubated 1h at 20°C before analysis. Individual SEC-MALLS chromatography runs of **(A)** PA-PB1(1-686), **(B)** RanBP5 and **(C)** PA-PB1(1-686) + RanBP5 were performed in the same buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol) using a S200increase (10/300 GL, GE Healthcare) column. UV absorbance at 280 nm is plotted against the elution volume. MALLS molecular mass estimates are plotted below the curve.

Supplementary Figure 6

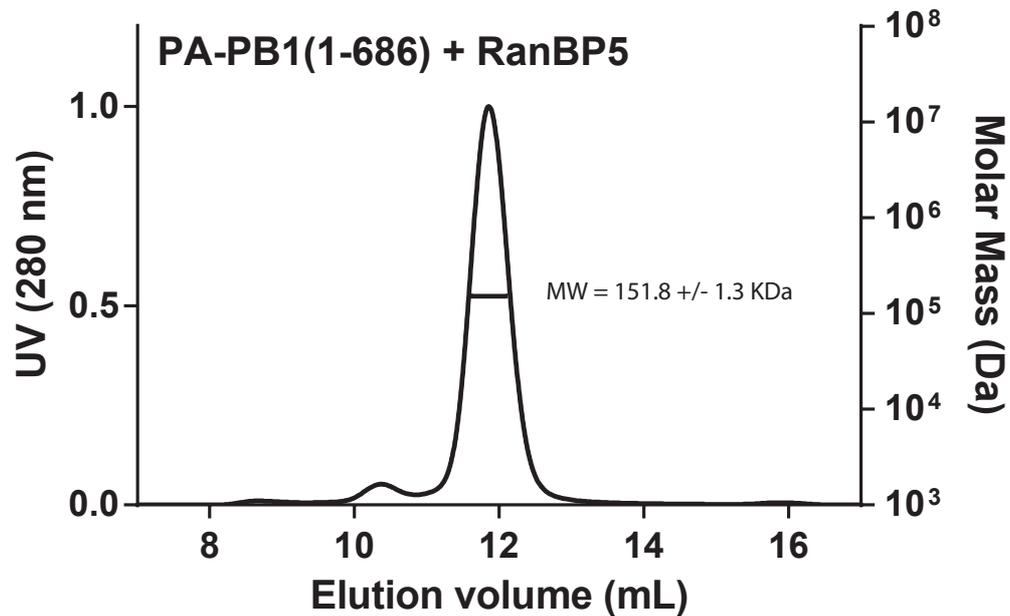
A



B



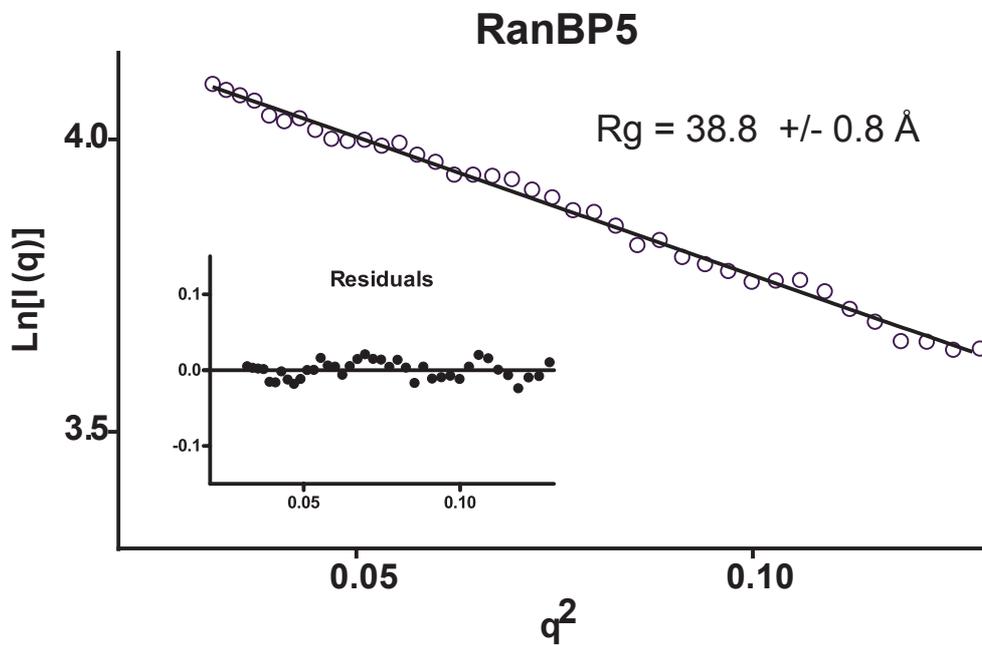
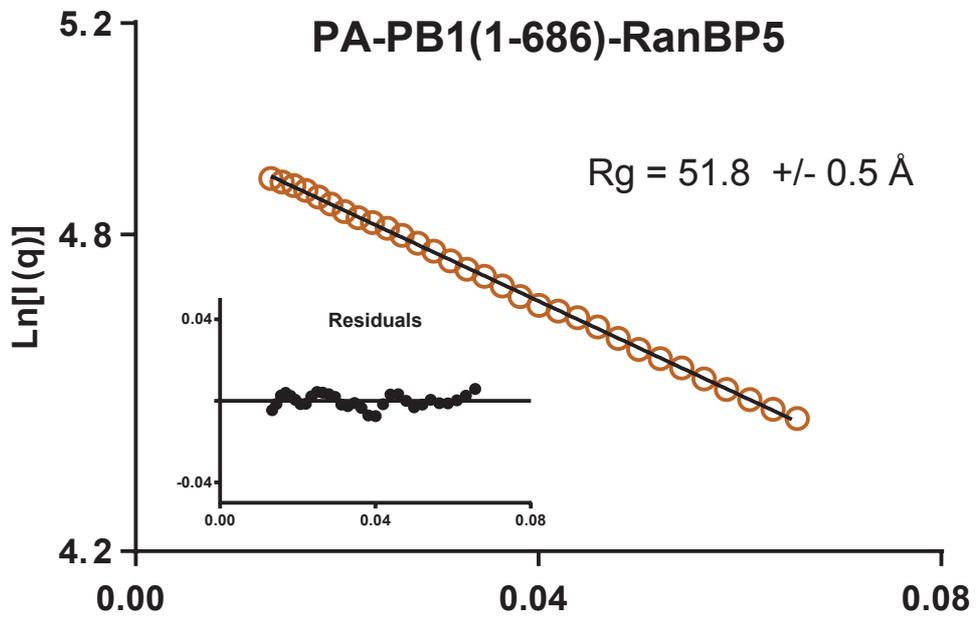
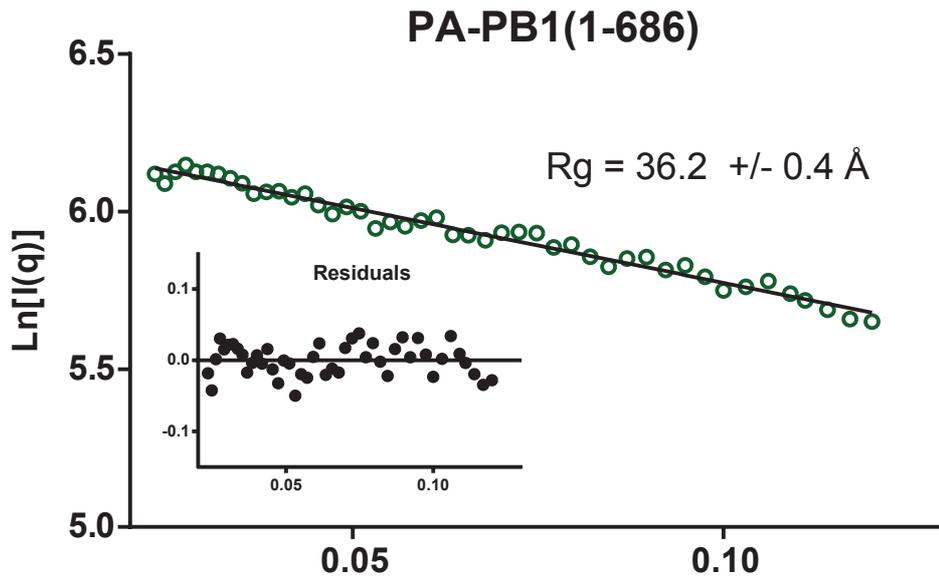
C



Supplementary figure 7

Rg determination by Guinier extrapolation. Guinier plots were calculated on low q regions and are linear for (A) PA-PB1(1-686), (B) PA-PB1(1-686)-RanBP5 and (C) RanBP5.

Supplementary figure 7

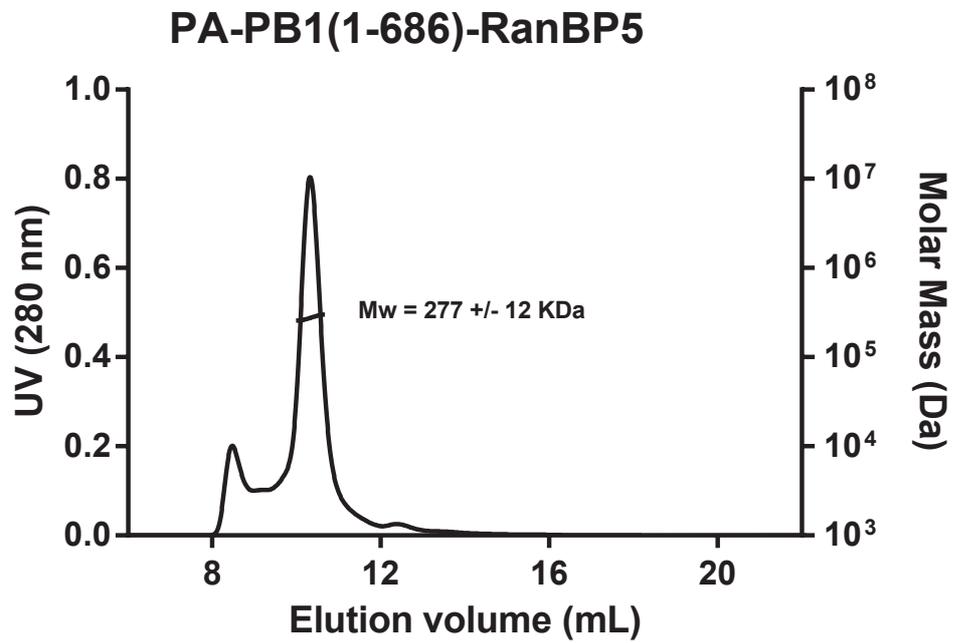


Supplementary Figure 8

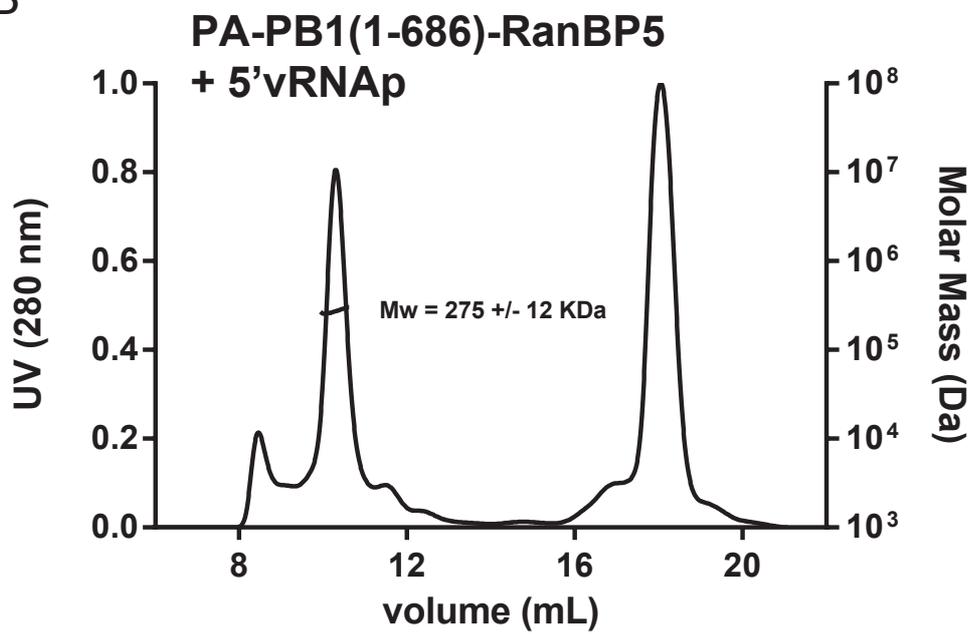
The 5'-vRNAp does not dissociate or bind to the PA-PB1(1-686)-RanBP5 complex. Prior to injection on SEC-MALLS, the RdRp-RanBP5 complex was incubated with a 3 fold molar concentration of 5'-vRNAp for 1h at 20°C. SEC-MALLS chromatograms are shown for both RdRp-RanBP5 without (**A**) and with (**B**) 5'-vRNAp. Both analyses were performed in the same buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 5 mM β -mercaptoethanol).

Supplementary Figure 8

A



B



CHAPTER 4: STRUCTURAL CHARACTERISATION OF RANBP5

Résumé du chapitre 4 en Français

L'assemblage de l'ARN polymérase ARN dépendante du virus de la grippe prend place dans le noyau de la cellule infectée. Ceci implique le transport des sous-unités du cytoplasme où elles sont exprimées vers le noyau où se déroule l'assemblage. Le transport de PA-PB1, un complexe précurseur de la polymérase est encore mal compris mais semble se dérouler grâce au recrutement de RanBP5 qui appartient à la classe des importines- β -3. D'un point de vue structurale, toutes les importines β partagent une architecture commune présentant une série de 23 motifs « HEAT » qui correspondent à des hélices α antiparallèles d'environ 15 à 20 résidus chacune. Leur structure tridimensionnelle tertiaire diffère cependant grandement entre différentes classes, mais aussi pour la même protéine selon son état fonctionnel (lié/dissocié). L'étude structurale de RanBP5 libre est donc présentée dans ce chapitre. L'expression a été mise en place en cellules d'insecte (High Five). La purification se fait en deux étapes majeures : suite à la centrifugation du lysat, une résine nickel suivie d'une chromatographie d'exclusion de taille permet d'obtenir une préparation purifiée de la protéine. Une analyse par MALLS et microscopie électronique montre une bonne homogénéité de l'échantillon, qui se comporte comme un monomère en solution. Ceci est aussi confirmé par SAXS qui permet d'estimer les dimensions de RanBP5. Celles-ci sont comparables avec celles d'un homologue d'importine β : la Kap121 issue de la levure qui partage 25 % d'homologie de séquence. Bien que comparable en taille, la courbe de diffusion est différente indiquant un repliement conformationnel différent.

Une fois purifiée, RanBP5 a aussi permis l'obtention de cristaux, dont la qualité de diffraction (3.3 Å) est comparable à celle d'importines β non liées. Plusieurs jeux de données complets ont pu être collectés et indexés avec un groupe d'espace P_{212121} . Le calcul du coefficient de Matthews estime la présence d'une molécule par unité asymétrique.

La résolution de la structure atomique est néanmoins toujours en cours pour des raisons de difficulté d'obtention des phases. En effet, le remplacement moléculaire s'avère difficile avec des structures homologues à la résolution actuelle. L'obtention des phases par diffraction anormale de la sélénométhionine est aussi compliquée du fait d'une cristallisation moins optimale. La résolution de la structure de RanBP5 impliquerait donc de résoudre sa forme complexée à Ran(GTP) qui permettrait peut-être d'augmenter la stabilité du complexe et donc sa qualité de diffraction.

NOTE OF INTRODUCTION TO CHAPTER 4

As previously described within **chapter 1.4.5**, the polymerase assembly which takes place in the nucleus requires the separate import of both PB2 by importin- α and PA/PB1 by RanBP5 which belongs to the importin- β family. Structural details of the importin- α / PB2 interaction exist and show how PB2 undergoes a conformational change upon binding the importin- α (Kuzuhara et al., 2009; Tarendeau et al., 2007, 2008). However, no structures have yet described the interaction of importin- β -3 (RanBP5) with PA-PB1. To date, only a limited amount of data exist describing the interaction of RanBP5 with the polymerase (Deng et al., 2006b; Hutchinson et al., 2011). Also, RanBP5, although belonging to the importin- β family has not yet been structurally characterised. In order to answer those questions, one of my thesis projects has focused on both an individual characterisation of RanBP5 which is presented in the following chapter, but also its interaction with the polymerase complex which has already been described in **chapter 3**.

4.1 OBJECTIVES

Solving the structure of ligand-free RanBP5 was the first objective in order to understand the different conformational steps RanBP5 undertakes to bind and release PA-PB1. However, any structural or biochemical work requires an essential prerequisite that is the expression of the protein in a soluble form. At first, recombinant expression was attempted in *E. coli* but yielded no soluble protein (results not shown). Insect cell recombinant expression was therefore undertaken.

4.2 EXPRESSION AND PURIFICATION OF RANBP5

RanBP5 is expressed in insect cells as a single protein with the use of the pFastBac-htb vector, which does not enable for the monitoring of CFP (used to estimate expression/stability of other constructs). Also, TEV is not produced in this system and the His-tag is separated from RanBP5 by a TEV cleavage site. As such, only YFP can be monitored during the expression of RanBP5 and this only relates to the baculovirus proliferation. YFP measurements would usually plateau at around 15000 au, 72-96h post-infection which is a third of the values observed for polymerase constructs, indicating a low rate of baculovirus proliferation. Expression of RanBP5 could only be confirmed through western blot analysis of small-scale purifications (not shown). The first large scale purification trials of RanBP5 yielded very little protein after the first nickel resin. Considerable increase in yield was observed only when glycerol (10% v/v) was added to the buffer immediately after lysis. Typical nickel resin purifications will yield around 10-15 mg of protein for 1L of culture and elute with a relatively high purity from the resin (**Figure 30A**). As expected, RanBP5 migrates just below the 130 kDa marker on SDS-PAGE gel which is coherent with its expected molecular weight of 120 kDa. Due to its important molecular weight, no migration difference is observed on gel after the TEV cleavage that is always performed after the nickel resin. Following the cleavage, TEV which carries a his-tag is separated from RanBP5 with a secondary nickel column flow through.

The second step of purification (**Figure 30B**), consisting in a SEC chromatography, only separates very little amounts of associated complexes or contaminants (fractions 23 to 59). Most of RanBP5 is eluted through a very thin and

symmetrical peak (fractions 66 to 86). Although not submitted to any ion exchange column, the end-purified product from the elution peak presents no visible nucleic acid contamination on a UV spectrum curve (220-310 nm, Figure 30C) with a calculated 260/280 ratio of 0,506.

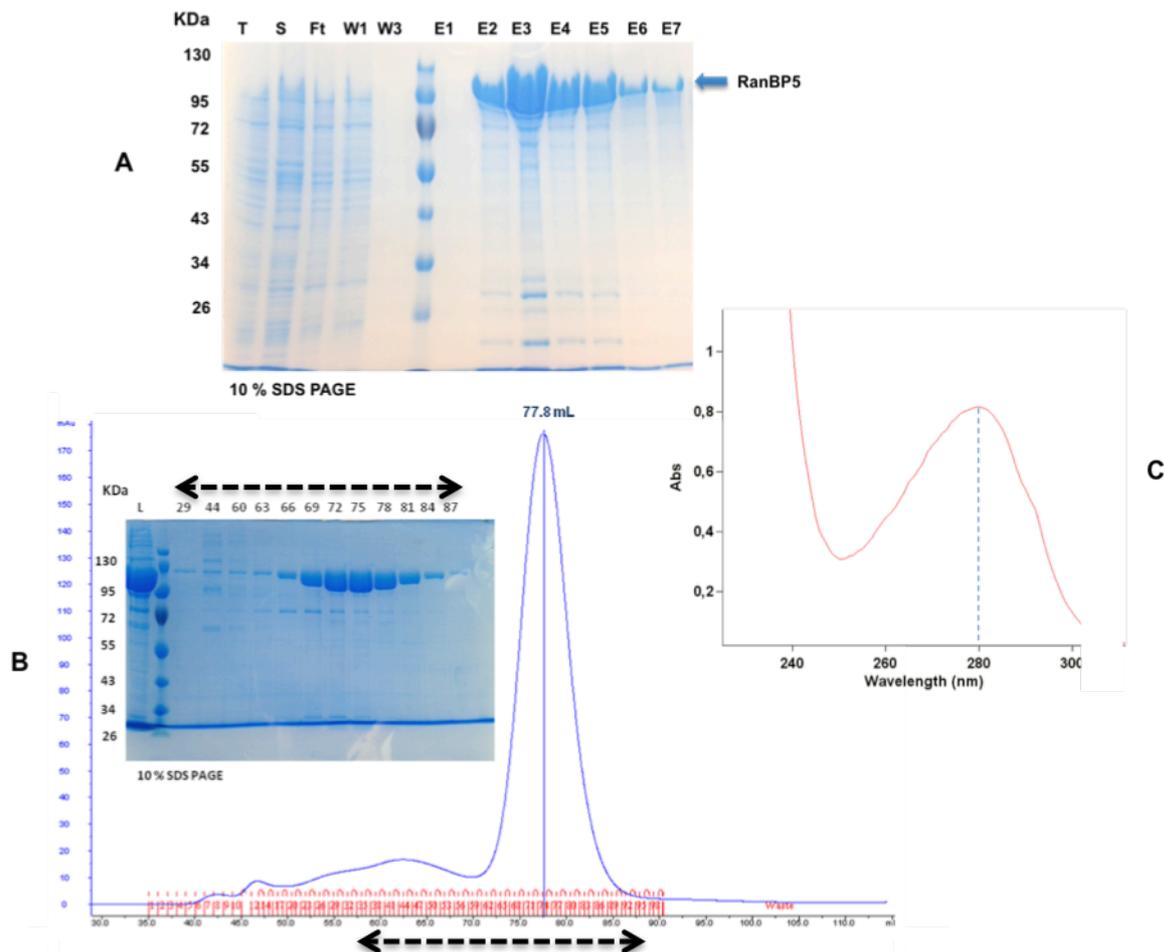


Figure 30: Two-step purification of RanBP5. (A) Nickel resin purification analysed by a 10% SDS PAGE. 8 μ L deposits were used to analyse the total lysate (T), post centrifugation supernatant (S), resin flow through (FT), wash steps (W) and elution fraction (E). (B) Chromatograph and SDS PAGE analysis with fraction numbers of the SEC run using a 120 ml S200 column. (C) UV spectrum performed on SEC peak fractions.

4.3 SAXS CHARACTERISATION

Online SEC-SAXS was performed on purified RanBP5 to evaluate the oligomeric state and also compare a RanBP5 envelope with closely related importin- β structures. When plotting the total scattering intensity with regards to the elution time (Figure 31A), a typical elution peak describes the elution behaviour of RanBP5, as would a UV measurement (280 nm). At every second during the measurement, a Guinier plot was performed to estimate an Rg value. This value, when combined with

the total intensity plot, gives us a real time R_g per point of measurement. It can be observed that prior to the peak, bigger R_g species are eluted, which indicates the partial presence of bigger complexes. However at the peak centre, a stable R_g estimate is observed which indicates a rather monodisperse sample. Based on this plot, stable R_g points (158-163) were pooled and merged using the PRIMUS program to produce a single $\ln I(q)$ data plot from which a Guinier curve could be calculated (Figure 31B). The Guinier estimation places the R_g at 3.89 nm and shows a truly linear curve, consistent with a non-aggregating protein. From this data, a $P(r)$ plot was calculated by GNOM with a fixed D_{max} at around 14 nm (Figure 31C).

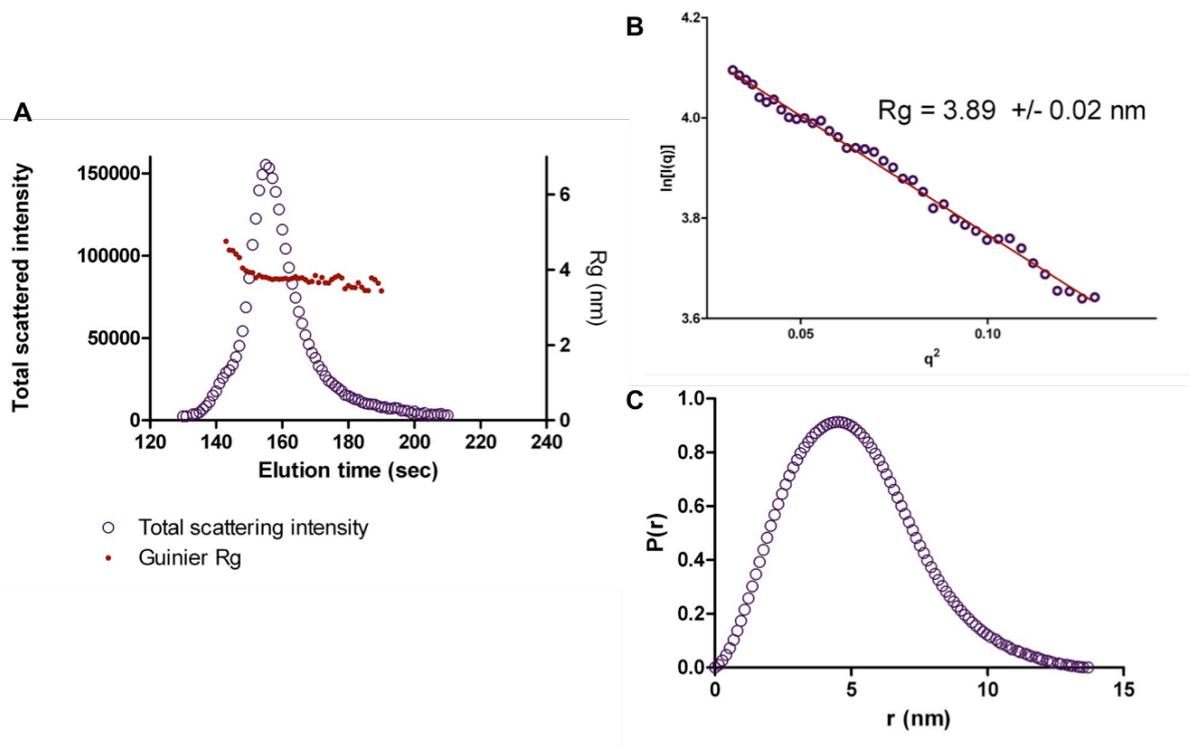


Figure 31: Online SAXS data analysis. (A) SEC elution plot with total scattering intensity of every frame and auto R_g prediction over time. (B) Guinier plot of peak measurements (158-163 sec) with R_g approximation. (C) $P(r)$ size distribution plot using GNOM.

Ab initio modelling was then undertaken using the raw data and $P(r)$ plot to produce an array of 15 models using DAMMIF. Once averaged by DAMAVER, the consistency of the averaged models towards the original $\ln I(q)$ data was checked using DAMMIN (Figure 32A). The model versus raw data fit presents a χ^2 approximation of 0.786 that shows a good fit between both curves and indicates a consistency between the final ab initio model and the original data. When visualising the averaged envelope (Figure 32B) and fitting the closest sequence related structure of a ligand-free importin- β (Kap121, pdb-id: **3w3T**), it appears that the pdb structure can largely fit within the envelope. Structural dimensions are respected as the 3W3T structure presents a maximum length of 12 nm, close to the RanBP5 D_{max} of 14

nm. There is however an important deviation between the Kap121 Crysol and RanBP5 experimental curve suggesting important conformational differences.

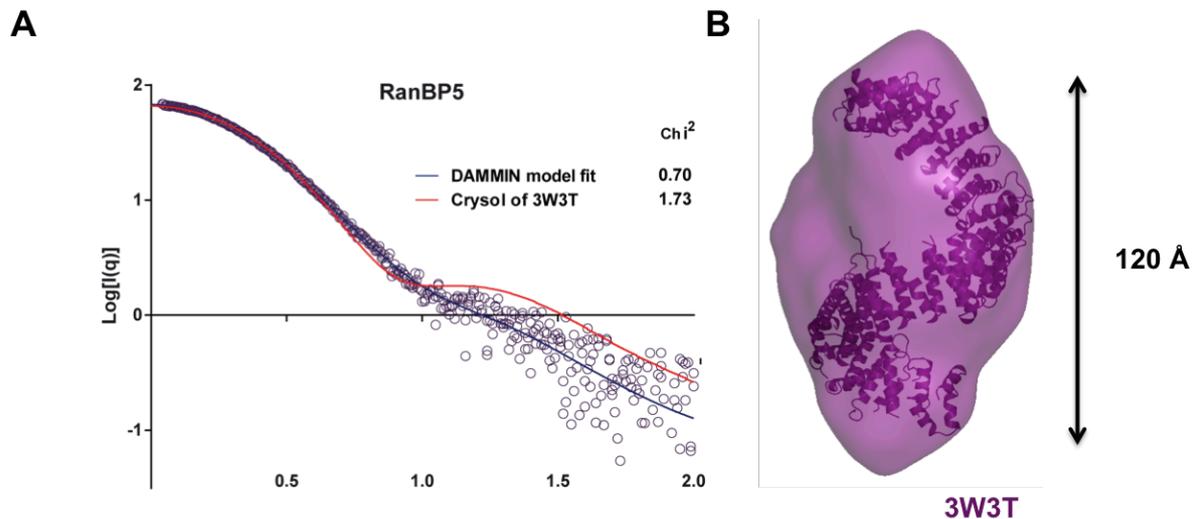


Figure 32: Ab initio modelling of RanBP5. (A) DAMMIN curve (in blue) using the damstart model as the starting envelope and CRY SOL curve (in red) of the Kap121 homologue (pdb id: 3W3T). Both curves are fitted against the raw $\ln I(q)$ plot (in violet) from frames 158 to 163. (B) DAMAVER envelope (purple) with fitted importin- β homologue (pdb id: 3W3T).

4.4 INITIAL CHARACTERISATION AND CRYSTALLISATION

As previously mentioned, in order to crystallise a given protein, one must produce it in high quantities, at high levels of purity and in a stable form.

To assess the level of purity in SEC peak fractions, EM negative staining grids were prepared in order to estimate the homogeneity of the sample. Both Sodium SilicoTungstate (SST) and Uranium Acetate (UrAc) were tested as contrasting agents; in the end, UrAc was used with RanBP5. As seen on an EM picture (Figure 33A) using a 40000X magnification, the sample is non-aggregated in clusters and seems homogenous in the distribution of particle sizes.

To estimate the stability of the protein once concentrated to 8 mg.ml^{-1} , a thermal shift assay (TSA) run was performed (Figure 33B). It shows one single melting stage with an approximate T_m at 56°C , which is rather stable for a protein of this size and well above the recommended 25°C temperature difference (ΔT) with the incubation temperature for crystallisation (Dupeux et al., 2011).

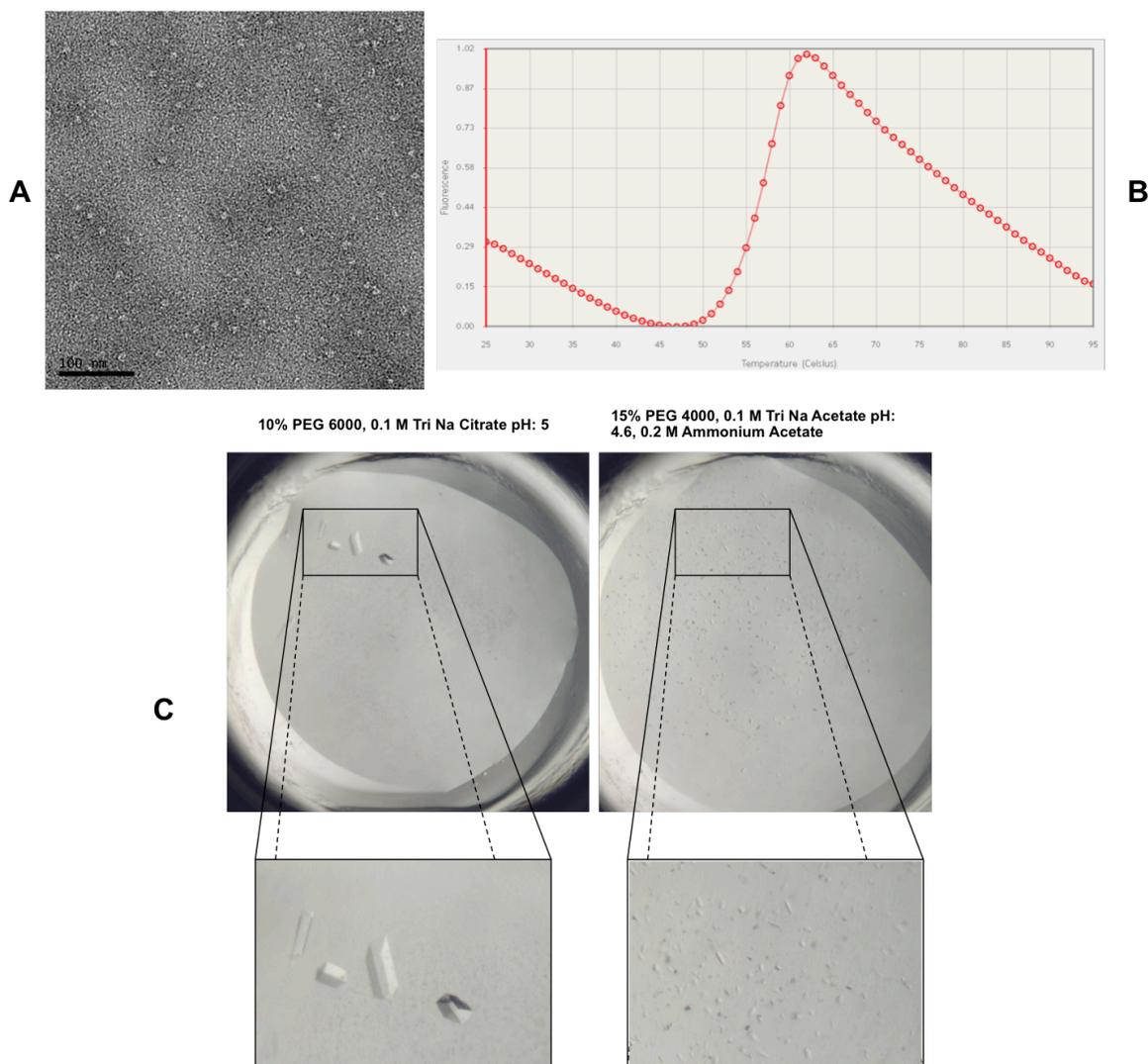


Figure 33: Quality control steps required for crystallisation. (A) Negative stain EM image taken with the Technai 12 at 40 000x, using UrAc as contrasting agent. (B) TSA result plotting the sypro fluorescence intensity as a function of temperature, the T_m is estimated as the midway point between the maximum and minimum fluorescence values. (C) Initial crystal hits 2 days after plate setup. Drops contain 100 nL of protein solution and 100 nL of precipitant. Pictures were taken by the Formulatrix imaging robot.

Based on these two previous results, crystallisation screens were undertaken using the EMBL high-throughput platform (HTX). Six initial screens were performed both at 20 °C and 4 °C covering a large range of salts, pH, pegs and other organic precipitants. In total 1152 different conditions were initially screened. Most either produced precipitated drops or clear drops. At 4 °C exclusively, two very comparable conditions yielded positive “hits”. One condition (10% PEG 6000, 0,1M Tri Na-Citrate pH 5) produced 4 small but sharp looking crystals while the other (15% PEG 4000, 0.1M Tri Na Acetate pH: 4.6, 0.2M Ammonium Acetate) produced a “carpet” of microcrystals (Figure 33C).

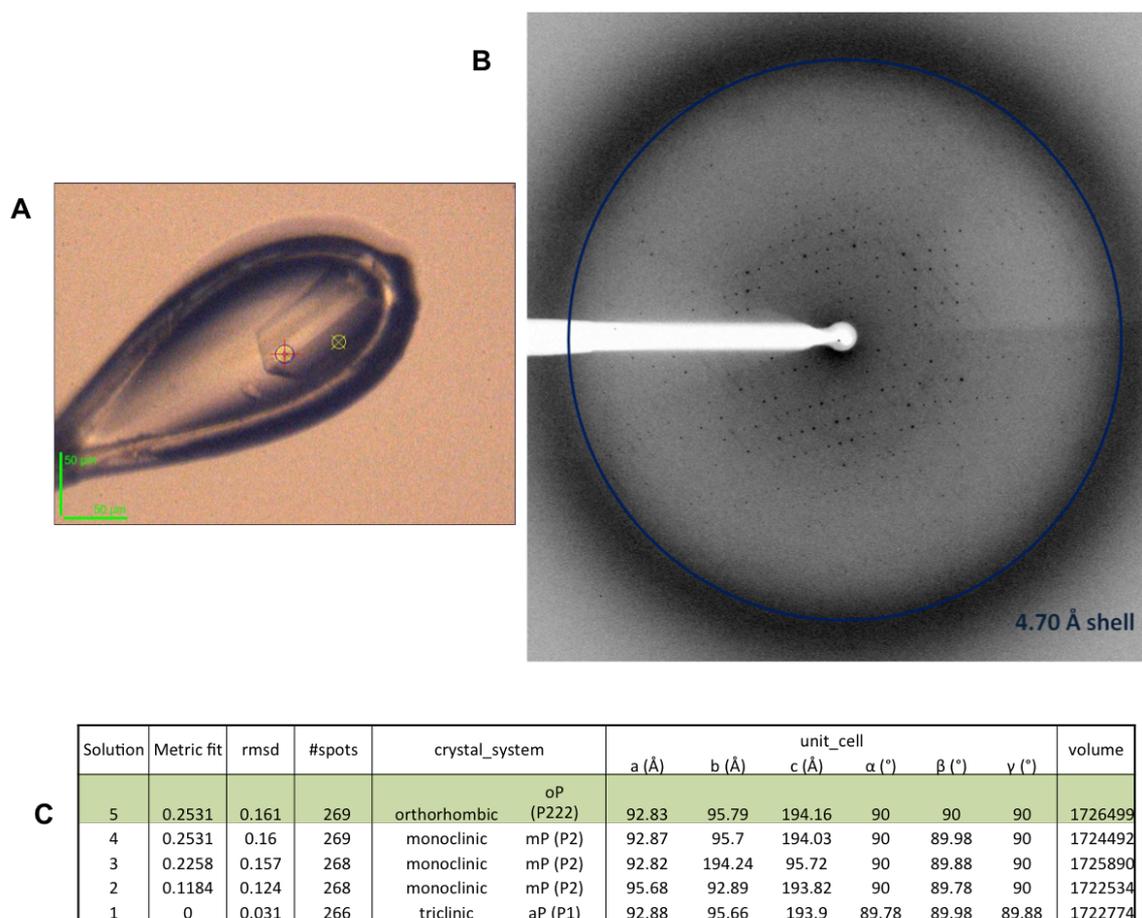


Figure 34: RanBP5 crystal characterisation on ID 30A-3 (ESRF, Grenoble). (A) Cryo-loop and crystal snapshot with the beam position indicated by a red cross, green scale bars represent 50 μm . (B) Image collected on a Mar225 detector. (C) Indexing result produced by the EDNA processing server, showing the most probable space groups.

The presence of crystals within crystallisation drops is a good start, however in some cases false results can appear because of salt crystal growth. In this case the crystals appeared quite quickly but increased in size over several days, which is more common for protein crystals. Also, upon manual observation no strong polarisation of light was observed which is usually the case with salt crystals. Three crystals were therefore mounted on cryo loops and cryo protected using 25 % glycerol. Although very small with a size below 100 μm in length (Figure 34A), one of the crystals diffracted to 4.3 \AA with visible spots at around 4.7 \AA (Figure 34B). The diffraction pattern is typical of a protein diffraction pattern as there are numerous spots, some of which are very close together. A salt crystal diffraction pattern usually has very little spots of very strong intensity. This is explained by the unit cell size, small unit cells such as those of salt molecules diffract with big distances between diffraction spots (reciprocal space), inversely bigger unit cells have shorter distances between the spots. The diffraction geometry and spot shape is also quite “clean” indicating no visible splitting of the crystal. Computer indexation was possible using

two images 90° apart and predicts the space group symmetry to be a P222 (Orthorhombic). As seen on the prediction result (Figure 34C), lower symmetry space groups such as P2 (Monoclinic) or P1 (Triclinic) also fit. However, the general rule is to attribute the highest symmetry space group while keeping a low penalty measurement (rmsd).

Although attempted, a complete data collection was not possible due to inherent difficulties in correctly centring the crystal on the X-ray beam and also due to its size, radiation damage quickly appeared. This experiment however provided a cell dimension estimate, which also enables for the calculation of Mathews coefficient. Computer calculations use the unit cell dimensions and theoretical molecular weight of the protein in order to calculate a water content percentage. This resulted in predicting the unit cell to have one molecule in the asymmetric unit with a water content of 60%, which is coherent with the observed low resolution.

4.5 CRYSTAL OPTIMISATION

With the objective of producing bigger and better diffracting crystals, several refined crystallisation screens were designed and prepared with a formulator robot (Formulatrix). As several crystal hits had appeared with pegs ranging from 4000 to 8000, a rational screen of PEG concentration/type versus pH and with or without ammonium acetate was designed (Figure 35). This screen enables to refine key conditions for proper crystallisation of RanBp5. First it demonstrates that RanBp5 can crystallise in all three PEG types (4-8000). However, sharp crystals are mostly obtained in 10-12 % PEG 6000 (Figure 35) as the other conditions generate curved crystals (PEG 4000) or smaller and thinner crystals (PEG 8000). The other critical parameter is pH as most crystals grow exclusively in pH 4.5 using Tri Sodium Citrate (results not shown). Finally, ammonium acetate tends to limit nucleation and therefore increases the chances for bigger crystals but also produces visually sharper crystals. Although the listed conditions above contribute to producing the best looking crystals, all other conditions were tested for diffraction as in some cases the best looking crystals are not necessarily the best diffracting ones. However, as a general rule of thumb for RanBP5 crystals, the nicest looking crystals generally produce the nicest diffraction.

Another important parameter controlling the nucleation was protein concentration. 3 to 5 mg/ml of protein were used in order to have a low nucleation and therefore produce bigger crystals.

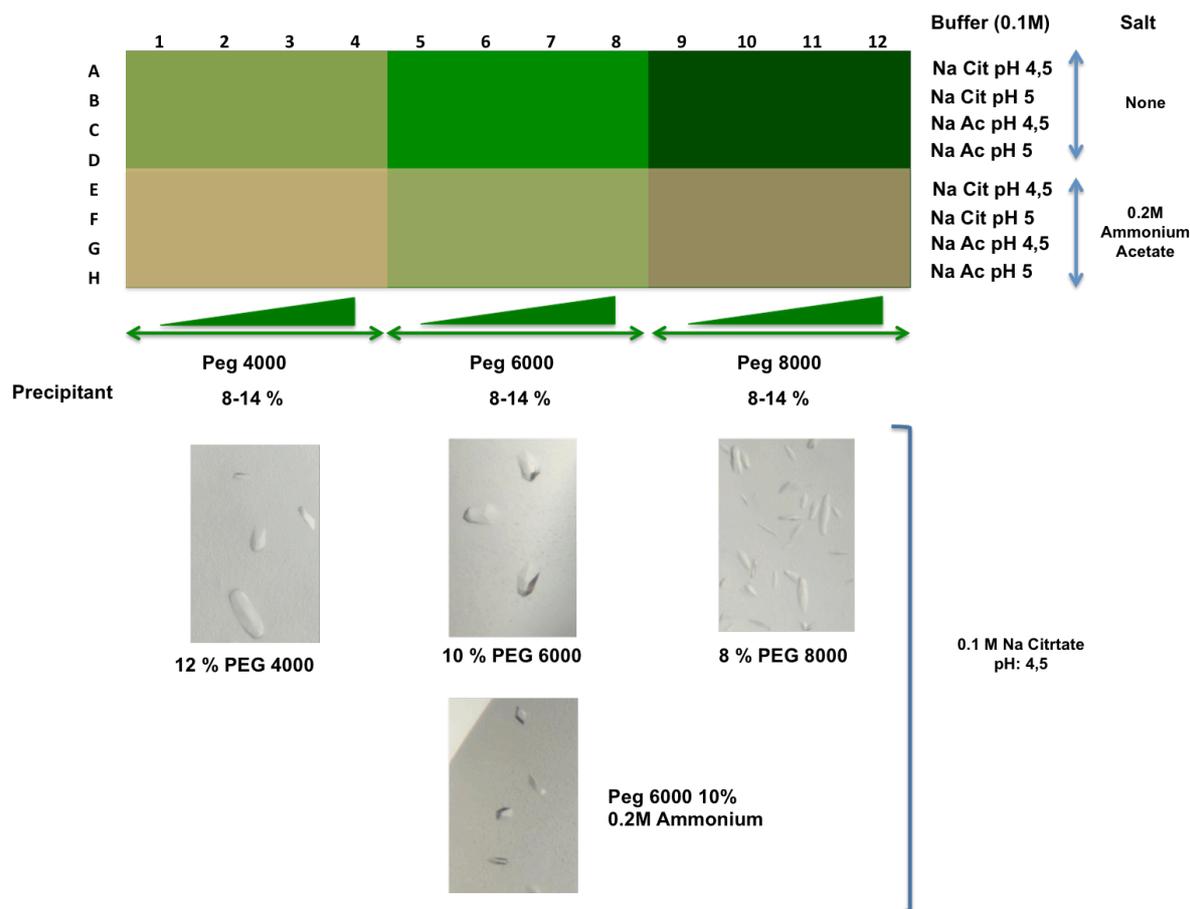


Figure 35: PEG and pH screening to produce RanBp5 crystals. (Above) Schematic view of the screening grid, PEG 4000, 6000 and 8000 are individually screened against two pH conditions with either sodium citrate or sodium acetate. Ammonium acetate complements half the screen. (Below) Positive crystal growth, all images are at the same scale.

From the initial optimised condition (12 % PEG 6000, 0.1M Na citrate pH: 4.5, 0.2M ammonium acetate), numerous additive screens were also made in the attempt to increase crystal size or diffracting resolution. To date, no additive chemicals were found to optimise the process. Macro seeding was also undertaken. Macro seeding is a technique which consists in using previously grown crystals reduced to fragments to initiate nucleation. This method can present several potentially beneficial effects. It can first initiate crystal growth in a condition that will not bring the protein to a supersaturated state, for instance a lower PEG content (8% PEG 6000). This in turn enables a perfect control of nucleation as only seeded crystals will grow. The second possible advantage is the potential for a cleaner crystal growth in the context where a protein can crystallise in several forms which can compete each other during crystal growth and contribute to twinning.

4.6 DATA COLLECTION

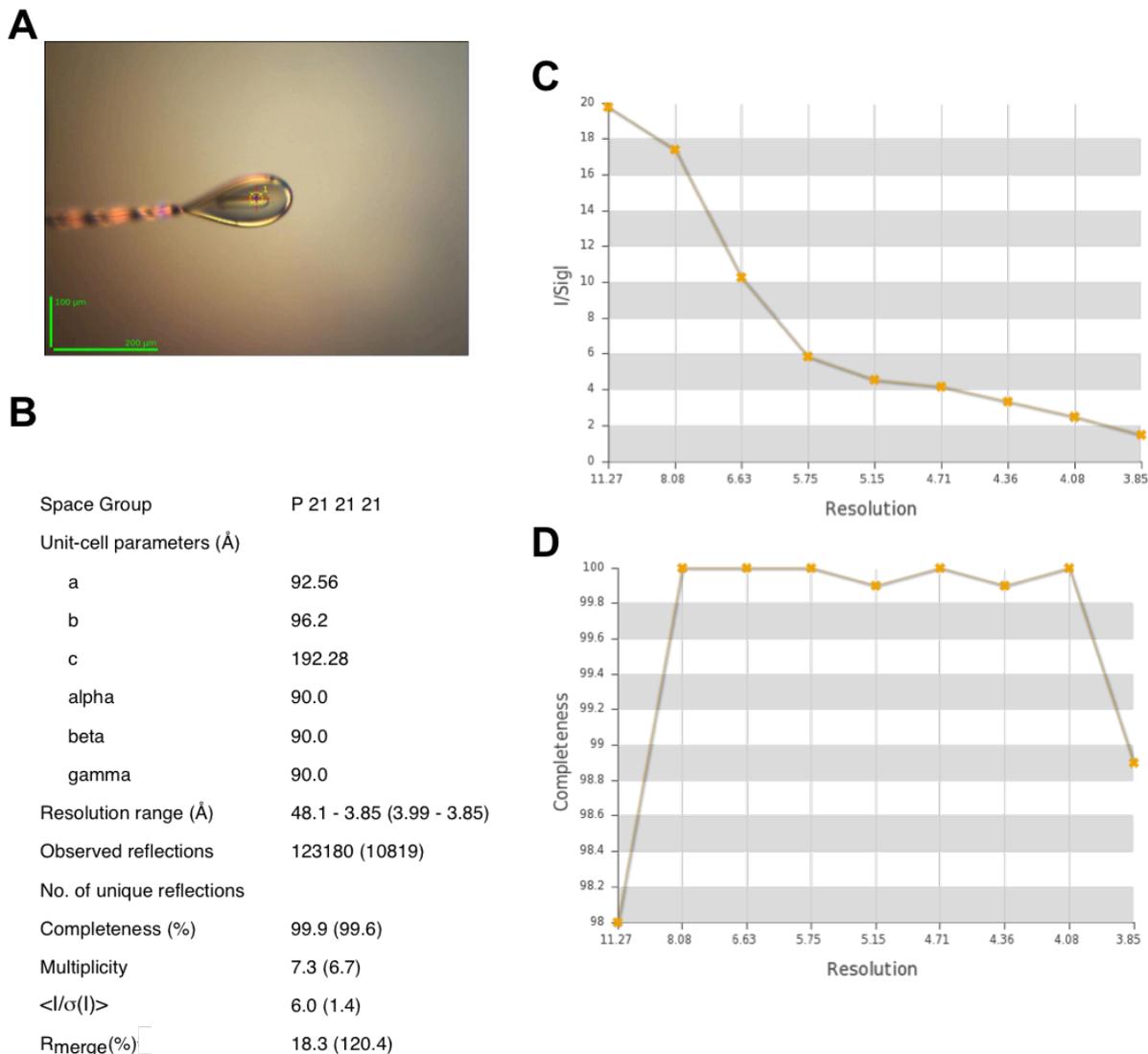


Figure 36: Initial dataset acquisition on ID 29 (ESRF, Grenoble). (A) Magnified view of the mounted crystal, the horizontal green scale represents 200 μm . (B) XDS indexing and integration report produced by the EDNA server. Bracket values describe the lower resolution shell statistics. (C) $I/\sigma(I)$ plotted against the observation resolution. (D) Completeness plotted against the observation resolution.

Following crystallisation and optimisation, a complete dataset was collected on the initial crystallisation condition. Although twice as big as the original crystals (Figure 36A), the outer edge resolution was quite low (4 Å) implicating once again a very high solvent content. Diffraction spots were so weak and thin that they were not easily visible on Pilatus detector images (not shown). However, indexing programs were able to evaluate the diffraction pattern. These programs enable an estimation of the cut-off resolution at around 3.9 Å, which is the point where the $I/\sigma(I)$ (signal to noise ratio) goes below a value of 2 (Figure 36C). As a result, data completeness only

extends to about the same resolution range before suddenly dropping (Figure 36D). Although indexed by the program, the outer shell (4.0 to 3.8 Å) or high-resolution shell contributes to increasing the R_{sym} figure of merit. It therefore reduces the quality of the dataset and must be deleted from the final integration stage. However, the overall dataset (from 50 to 3.9 Å) has a good completeness and multiplicity (Figure 36B). Cell parameters differ slightly from the initial parameters as they are obtained from a refined indexing on all the images.

Following indexing, the entire dataset was integrated and scaled using XDS. As beta importins all adopt a heat repeat structure. Molecular replacement (MR) was attempted on multiple homologues by the Mr.Bump software (Keegan and Winn, 2007). This program finds the closest sequence homologues on the pdb and uses Phaser (McCoy et al., 2007) as a molecular replacement program and several downstream refinement programs such as re_{mac} (Murshudov et al., 1997). The closest sequence homologue (25% sequence identity) is Kap121p which is a yeast β-3 importin crystallised in different forms (Kobayashi and Matsuura, 2013). When using full β-importin structures, no valid phasing solutions were found. This was also true when using truncated search models although the figure of merit score improved (Z score in phaser) in particular when using N-terminal truncation of the Kap121p models, which implies a certain similarity with the search model. In any case, the incapability of phasing using MR can be attributed to important structural differences between the homologue search model and the structure of RanBP5, but also to the dataset resolution. Indeed, resolutions beyond 3 Å are difficult to phase through MR as Patterson maps become very noisy. Resolution improvement is therefore required in order to attempt further MR trials.

4.7 INCREASING RESOLUTION THROUGH *IN-SITU* PROTEOLYSIS

In some cases, slight degradations of unstructured loops or terminal tails can improve overall resolution by improving the crystal packing. This can be achieved through the direct addition of protease during the crystal drop setup and is known as *in-situ* proteolysis (Dong et al., 2007; Wernimont and Edwards, 2009). In order to assess the effect of small quantities of proteases, several kinetic proteolysis were made and analysed by SDS PAGE. Four commonly used proteases in protein crystallography were tested in the same conditions. Every protease was added at a concentration of 1/500th (w/w) to a sample of RanBP5 at 7mg/ml (concentration of crystallisation) and placed at 37 °C. At given time intervals (5 to 60 minutes), the reaction was stopped by adding SDS loading buffer and heating the tubes 5 min at

95 °C. As seen on the SDS-PAGE (Figure 37A), papain, which is a cysteine protease, over-degrades RanBP5 very quickly and therefore is not suited for in situ proteolysis. Elastase, which belongs to the serine protease class (targets A, G and V residues preferentially), generates several degradations below the 72 kDa, 55 kDa and 26 kDa. However, none of those remain truly stable during the process. Trypsin and chymotrypsin were tested as well; these are also serine class inhibitors targeting K / R residues (trypsin) and Y / W / F (chymotrypsin) respectively. These two enzymes are the most commonly used enzymes for *in situ* proteolysis as they weakly degrade proteins. This is visible on the SDS PAGE analysis. Chymotrypsin generates a small degradation below the RanBP5 native band, which could correspond to a small terminal truncation. Trypsin also behaves in a comparable fashion but additionally generates another band at around 72 kDa.

Based on the limited proteolysis results, crystal drops were set up with both trypsin and chymotrypsin. As *in situ* proteolysis can drastically modify the crystallising behaviour of a given protein, broad condition screening was undertaken. Carpet crystal growth was observed in the usual condition (10 % PEG 6000, Na citrate pH: 4.5) when adding chymotrypsin and, additionally, in a new condition (20% PEG 3350, 0.1 M Ammonium Formate) where a single small crystal was grown (Figure 37B). Although very small, the crystal was “fished” for diffraction testing. Differently from previous crystals which had all been flash frozen with either glycerol or ethylene glycol, this crystal was flash frozen in 40% PEG 3350 as PEG 3350 is the main precipitant of this condition.

Upon characterisation on the beamline (ID 29), clean diffraction spots could be observed over 3.5 Å (Figure 38A), which was quite an important increase when compared to native RanBP5 crystals. Upon indexing/integration (Figure 38B), one of the first observations is the fact that the space group remains the same (orthorhombic, P_{212121}) with very slight differences in the cell A, B, C dimensions (native: 92.6 / 96.2 / 192.3; with chymotrypsin: 91.2 / 95.4 / 189.2). The chymotrypsin grown RanBP5 crystal has a slightly smaller unit cell. This indicates potentially more efficient packing that is either the result of in situ proteolysis, or of the higher PEG content within the crystallisation condition. The overall dataset (2000 images) is of a much higher quality than those previously acquired. Reflections were recorded up to 3.5-3.4 Å with completeness close to 100 % in all the resolution shells below 5.5 Å (Figure 38C). The dataset resolution limit is around 3.35 Å where the I/σ drops below a value of 2 (Figure 38D).

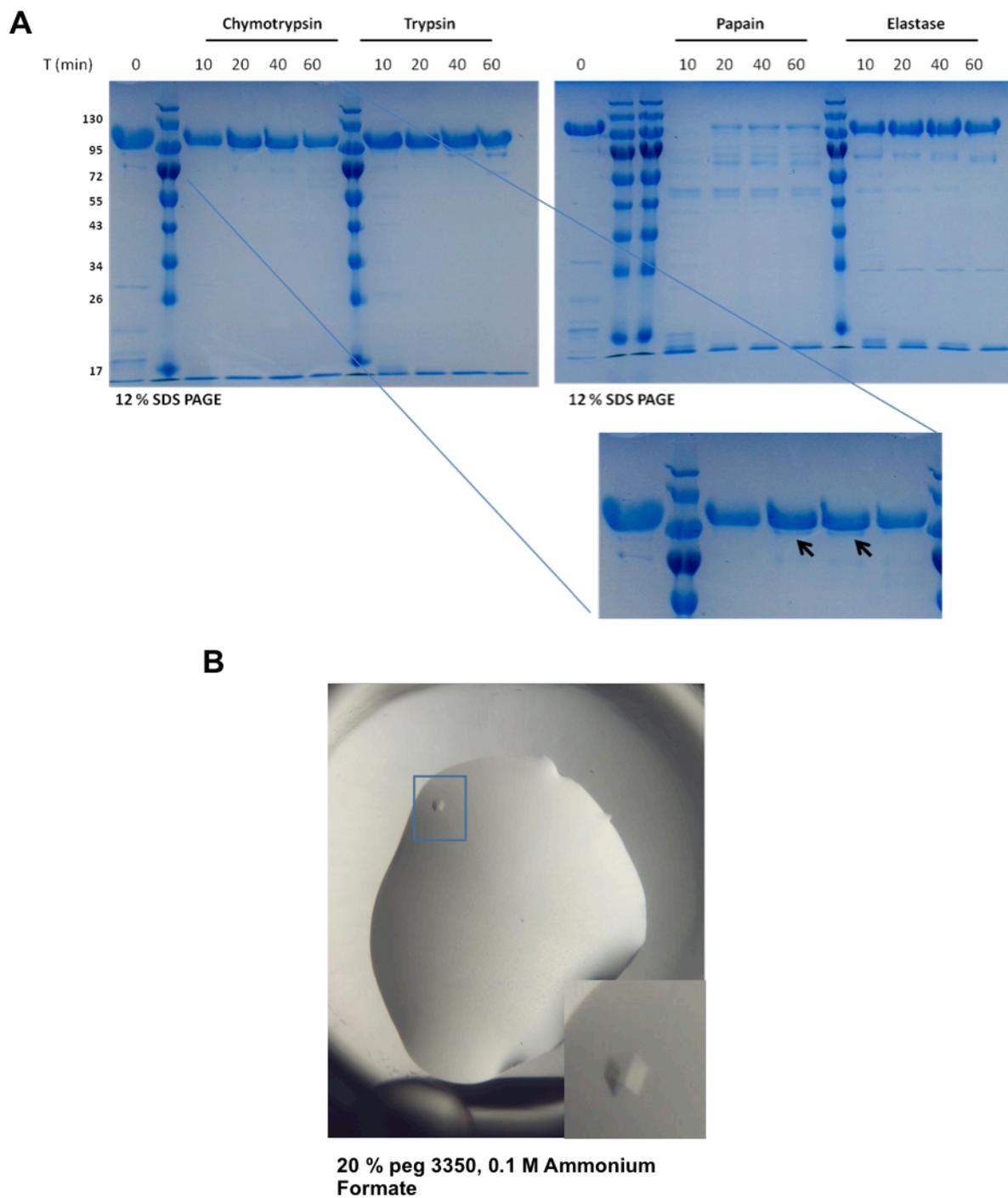


Figure 37: *In situ* proteolysis of RanBP5. (A) Kinetic proteolysis realised with chymotrypsin, trypsin, papain and elastase at a constant ratio of 1/500th (w/w) on RanBP5 at 37 °C over the course of 1h. Black arrows highlight transient degradation. (B) Magnified picture of an *in situ* proteolysis grown RanBP5 crystal.

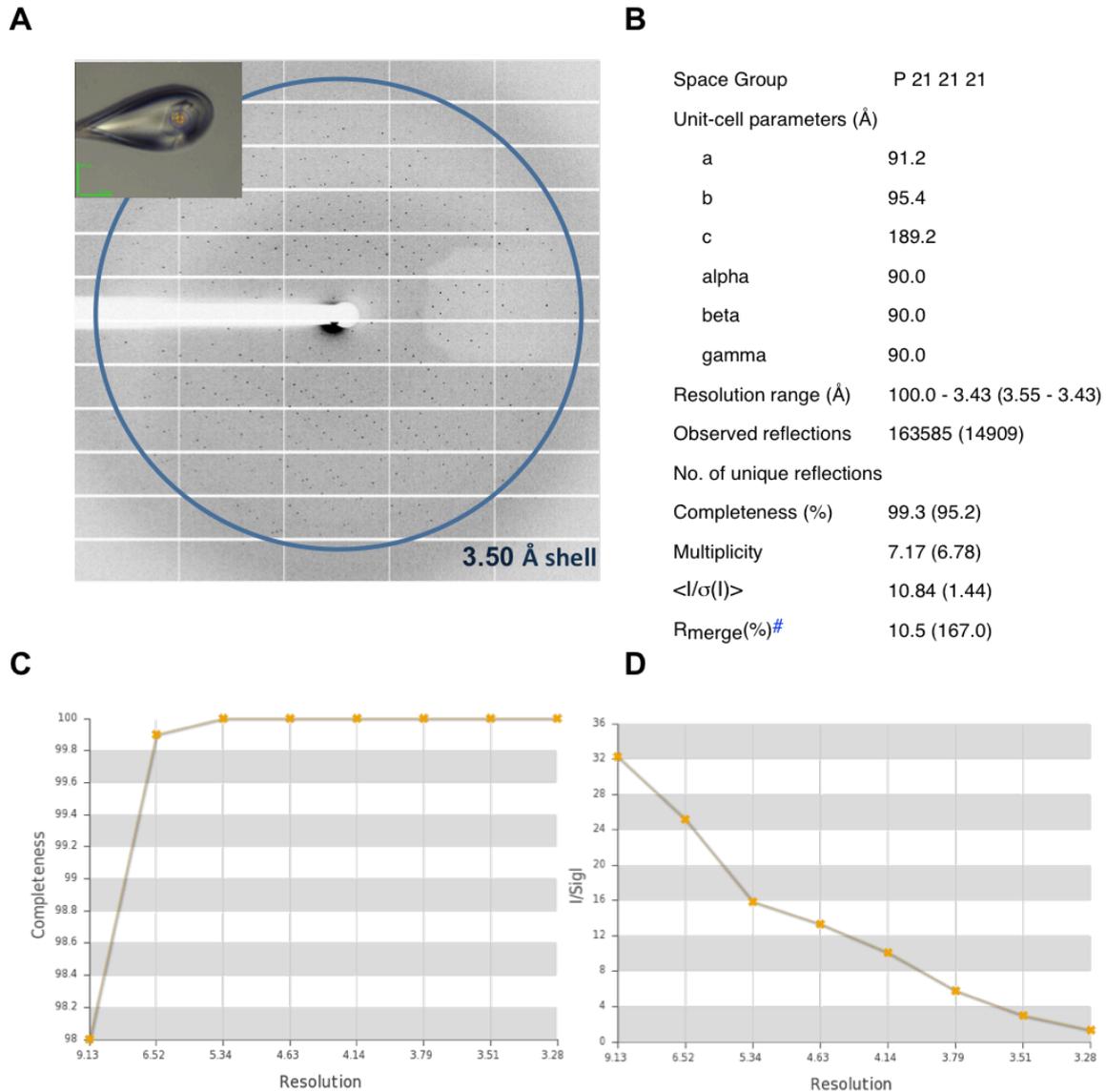


Figure 38: Data collection on RanBP5 crystal grown with in situ proteolysis (ID 29, ESRF Grenoble). (A) Diffraction image with a 3,5 Å resolution shell collected on a Pillatus 6M detector. The mounted crystal is shown on the upper left, green scale bars represent 50 μm. (B) Indexing and integration statistics from XDS. Bracket values describe the lower resolution shell statistics. (C) Completeness versus resolution plot. (D) I/σ versus resolution plot.

Although considerably better than previous diffraction experiments, MR attempts with the chymotrypsin crystal fell short of finding a solution to the phase problem. Reproduction of the crystallisation has also been a considerable problem as the condition in PEG 3350, which produced the crystal, came from a commercial screen and could not be reproduced when hand made. The contribution of chymotrypsin to the overall crystallisation also asks the question of whether or not the RanBP5 crystals are partial or important digestions. In order to answer this question, a handmade drop with a “carpet” of small crystals had its protein/precipitant mother liquor removed, which only left the crystals stuck to the

glass surface. Those crystals were washed twice with pure precipitant then re-suspended in the protein buffer. An SDS-PAGE gel was run with the sample and a freshly produced RanBP5 prep sample concentrated at 5 mg/ml. The gel was revealed using a silver nitrate kit (Biorad). As visible on the gel (Figure 39), the re-suspended crystals are exclusively composed of full RanBP5 migrating at around 120 kDa. This result indicates that degradation may only occur partially on the C-terminus or N-terminus or that the protease degrades other elements within the protein prep, indirectly enhancing crystallogenes.

Based on the following results, the only solution to solving the phase problem is to use experimental phasing methods.

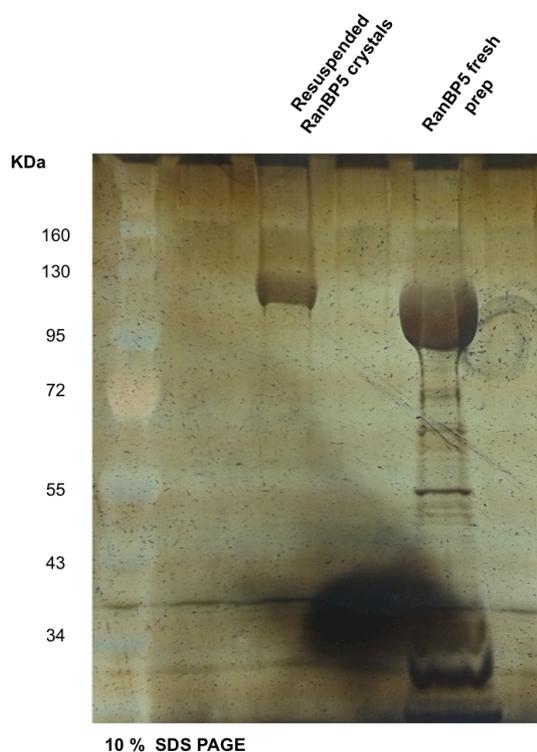


Figure 39: SDS PAGE analysis of RanBP5 crystal content using a silver nitrate stain.

4.8 EXPERIMENTAL PHASING OF RANBP5

Experimental phasing can overcome in some cases the phase problem, especially when no suitable MR search model can be found. Resolution is not as much of a limiting factor when phasing experimentally, as phases can be determined for lower resolutions. The important factors are high multiplicity and absolute completeness within the dataset, which means having intense and clean diffraction within all resolution shells with a minimum of radiation damage.

4.8.1 SAD PHASING OF RANBP5

Single Anomalous Dispersion as previously described is the most commonly used experimental phasing method and relies on the incorporation of selenomethionines instead of methionines.

Mostly used for prokaryotic protein expression, SAD can also be used for insect cell expressed proteins but requires a very strict expression protocol (Barton et al., 2006; Cronin et al., 2007). Insect cells are grown in their normal media (Xpress Five, Life Technologies), then infected with the baculovirus and incubated for 10 to 12h enabling baculovirus integration. Following this stage, cultures are gently centrifuged and re-suspended in methionine depleted media (Expression Systems) before complementation with selenomethionine. Cultures are then monitored as usual until harvest. Selenomethionine is slightly toxic for insect cells, thus reducing the overall expression yields by a factor of 2. Purification was undertaken using the exact same protocol as for the production of native crystals. Selenomethionine crystals were produced in PEG 6000 screens in the presence and absence of chymotrypsin. In all cases, crystals remained quite small and some appeared twinned. All were cryo-mounted using 40 % PEG 3350 as a cryo-protectant.

When tested on the beamline (ID 29 and ID 23-1) most crystals diffracted very poorly, if at all. Out of 15 tested crystals, 3 diffracted to a reasonable resolution (4 Å) for a full data collection. One full dataset was collected on a slightly twinned and broken crystal (Figure 40A). Although reflections have been indexed up to 3.8 Å in resolution, most are visible at around 5 Å on the images. Another element is the apparent slight twinning of diffraction reflections which does contribute indexing errors. Overall a complete dataset was collected and integrated with the same space group parameters obtained for PEG 6000 crystals. Good completeness and redundancy was observed over the entire dataset (Figure 40B).

Prior to collecting the reflection images, an energy scan was undertaken on the crystal to confirm the presence of selenium. In practice a phosphor screen detector is inserted after the crystal to measure the x-ray fluorescence at the excitation energy of selenium (12.620 to 12.740 KeV). As seen on the fluorescence scan produced by the CHOOCH (Evans and Pettifer, 2001) program (Figure 40C), a fluorescence peak is obtained at 12.6623 KeV, which confirms the presence of selenium within the crystal. From this scan an anomalous signal difference can be predicted for f' and f'' . To maximise the anomalous signal, and therefore the

amplitude difference between f' and f'' , the data collection was performed at the peak inflexion (12.662 KeV).

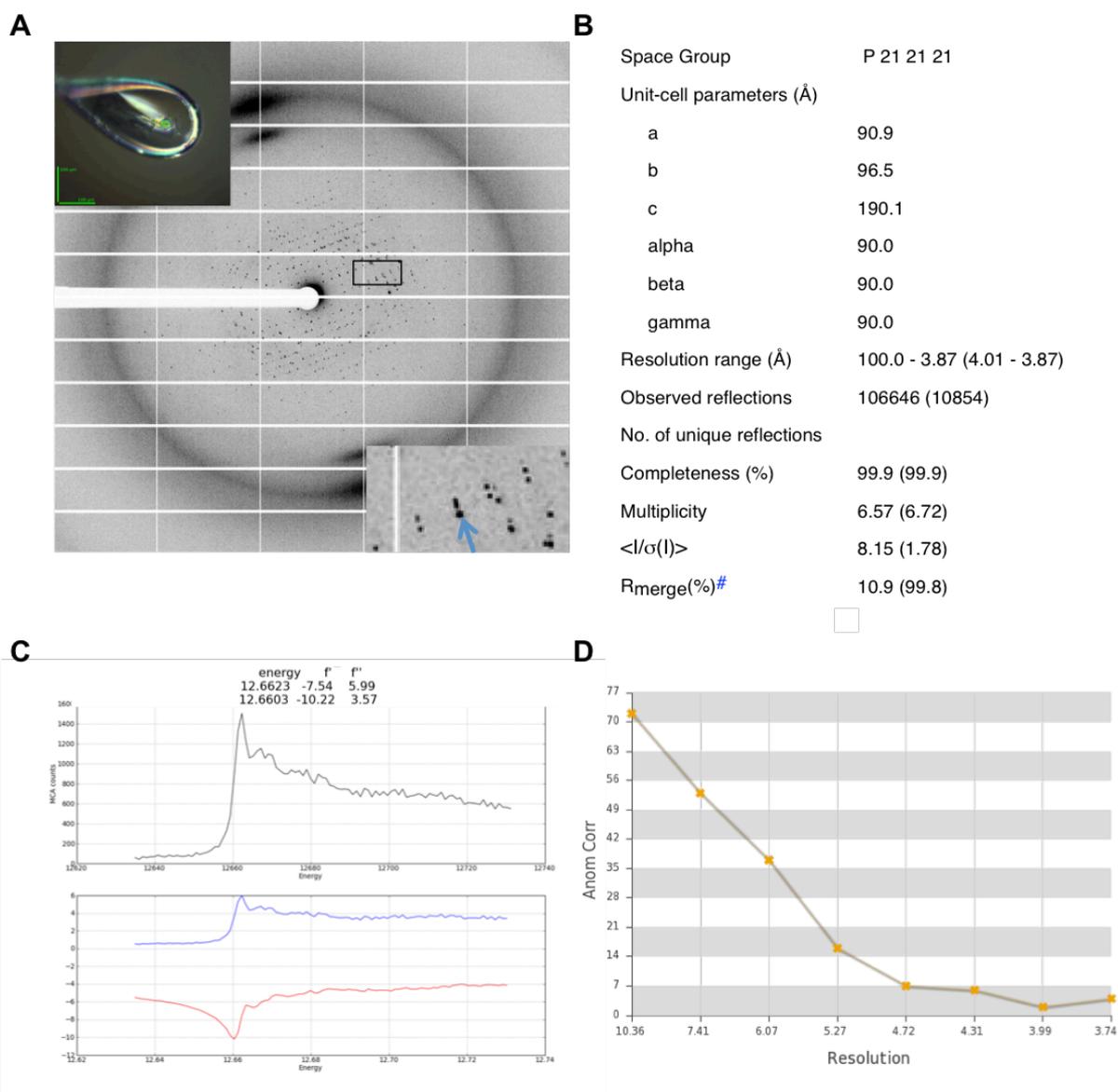


Figure 40: SAD data collection on RanBP5. (A) Diffraction image on the Pillatus 6M detector and crystal picture with green scale bars representing 100 μm . (ID 29, ESRF Grenoble). **(B)** Indexing/Integration statistics from XDS. Bracket values describe the lower resolution shell statistics. **(C)** Selenium energy scan with anomalous signal difference prediction. **(D)** Anomalous correction plotted against resolution (XDS).

Once the data was integrated, anomalous signal can also be quantified by XDS during integration. Anomalous correction, a percentage of correlation between random half-sets of anomalous intensity differences, can be plotted against the resolution shells (Figure 40D). Above a value of 10, it indicates the presence of anomalous data. As observed on the plot, although reflections were integrated to 3.8 Å, anomalous data only extends to about 5 Å.

Upon integration, selenomethionine sites and practical resolution for phasing can be estimated using the HKL2MAP program (Pape and Schneider, 2004) which recursively runs the SHELX sub-programs (Sheldrick, 2010). HKL2MAP estimated the presence of 39 selenomethionines at a cut-off resolution of 6 Å. We expected to find 42 methionines in a full RanBP5 molecule. However due to the poor resolution of the anomalous data, no clear phasing solution could be distinguished using SHELX (Sheldrick, 2008).

In order to potentially improve the odds of finding a phasing solution, Single Isomorphous Replacement using Anomalous Signal (SIRAS) was used. The method consists in combining both the 3.3 Å resolution *in situ* proteolysis dataset with selenomethionine dataset at 3.8 Å, in order to have a sharper anomalous signal estimate. SHARP (Bricogne et al., 2003) was used to calculate a phasing solution but did not converge towards a strong phasing solution at the time of the writing of this thesis.

It is interesting to note that while 2 of the diffraction selenomethionine crystals conserved the native P_{212121} unit cell ($a/b/c=92/96/192$), two other crystals, diffracting at around 4.5 Å, had a P_{212121} unit cell with distinctly different dimensions ($a/b/c=97/125/149$). This phenomenon was never observed with native RanBP5 crystals implying a different crystal packing induced by the presence of selenomethionine. These observations also give a clue to why more twinning is observed in selenomethionine crystals as the two existing unit cells may compete during crystal growth.

In any case, bigger and better diffracting selenomethionine crystals will be required to solve the phase problem.

4.8.2 SIR AND MIR PHASING

At present, important focus is being put on obtaining bigger RanBP5 crystals in order to start testing different heavy atom soaks. The process will take time to fulfil as only certain heavy atom salts interact with a given protein and out of those, many will compromise the crystal integrity as they can destabilise the protein.

In the event that a derivative crystal dataset is collected, it will be combined with the SAD dataset to improve phasing solutions.

4.9 CONCLUSION AND PERSPECTIVES

When expressed in its native form in insect cells, RanBP5 can easily be expressed and purified to a relative purity through a two-step purification process. The protein is rather stable, homogenous and monodisperse as shown by SAXS and thermofluor. Although forming crystals in a restricted number of crystallisation conditions, the process and the parameters controlling crystallogenesis are still poorly understood. Some factors such as adding small quantities of chymotrypsin have improved the odds of crystallogenesis while some purifications fail to produce crystals for no obvious reasons. The produced crystals therefore remain small and of relatively low diffracting quality although some rare crystals have yielded good quality diffraction relative to their size. In turn, this has limited our ability to use MR or experimental methods to solve the phase problem. Precise tuning of crystallogenesis may eventually enable a structure resolution by SAD or SIR.

It is important to note however that most published importin- β structures were published through a complex formation of either RanGTP or an NLS cargo (Kobayashi and Matsuura, 2013; Kobayashi et al., 2015; Lee et al., 2005). In all cases the stabilising effect increased the overall resolution enabling a solution to the phase problem. For some cases loop truncations were required for ligand-free β -importin crystallisation (Kobayashi and Matsuura, 2013; Soniat et al., 2013). These observations indicate that in order to solve the structure of ligand-free RanBP5, a structural determination of NLS or RanGTP bound RanBP5 will be required. The cloning and expression of RanBP5/RanGTP and RanBP5/PB1(187-211) is ongoing in order to solve the phasing problem, but also with the final aim to explain the binding and dissociation kinetics of RanBP5 and the RdRp through structural and biochemical characterisation.

CHAPTER 5: MICRO-SCALE CONSTRUCT SCREENING USING AMBR

Résumé du chapitre 5 en Français

L'expression de protéines en cellules d'insecte peut se révéler très efficace pour obtenir certains complexes protéiques pour lesquels l'expression en bactéries échoue systématiquement. Cependant sa mise en place impose des contraintes techniques plus lourdes comme le contrôle de la densité cellulaire, du pH et de l'oxygénation du milieu de culture. Ces contraintes couplées à l'émergence relativement récente de l'expression recombinante en cellules d'insecte ont limité le développement de méthodes d'expression en haut débit, couramment utilisées avec les bactéries. Dans ce chapitre, nous avons utilisé nos constructions d'ARN polymérase tronquée et les partenaires cellulaires (Red-Smu et RanBP5) pour développer une méthodologie de criblage de protéines recombinantes dans le but d'estimer leur qualité en vue d'une étude structurale. Pour ce faire 10 constructions différentes ont été testées. Chacune de ces constructions a permis l'infection simultanée de 20 micro-fermenteurs spécialement conçus pour les cultures en suspension de cellules eucaryotes. Les cultures ont été maintenues par AMBR (Automated Micro Bio-Reactor), un automate capable de contrôler les paramètres de culture et d'effectuer des prélèvements cellulaires à intervalles récurrents. Grâce au suivi de la fluorescence YFP et CFP chaque culture a pu être stoppée à son maximum d'expression. La phase automatisée d'expression est suivie d'un protocole de micro-purification par résine d'affinité et chromatographie d'exclusion de taille. Les résultats ont mis en évidence qu'une analyse à micro échelle du profil de gel filtration des constructions combinée à un gel SDS-PAGE permet une estimation précise de plusieurs paramètres. Ces paramètres incluent le taux d'expression, la pureté, et la solubilité de chaque construction. Notre démarche a aussi impliqué une caractérisation de la température de fusion (T_m) par thermofluor qui permet d'estimer la stabilité relative de chaque construction.

Ces résultats feront partie d'un article présentant les différentes applications possibles d'expression de cellules d'insecte en micro fermenteurs utilisant la plateforme AMBR.

NOTE OF INTRODUCTION TO CHAPTER 5

The following chapter describes a direct application of the construct library described in **chapter 3**. The use of micro-scale and automated expression control is described and combined with small-scale purification protocols. In all, the results describe one the first high-throughput methodologies for insect cell-culture and soluble recombinant protein analysis.

This chapter will incorporate part of a methods paper focusing on the AMBR expression platform applications and will be published before the end of 2015. The future article will present three different applications ranging from small-scale expression/purification screening, to micro scale crystallogenesis and influenza Virus Like Particle (VLP) production.

5.1 INTRODUCTION

For over forty years since its discovery, the influenza RNA-dependent RNA-polymerase (RdRp) had remained out of reach to crystallographers. Yet crucial atomic insight within this viral polymerase was needed to understand the mechanisms of transcription and replication, both taking place in the nucleus of infected cells. Expression problems were the key bottleneck to studying the structure of this multi-subunit complex, composed of PA, PB1 and PB2. Classical expression methods using prokaryotic cells or mammalian cell co-expression failed to produce the heterotrimer in either soluble form or sufficient quantities for structural biology. This all changed with the introduction of a polyprotein expression strategy in insect cells using the Multibac system (Crépin et al., 2015; Nie et al., 2014; Vijayachandran et al., 2011). In 2014, the structures of full influenza B (Reich et al., 2014) and bat strain H17N10 influenza A (Pflug et al., 2014) polymerases were solved by X-ray crystallography providing a unique insight into the polymerase architecture. However, to date, no full polymerase structures have been described for either human (H3N2) or highly pathogenic avian (H5N1) strains, which would be crucial in the design of broad spectrum inhibiting compounds. Also, the RdRp is believed to recruit a plethora of cellular partners when infecting host cells. Some of these include α - and β -importins, which enable nuclear assembly of the polymerase subunits. As such, RanBP5 exports PA-PB1 to the nucleus (Deng et al., 2006b; Hutchinson et al., 2011) and PB2 is transported separately by importin- α (Mukaigawa and Nayak, 1991; Tarendeau et al., 2007). More recently, studies have also shown the polymerase to interact with the RED-SMU1 heterocomplex which is implicated in mRNA splicing of NS1/NS2 genes (Fournier et al., 2014). RED (557 aa) has been identified as part of the spliceosome (Neubauer et al., 1998; Zhou et al., 2002) and owes its name from highly rich arginine (R), glutamic acid (E) and aspartic acid (D) stretches (Assier et al., 1999) in its core region (res 334 to 375). Little is known of the structure or function of SMU1.

Considering the unresolved problem of full polymerase expression, we have designed a rational screen of truncated core polymerase constructs, using the H3N2 strain. In all, we investigate and compare the expression yield and purification recovery of truncated human polymerase constructs alongside with RED-SMU1 and RanBP5. To do so, we have used AMBR (Tap biosystems). AMBR is an automated micro-scale fermenter platform capable of running 24 expression experiments simultaneously. This platform provides higher throughput ability towards tedious

manual insect cell suspension-culture maintenance, whilst maintaining optimal growth parameters. As such, when harvested, all cultures of a same batch will ensure maximal expression yield per construct. With these small-scale expression cultures, we have then used small-scale nickel resin purification and an äkta micro (GE healthcare) S200 gel filtration. All with the intent to provide an easy and direct protocol towards obtaining purified recombinant protein. In this application, we demonstrate the ability to compare qualitatively and quantitatively the output from every construct. These parameters can then drive the choice towards the most appropriate construct choice for structural biology applications.

5.2 CONSTRUCT PANEL OVERVIEW

Overall, 10 different constructs were tested (Figure 41A); including 6 RdRp truncation constructs and 4 cellular partners RED-SMU1 and RanBP5 constructs. Polymerase truncation constructs were all designed with the aim to express a minimal polymerase assembly composed of PA-PB1 with or without the endonuclease domain which clearly forms a separate domain from the PA-PB1 core, as seen on the X-Ray structure of Bat influenza A polymerase (Figure 41B). A His-tag positioning was also alternatively introduced in the C- and N-terminal extremities. Other constructs including a PB2 truncation extending to the residue 36 were tested with or without an endonuclease domain. PB2(1-36) was identified as a minimal PB1-interacting domain (Sugiyama et al., 2009) and hence may contribute to a higher stability of the RdRp complex. RED-SMU1 was both tested in its complete form and with two C-terminal random truncations of RED based on the literature (Fournier et al., 2014). RanBP5 was tested in its complete form and was expressed as the only non-polyprotein construct.

RdRp and RED-SMU1 constructs were expressed within the polyprotein expression system. This strategy uses the expression of a single polypeptide chain in which the genes of interest are inserted between TEV and CFP genes. Peptidic linkers in between the different genes present a TEV cleavage sites. As the polyprotein is expressed, TEV self cleaves itself from the polyprotein and processes the linkers in between all the other proteins. This system ensures a direct stoichiometry in the expression of all the different partners. CFP is the last protein within the polypeptide chain. The measurement of CFP fluorescence enables a direct yield estimate of overall polyprotein expression. CFP fluorescence can also be rationalised with YFP, which is expressed by the baculovirus and under the control of another polyhedrin promoter.

For every construct, the AMBR experiments were realised both in Sf21 and High Five cells. Due to a consistent reduced yield in expression with the Sf21 cells, purification output results are only shown for the High Five AMBR experiments.

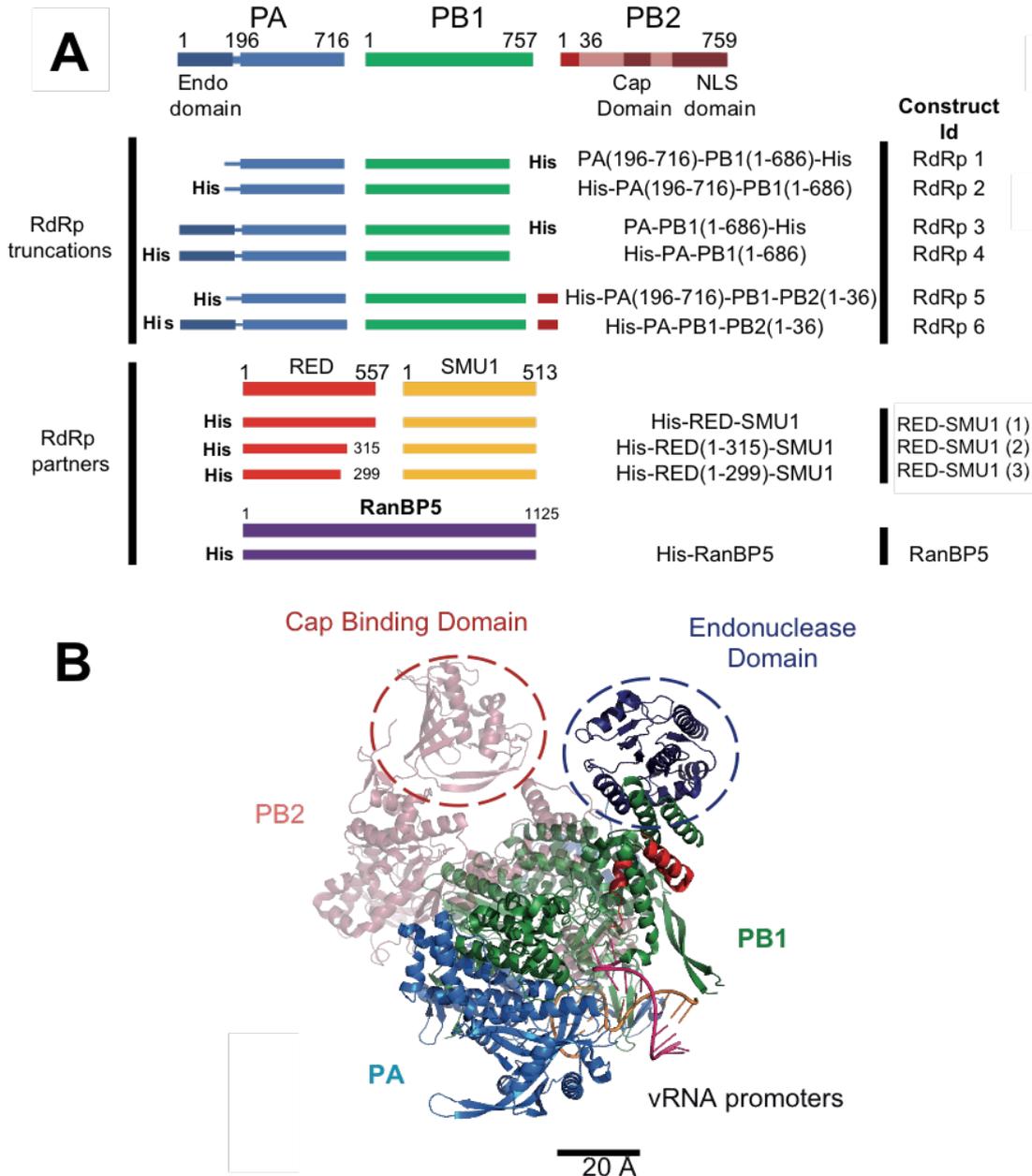


Figure 41: Constructs tested using AMBR. (A) Schematic view of the construct truncation library used to produce 10 independent expression and purification experiments. Polymerase truncations deprived (PA(196-716)-PB1(1-686)) or including (PA-PB1(1-686)) the PA endonuclease domain with alternate C-terminus or N-terminus His-tags (RdRp 1, 2, 3 and 4). Heterotrimer truncations possessing the N-terminal domain of PB2 and once again deprived (PA(196-716)-PB1-PB2(1-36)) or possessing (PA-PB1-PB2(1-36)) the endonuclease domain (RdRp 5 and 6 respectively). Cellular partner proteins include RED-SMU1 heterodimers with full and C-terminal truncations of RED. Also included within the experiment is RanBP5, which is expressed as a single protein. (B) X-ray structure (pdb id: 4WSB) of the bat flu A polymerase with subunits PA, PB1 and PB2 coloured with the same colour code as the RdRp construct diagram. 3' and 5' vRNA promoters are coloured in purple and yellow respectively. Endonuclease and

cap binding domains are shown with dashed circles. PB2 domains which were not included within the AMBR constructs are depicted in faded red.

As a qualitative control for the whole production process leading to a purified product, the RED-SMU1(3) construct was duplicated during the early baculovirus generation stage.

5.3 AMBR PURIFICATION RESULTS

All the tested constructs were eluted from SEC purification using the äkta micro purifier. Peak fractions were analysed by a gradient 4-15 % SDS PAGE gel (Figure 42A). As seen on the gel, all constructs present visible bands with varying intensities. RdRp constructs all purify as a complex and show the expected bands for full PA, full PB1, PA(196-716) and PB1(1-686). PB2(1-36) is not detectable using a Coomassie blue coloration as its molecular weight is below the 15 kDa marker. RdRp6 shows a single band, which in effect is the overlay of both full PA, and full PB1 bands as revealed by another isocratic 12 % SDS PAGE gel of the entire SEC elution (Figure 42c). RED-SMU1 constructs also purify as a complex with the RED subunit migrating at 72 kDa when full and migrating below the 50 kDa marker when cleaved (RED-SMU1(2) and RED-SMU1(3)). SMU1 migrates as two characteristic smeared bands (Fournier et al., 2014) next to the 50 kDa marker. RanBP5 produces a single band, which migrates between the 100 and 150 kDa marker. All polyprotein constructs have a detectable TEV double band presence close to the 25 kDa marker, it is expected, as TEV was also His-tagged as an internal control and is difficult to separate using only SEC.

When analysing the RdRp constructs, important effects on the final yield can be observed by comparing RdRp 1&2 and RdRp 3&4 which differ on their His-tag placement. In the first case, the His-tag is fused to the N-terminus of PA. In the second case it is placed on the C-terminus of PB1. As such, a C-terminal placement on PB1 significantly reduces overall purification yields in both cases. However, as mostly visible with RdRp 2 and RdRp 4, the band corresponding to His-PA(196-716) or full His-PA has a stronger intensity than the PB1(1-686) band, indicating the presence of excess PA within the purified product. The opposite is not true when the his-tag is placed on PB1(1-686) as both bands are of similar intensity. Removing the endonuclease also contributes to improving final purification yields, mostly when PB2(1-36) is absent. This can easily be seen when comparing the band intensity of PB1(1-686) throughout RdRp constructs 1 and 3. Finally, the presence of PB2 greatly diminishes the overall expression of all of the polyprotein subunits.

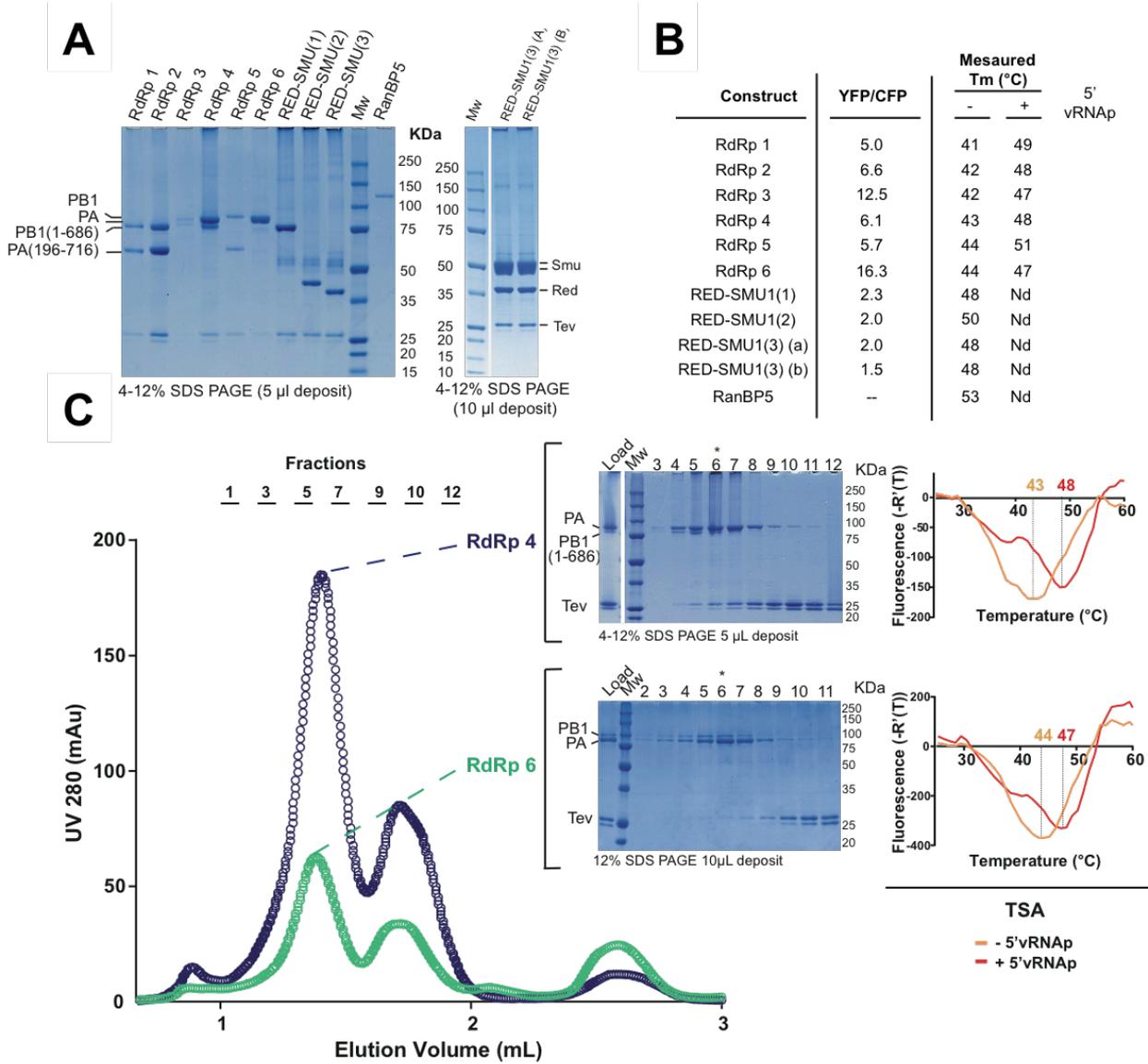


Figure 42: Small-scale purification and characterisation of AMBR cultures. (A) AMBR purified end product analysis by SDS PAGE. (Gel on the left) SEC peak elution fractions from all constructs were deposited in equivalent quantities on a gradient 4-12% SDS gel. (Gel on the right) Control duplicate AMBR experiment of the RED-SMU1(3) construct is shown with higher quantities deposited (10µL). (B) Table of YFP/CFP reporter gene quantification and SEC peak thermal shift assay results for all AMBR constructs. YFP fluorescence ($\lambda_{excitation/emission} = 488/525$) values directly report polyhedrin promoter activity. CFP fluorescence ($\lambda_{excitation/emission} = 430/492$) correlates to the expression of the entire polyprotein constructs, as CFP is the last expressed protein. The relative ratio of YFP/CFP values is indicative of the polyprotein expression efficacy. Thermal shift assay T_m estimates are indicative of the purified sample stability. For RdRp constructs, increases in T_m upon 5'vRNA binding are also measured. (C) SEC chromatogram comparison for RdRp 4 and 6. Äkta micro purification runs of RdRp 4 (blue) and 6 (green) constructs are shown with 280 nm UV absorbance plotted against the elution volume. Corresponding SDS PAGE with the peak fraction (*) and thermal shift assay analysis are also shown on the right of the chromatogram.

This difference is even more visible when comparing the äkta micro elution curves of RdRp 4 and RdRp 6 (Figure 42c). Both constructs are N-terminally His-

tagged and possess the PA endonuclease. However integrating PB2 within the polyprotein not only reduces polymerase but also TEV expression. This result indicates that integrating PB2(1-36) brings upon a translational or post-translational problem for the polyprotein expression and is analogous to the expression problem described at length in **chapter 3**.

All RED-SMU1 constructs are purified at the same level of purity and with comparable yields despite truncations in the RED protein. RanBP5 is expressed in relatively low yields when compared to the other proteins but is observed as a single band migrating between the 100 and 150 kDa marker. Duplication of the RED-SMU1(3) expression and analysis with higher deposits on the SDS PAGE (**Figure 42A**) show identical band intensities for RED(1-299), SMU1 and TEV. Indicating a strong consistency in expression and purification methodology.

YFP and CFP measurements can be rationalised together to produce the YFP/CFP ratio, which reports the relative expression efficacy of a given construct (**Figure 43B**). RED-SMU1 constructs are the most easily expressed constructs within the panel with a low YFP/CFP ratio around 2 at the time of harvest. Within the polymerase constructs, most YFP/CFP ratios are found close to 6. The highest measured ratios are for constructs RdRp 3 and 6 which both have a PA endonuclease domain and either a PB1 C-terminal His-tag or PB2(1-36).

Thermofluor experiments as shown for constructs RdRp 4&6 (**Figure 42C**) were realised on all constructs using the peak fraction to determine a T_m (**Figure 42B**). RdRp constructs T_m was also measured in the presence of 5'vRNAp as previous research has highlighted the increase in T_m attributed to 5'vRNAp binding (Brownlee and Sharps, 2002). From all the constructs tested, RanBP5 is the most stable protein with a T_m of 53°C. RED-SMU1(1) comprising a full RED(1-557) has a T_m of 48°C comparable to that of RED-SMU1(3). RED-SMU1(2) shows a 2°C increase in T_m and is the most stable out of all RED-SMU1 constructs. Polymerases vary greatly in their respective T_m with and without 5'vRNAp. 3 to 8°C stabilisation upon 5'vRNA binding can be observed depending on the constructs. For identical constructs RdRp 1 to 4 with PA (N-terminus) or PB1 (C-terminus) His-tag placements, a consistent 1°C increase in T_m can be observed without 5'vRNAp between RdRp 1&2 and RdRp 3&4. This difference is possibly explained by the presence of co-purified His-PA (196-716) or full His-PA (RdRp 2 and 4 respectfully), which may impact on the melting curve. However upon 5'vRNAp binding the endonuclease deficient construct RdRp 1 with a C-terminal His-tag is the most stable dimeric form of polymerase with a T_m of 49°C. RdRp 5&6 constructs including PB2(1-36) also show an increased T_m at 44°C when unbound to 5'vRNAp indicating that PB2 increases the overall

construct stability. Again, upon 5'vRNAP binding, the lack of the endonuclease domain generates the most stable of all polymerase constructs (RdRp5) with a T_m of 51°C.

5.4 CONCLUSION

Micro-scale expression using the AMBR system and purification with the äkta micro can efficiently be used to assess different parameters that are crucial for the rational choice of constructs for structural characterisation. Constructs can be selected on their expression and solubility levels as shown by the CFP/YFP measurements and SDS-PAGE peak analysis. Purity estimation can also be undertaken with the SDS-PAGE and SEC chromatography profile that can also provide an estimation of oligomeric state. Thermal shift assay performed on the peak fraction can also generate key information on the stability of the purified product. It can as well be used as a qualitative tool for the analysis of binding additives.

Using AMBR, we demonstrate that human strain (H3N2) influenza complexes can be expressed when removing the C-terminus of PB2 as previously shown in **chapter 3**. We highlight the beneficial effect of the endonuclease removal on yield and stability. Interestingly we also show the stabilising effect of the N-terminal domain of PB2(1-36) which has been shown to closely interact with PA-PB1. Also, we confirm the stabilising effect of 5'vRNAP binding on the truncated polymerases, which implies an active energy driven binding of the vRNA. Finally we show the ability to express cellular partners RED-SMU1 and RanBP5, which are purified to high levels of purity and in a reproducible manner as shown by the duplicated expression process of RED-SMU1(3).

In all, we present a methodology opening the domain of medium to high throughput screening of insect cell recombinant expression. This relatively simple workflow tremendously accelerates the search for high value protein candidates and can be exploited both qualitatively and quantitatively. Such a method is also greatly suited for polyprotein screening and will be used to tackle future multi protein complex assemblies.

GENERAL CONCLUSION

The heterotrimeric RNA-dependent RNA-polymerase is the main core component of the influenza virus infectious cycle. Together with the nucleoprotein and vRNA it performs both transcription and replication of the vRNA genome in the nucleus of infected cells. These two activities control both the production of novel viral proteins and vRNA. As such, the polymerase activity controls all the processes leading to novel virus synthesis during infection making it an ideal target for inhibitor drug design.

At the very start of this research project no expression technology could propose a solution to express the RNA polymerase in sufficient quantities for structural biology. To try and overcome this problem, our group in collaboration with others, setup the polyprotein expression strategy to express influenza A polymerase. Although un-successful in the expression of full PA-PB1-PB2, our group managed to show a successful proof of concept by expressing a truncated version of the polymerase, which was both endonuclease and *de-novo* synthesis active. With this construct, PB1 could be expressed for the very first time in a soluble form. Tremendous effort was then invested in finding a suitable truncated construct for the influenza A polymerase. Over 4 years, our group has dissected H5N1 and H3N2 polymerases with the use of over 30 construct variations, all with the intent to obtain a crystal structure of the polymerase. Although some constructs (notably PA(196-716)-PB1(1-686)) showed a very promising biochemical behaviour and even produced protein crystals. No high-resolution diffracting crystals were ever produced with H5N1 or H3N2 strains when truncating the polymerase, even after tedious optimisations steps. This also remains true to this date in the rest of the world.

Many of the purified constructs prove to be difficult proteins in solution. Even when well expressed and easy to purify, many of the truncated constructs (especially with PA or PB2 truncation) behave as dimers in solution when sufficiently concentrated.. The polymerase is composed of highly charged subunits with different polarities; in the context of a truncated RdRp this induces a trans molecular charge compensation which probably is the cause of this dimerisation. V-RNA-p binding induces important conformational changes which induce a monomerisation of the polymerase. It is however difficult to conclude that this oligomeric behaviour is of biological significance. We clearly observe PA-PB1 as a dimer and no full polymerase has ever been reported to form dimers.

Crystal structures of full Bat-A and B polymerases were eventually solved in 2014 by our collaborator Stephen Cusack using the same polyprotein approach. These structures provide a tremendous understanding of the overall polymerase

architecture and cap-snatching mechanism. However, the intricate complexity of the polymerase quaternary structure also opens new questions regarding its assembly process, which possibly requires multiple cellular partners.

In this thesis, the underlying problems of expression of full H3N2 and H5N1 polymerases are advanced: Notably the role played by the PB2 subunit in reducing the expression of the entire polyprotein. No clear answer can be given as for why PB2 is such a problem for the expression of these strains compared to Bat-A or B polymerase. One possible answer to this problem may lie in the required co-expression of other cellular factors, notably Hsp90 which may play a role in the stabilisation of PB2 (Chase et al., 2008; Naito et al., 2007). Also, as a potential solution, one could consider vRNAP bound polymerase synthesis by engineering a cellular transcription of vRNA in parallel to the polyprotein system. In any case, the produced truncation library has brought other answers on the polymerase and notably on its precursor core complex PA-PB1. We have demonstrated that the complex exists as a monomer in solution and possesses *de novo* synthesis activity. We have also shown that it binds with a very high affinity to the 5'vRNAP sequence and that 3'vRNAP requires the presence of PB2. We have proven that truncating the endonuclease domain from the polymerase is feasible. Nonetheless, PB1 requires the long and non-structured hinge region of PA (residues 196-257) to remain soluble. This result is also quite important as it validates the theory that PA-N155 and PA-N182 could assemble with PB1 to form separate polymerase entities whose function would still need to be investigated (Muramoto et al., 2013).

Expansion of the polyprotein expression technology has enabled for the co-expression of RanBP5 with PA-PB1, which further expands the current possibilities to study polymerase and host cell factor interaction.

Using multiple biophysical methods such as SEC-MALLS, EM and SAXS we have further studied how the PA-PB1-RanBP5 complex behaves in solution. PA-PB1-RanBP5 forms a 1:1:1 complex. RanBP5 has been shown to tightly bind to with PA-PB1 and inhibit the 5'vRNAP binding. To further investigate the nature of this binding process, mutations and dissociation studies will be performed. Also, a cross linking mass spec investigation of the complex may advance the structural knowledge and enable for a precise model generation.

The crystallographic characterisation of RanBP5 and its interaction with the NLS domain of PB1 is underway. The unbound form of RanBP5 can be expressed and purified alone and yields medium quality diffracting crystals which contain the full protein. Solving the phase problem for the unbound RanBP5 form may require

solving the structures of NLS or Ran(GTP) bound forms, that is underway but will still yield key structural data, as important conformational differences are expected between the different forms. Solving the X-ray structure of RanBP5 bound to the PB1-NLS will be a key step towards finding any polymerase assembly inhibitors. Putting together an atomic model for RanBP5 with the data of the full import complex will not only broaden our understanding of how the polymerase assembles in the nucleus but also enrich the field of importin/exportin studies. RanBP5, is also involved in the transport of other proteins associated to other viral proteins such as rev from HIV and the L1-L2-L3 from the Human Papilloma Virus (Nelson et al., 2003; Darshan et al., 2004; Klucsevsek et al., 2006). Applying the same rational expression approach to these complexes may also successfully reconstitute these complexes.

Finally, the large panel of polymerase constructs has given us the opportunity to setup some of the very first high throughput expression technology with baculovirus infected cells insect cells. With the AMBR robot, we demonstrate that using small-scale expression, affinity binding and SEC can produce interpretable results which can then drive a rational choice towards the most promising constructs. This experiment was undertaken as a proof of concept, it will now be applied at a higher frequency for construct libraries which were previously untested. One could imagine such a platform being used to screen co-expressed polymerases with potential cellular partners and directly identify complexes that form recombinantly. Micro-scale purification also allows for further structural studies such as EM and micro-crystallisation experiments.

On a final note, this thesis but also more broadly the numerous internships beforehand have clearly defined my path as a researcher. I have had the unique chance to collaborate with our industrial partner (Roche Pharmaceuticals), which gave me the opportunity to get a first-hand experience of working within the industry (January to March 2013 in Nutley, USA). Also, my PhD has made use of multiple biophysical and biochemical methods which have enriched not only my technical abilities but also my critical perspective on how research can be done.

LIST OF FIGURES

Figure 1: Host distribution of type A influenza stains.....	7
Figure 2: Number and percentage of respiratory specimens testing positive for influenza in the year 2010/2011, by type and week	9
Table 1: General overview of currently used influenza inhibitors.	16
Figure 3: Influenza A viral architecture.	17
Figure 4: Organisation of the influenza A genomic segments	19
Figure 5: Structural overview of haemagglutinin from an H3N2 influenza strain.....	22
Figure 6: Surface representation of H1N1 NP	23
Figure 7: Neuraminidase structural features	25
Figure 8: Structures from segment 7 proteins.....	27
Figure 9: Structural features of NS1	30
Figure 10: Influenza A virus life cycle.....	34
Figure 11: Reassortment events driving the emergence of pandemic influenza A	37
Figure 12: EM and cryo-EM models of RdRps isolated or within the context a vRNP	40
Figure 13: Domain structures of the influenza A RdRp	41
Figure 14: influenza A H17N10 overall structure	46
Figure 15: Structure of the PB1 polymerase active site	48
Figure 16: "Cap-snatching mechanism"	50
Figure 17: "Trans" model for vRNP synthesis	52
Figure 18: Nuclear import routes of full RNPs and single constituents through the NPC	54
Figure 19: Kap121p structural architecture	56
Figure 20: Cassette cloning sites within the K7-pMarQ vector	61
Figure 21: Schematic representation of the anisotropy of a fluorescent DNA molecule	67
Figure 22: Schematic view of the SEC-MALLS-RI system and data analysis	70
Figure 23: SAXS schematic principle and experimental setup.	71
Figure 24: Example of a buffer subtraction on I(q) raw data curves.....	72
Figure 25: Guinier transform	73
Figure 26: P(r) function representations for different types of protein	74
Figure 27: Crystallogensis phase diagram	76
Figure 28: Braggs Law	78
Figure 29: AMBR platform overview	83
Figure 30: Two-step purification of RanBP5	94
Figure 31: Online SAXS data analysis	95
Figure 32: Ab initio modelling of RanBP5	96
Figure 33: Quality control steps required for crystallisation	97
Figure 34: RanBP5 crystal characterisation on ID 30A-3.....	98
Figure 35: PEG and pH screening to produce RanBp5 crystals	100
Figure 36: Initial dataset acquisition on ID 29.....	101
Figure 37: <i>In situ</i> proteolysis of RanBP5.....	104
Figure 38: Data collection on RanBP5 crystal grown with in situ proteolysis.....	105
Figure 39: SDS PAGE analysis of RanBP5 crystal content using a silver nitrate stain.....	106
Figure 40: SAD data collection on RanBP5.....	108
Figure 41: Constructs tested using AMBR.....	Error! Bookmark not defined.
Figure 42: Small-scale purification and characterisation of AMBR cultures	119

LIST OF ABBREVIATIONS

ATP	Adenosine TriPhosphate
AMBR	Automated Micro Bio-Reactor
AU	Arbitrary Unit
BME	Beta-mercaptoethanol
CDC	Center for disease Control
CFP	Cyan Fluorescent Protein
CTP	Cytidine TriPhosphate
DNA	Deoxyribonucleic Acid
EMBL	European Molecular Biology Laboratory
ER	Endothelial Reticulum
ESPRIT	Expression of Soluble Proteins by Random Incremental Truncation
ESRF	European Synchrotron Radiation Facility
FBA	Filter Binding Assay
GDP	Growth Domestic Product
GTP	Guanosine TriPhosphate
HA	Hemagglutinin
HEAT	Huntingtin, elongation factor 3 (EF3), protein phosphatase 2A (PP2A), and the yeast kinase TOR1
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
IVA	Influenza A Virus
IVB	Influenza B Virus
kDa	Kilo Dalton
KeV	Kilo electro-Volt
M1/M2	Matrix protein
MAD	Multiple Anomalous Dispersion
MALLS	Multi Angle Laser Light Scattering
MIR	Multiple Isomorphous Replacement
MIRAS	Multiple Isomorphous Replacement with Anomalous Scattering
MR	Molecular Replacement
NA	Neuraminidase
NEP	Nuclear Export Protein
NLS	Nuclear Localisation Signal
NP	NucleoProtein
NS1/2	Non Structural Protein
PA	Protein Acid
PAGE	Polyacrylamide Gel
PB1/PB2	Protein Basic
PBS	Phosphate Buffered Saline

PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
RdRp	RNA-dependent RNA-polymerase
R _g	Radius of Gyration
RNA	Ribonucleic Acid
c-RNA	complementary Ribonucleic Acid
v-RNA	viral Ribonucleic Acid
RNP	RiboNucleoProtein
SAD	Single Anomalous Dispersion
SAXS	Small Angle X-Ray Scattering
SDS	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SIR	Single Isomorphous Replacement
SIRAS	Single Isomorphous Replacement with Anomalous Scattering
ssDNA/RNA	Single strand DNA/RNA
TEV	Tobacco Etch Virus
T _m	Melting Temperature
TSA	Thermal Shift Assay
UTP	Uridine TriPhosphate
UTR	UnTranslated Region
UV	Ultra Violet
VLP	Virus Like Particle
WHO	World Health Organisation
YFP	Yellow Fluorescent Protein
YOA	Years of Age

BIBLIOGRAPHY

- Akarsu, H., Burmeister, W.P., Petosa, C., Petit, I., Müller, C.W., Ruigrok, R.W.H., and Baudin, F. (2003). Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). *EMBO J.* 22, 4646–4655.
- Albertini, A.A.V., Wernimont, A.K., Muziol, T., Ravelli, R.B.G., Clapier, C.R., Schoehn, G., Weissenhorn, W., and Ruigrok, R.W.H. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 313, 360–363.
- Albo, C., Valencia, A., and Portela, A. (1995). Identification of an RNA binding region within the N-terminal third of the influenza A virus nucleoprotein. *J. Virol.* 69, 3799–3806.
- Amorim, M.J., Bruce, E.A., Read, E.K.C., Foeglein, A., Mahen, R., Stuart, A.D., and Digard, P. (2011). A Rab11- and microtubule-dependent mechanism for cytoplasmic transport of influenza A virus viral RNA. *J. Virol.* 85, 4143–4156.
- Area, E., Martín-Benito, J., Gastaminza, P., Torreira, E., Valpuesta, J.M., Carrascosa, J.L., and Ortín, J. (2004). 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl. Acad. Sci. U. S. A.* 101, 308–313.
- Arnold, M., Nath, A., Hauber, J., and Kehlenbach, R.H. (2006). Multiple importins function as nuclear transport receptors for the Rev protein of human immunodeficiency virus type 1. *J. Biol. Chem.* 281, 20883–20890.
- Arranz, R., Coloma, R., Chichón, F.J., Conesa, J.J., Carrascosa, J.L., Valpuesta, J.M., Ortín, J., and Martín-Benito, J. (2012). The Structure of Native Influenza Virion Ribonucleoproteins. *Science* 338, 1634–1637.
- Assier, E., Bouzinba-Segard, H., Stolzenberg, M.C., Stephens, R., Bardos, J., Freemont, P., Charron, D., Trowsdale, J., and Rich, T. (1999). Isolation, sequencing and expression of RED, a novel human gene encoding an acidic-basic dipeptide repeat. *Gene* 230, 145–154.
- Avalos, R.T., Yu, Z., and Nayak, D.P. (1997). Association of influenza virus NP and M1 proteins with cellular cytoskeletal elements in influenza virus-infected cells. *J. Virol.* 71, 2947–2958.
- Avilov, S.V., Moisy, D., Naffakh, N., and Cusack, S. (2012). Influenza A virus progeny vRNP trafficking in live infected cells studied with the virus-encoded fluorescently tagged PB2 protein. *Vaccine* 30, 7411–7417.
- Baake, M., Bäuerle, M., Doenecke, D., and Albig, W. (2001). Core histones and linker histones are imported into the nucleus by different pathways. *Eur. J. Cell Biol.* 80, 669–677.
- Babu, Y.S., Chand, P., Bantia, S., Kotian, P., Dehghani, A., El-Kattan, Y., Lin, T.H., Hutchison, T.L., Elliott, A.J., Parker, C.D., et al. (2000). BCX-1812 (RWJ-270201): discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* 43, 3482–3486.
- Bai, X., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S.H.W., and Shi, Y. (2015). An atomic structure of human γ -secretase. *Nature* 525, 212–217.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* 35, 235–241.
- Bancroft, C.T., and Parslow, T.G. (2002). Evidence for segment-nonspecific packaging of the influenza A virus genome. *J. Virol.* 76, 7133–7139.
- Barton, W.A., Tzvetkova-Robev, D., Erdjument-Bromage, H., Tempst, P., and Nikolov, D.B. (2006). Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. *Protein Sci. Publ. Protein Soc.* 15, 2008–2013.
- Baudin, F., Bach, C., Cusack, S., and Ruigrok, R.W. (1994). Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO J.* 13, 3158–3165.
- Baudin, F., Petit, I., Weissenhorn, W., and Ruigrok, R.W.H. (2001). In Vitro Dissection of the Membrane and RNP Binding Activities of Influenza Virus M1 Protein. *Virology* 281, 102–108.

- Beaton, A.R., and Krug, R.M. (1986). Transcription antitermination during influenza viral template RNA synthesis requires the nucleocapsid protein and the absence of a 5' capped end. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6282–6286.
- Bedford, T., Cobey, S., Beerli, P., and Pascual, M. (2010). Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathog* **6**, e1000918.
- Belshe, R.B., Burk, B., Newman, F., Cerruti, R.L., and Sim, I.S. (1989). Resistance of influenza A virus to amantadine and rimantadine: results of one decade of surveillance. *J. Infect. Dis.* **159**, 430–435.
- Berger, I., Garzoni, F., Chaillet, M., Haffke, M., Gupta, K., and Aubert, A. (2013). The multiBac protein complex production platform at the EMBL. *J. Vis. Exp. JoVE* e50159.
- Biswas, S.K., and Nayak, D.P. (1994). Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. *J. Virol.* **68**, 1819–1826.
- Biswas, S.K., Boutz, P.L., and Nayak, D.P. (1998). Influenza virus nucleoprotein interacts with influenza virus polymerase proteins. *J. Virol.* **72**, 5493–5501.
- Blaas, D., Patzelt, E., and Kuechler, E. (1982). Identification of the cap binding protein of influenza virus. *Nucleic Acids Res.* **10**, 4803–4812.
- Blok, J., Air, G.M., Laver, W.G., Ward, C.W., Lilley, G.G., Woods, E.F., Roxburgh, C.M., and Inglis, A.S. (1982). Studies on the size, chemical composition, and partial sequence of the neuraminidase (NA) from type A influenza viruses show that the N-terminal region of the NA is not processed and serves to anchor the NA in the viral membrane. *Virology* **119**, 109–121.
- Boekema, E.J., Folea, M., and Kouřil, R. (2009). Single particle electron microscopy. *Photosynth. Res.* **102**, 189–196.
- Boivin, S., Cusack, S., Ruigrok, R.W.H., and Hart, D.J. (2010). Influenza A Virus Polymerase: Structural Insights into Replication and Host Adaptation Mechanisms. *J. Biol. Chem.* **285**, 28411–28417.
- Bornholdt, Z.A., and Prasad, B.V.V. (2006). X-ray structure of influenza virus NS1 effector domain. *Nat. Struct. Mol. Biol.* **13**, 559–560.
- Boulo, S., Akarsu, H., Ruigrok, R.W.H., and Baudin, F. (2007). Nuclear traffic of influenza virus proteins and ribonucleoprotein complexes. *Virus Res.* **124**, 12–21.
- Bouloy, M., Plotch, S.J., and Krug, R.M. (1978). Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc. Natl. Acad. Sci. U. S. A.* **75**, 4886–4890.
- Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around *SHARP* 2.0. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 2023–2030.
- Broor, S., Krishnan, A., Roy, D.S., Dhakad, S., Kaushik, S., Mir, M.A., Singh, Y., Moen, A., Chadha, M., Mishra, A.C., et al. (2012). Dynamic patterns of circulating seasonal and pandemic A(H1N1)pdm09 influenza viruses from 2007–2010 in and around Delhi, India. *PLoS One* **7**, e29129.
- Brownlee, G.G., and Sharps, J.L. (2002). The RNA polymerase of influenza a virus is stabilized by interaction with its viral RNA promoter. *J. Virol.* **76**, 7103–7113.
- Bruce, E.A., Digard, P., and Stuart, A.D. (2010). The Rab11 pathway is required for influenza A virus budding and filament formation. *J. Virol.* **84**, 5848–5859.
- Bullido, R., Gómez-Puertas, P., Saiz, M.J., and Portela, A. (2001). Influenza A virus NEP (NS2 protein) downregulates RNA synthesis of model template RNAs. *J. Virol.* **75**, 4912–4917.

- Bush, R.M., Bender, C.A., Subbarao, K., Cox, N.J., and Fitch, W.M. (1999). Predicting the evolution of human influenza A. *Science* 286, 1921–1925.
- Cannell, J.J., Vieth, R., Umhau, J.C., Holick, M.F., Grant, W.B., Madronich, S., Garland, C.F., and Giovannucci, E. (2006). Epidemic influenza and vitamin D. *Epidemiol. Infect.* 134, 1129–1140.
- Carrat, F., Vergu, E., Ferguson, N.M., Lemaître, M., Cauchemez, S., Leach, S., and Valleron, A.-J. (2008). Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies. *Am. J. Epidemiol.* 167, 775–785.
- Carrillo, B., Choi, J.-M., Bornholdt, Z.A., Sankaran, B., Rice, A.P., and Prasad, B.V.V. (2014). The influenza A virus protein NS1 displays structural polymorphism. *J. Virol.* 88, 4113–4122.
- (CDC) (2006). High levels of adamantane resistance among influenza A (H3N2) viruses and interim guidelines for use of antiviral agents--United States, 2005-06 influenza season. *MMWR Morb. Mortal. Wkly. Rep.* 55, 44–46.
- CDC (2010). 2009 H1N1: Overview of a Pandemic. Available at <http://www.cdc.gov/h1n1flu/yearinreview/yir5.htm>.
- CDC (2011). Update: Influenza Activity. Available at <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6021a5.htm>. *Morb. Mortal. Wkly. Rep.* 60, 706.
- Chakrabarti, A.K., and Pasricha, G. (2013). An insight into the PB1F2 protein and its multifunctional role in enhancing the pathogenicity of the influenza A viruses. *Virology* 440, 97–104.
- Chan, J., Holmes, A., and Rabadan, R. (2010). Network Analysis of Global Influenza Spread. *PLoS Comput Biol* 6, e1001005.
- Chan, J.F.-W., To, K.K.-W., Tse, H., Jin, D.-Y., and Yuen, K.-Y. (2013). Interspecies transmission and emergence of novel viruses: lessons from bats and birds. *Trends Microbiol.* 21, 544–555.
- Chang, S., Sun, D., Liang, H., Wang, J., Li, J., Guo, L., Wang, X., Guan, C., Boruah, B.M., Yuan, L., et al. (2015). Cryo-EM structure of influenza virus RNA polymerase complex at 4.3 Å resolution. *Mol. Cell* 57, 925–935.
- Chase, G., Deng, T., Fodor, E., Leung, B.W., Mayer, D., Schwemmler, M., and Brownlee, G. (2008). Hsp90 inhibitors reduce influenza virus replication in cell culture. *Virology* 377, 431–439.
- Chen, B.J., Leser, G.P., Jackson, D., and Lamb, R.A. (2008). The influenza virus M2 protein cytoplasmic tail interacts with the M1 protein and influences virus assembly at the site of virus budding. *J. Virol.* 82, 10059–10070.
- Chen, W., Calvo, P.A., Malide, D., Gibbs, J., Schubert, U., Bacik, I., Basta, S., O'Neill, R., Schickli, J., Palese, P., et al. (2001). A novel influenza A virus mitochondrial protein that induces cell death. *Nat. Med.* 7, 1306–1312.
- Chenavas, S., Estrozi, L.F., Slama-Schwok, A., Delmas, B., Di Primo, C., Baudin, F., Li, X., Crépin, T., and Ruigrok, R.W.H. (2013). Monomeric Nucleoprotein of Influenza A Virus. *PLoS Pathog.* 9.
- Cheng, A., Wong, S.M., and Yuan, Y.A. (2009). Structural basis for dsRNA recognition by NS1 protein of influenza A virus. *Cell Res.* 19, 187–195.
- Chou, Y., Heaton, N.S., Gao, Q., Palese, P., Singer, R.H., Singer, R., and Lionnet, T. (2013). Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog.* 9, e1003358.
- Chu, C.M., Dawson, I.M., and Elford, W.J. (1949). Filamentous forms associated with newly isolated influenza virus. *Lancet* 1, 602.

- Chung, K.M., Cha, S.-S., and Jang, S.K. (2008). A novel function of karyopherin beta3 associated with apolipoprotein A-I secretion. *Mol. Cells* 26, 291–298.
- Claas, E.C., de Jong, J.C., van Beek, R., Rimmelzwaan, G.F., and Osterhaus, A.D. (1998a). Human influenza virus A/HongKong/156/97 (H5N1) infection. *Vaccine* 16, 977–978.
- Claas, E.C., Osterhaus, A.D., van Beek, R., De Jong, J.C., Rimmelzwaan, G.F., Senne, D.A., Krauss, S., Shortridge, K.F., and Webster, R.G. (1998b). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet* 351, 472–477.
- Colman, P.M., Varghese, J.N., and Laver, W.G. (1983). Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* 303, 41–44.
- Coloma, R., Valpuesta, J.M., Arranz, R., Carrascosa, J.L., Ortín, J., and Martín-Benito, J. (2009). The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog.* 5, e1000491.
- Compans, R.W., Content, J., and Duesberg, P.H. (1972). Structure of the ribonucleoprotein of influenza virus. *J. Virol.* 10, 795–800.
- Connor, R.J., Kawaoka, Y., Webster, R.G., and Paulson, J.C. (1994). Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* 205, 17–23.
- Cook, A., Bono, F., Jinek, M., and Conti, E. (2007). Structural biology of nucleocytoplasmic transport. *Annu. Rev. Biochem.* 76, 647–671.
- Crépin, T., Dias, A., Palencia, A., Swale, C., Cusack, S., and Ruigrok, R.W.H. (2010). Mutational and metal binding analysis of the endonuclease domain of the influenza virus polymerase PA subunit. *J. Virol.* 84, 9096–9104.
- Crépin, T., Swale, C., Monod, A., Garzoni, F., Chaillet, M., and Berger, I. (2015). Polyproteins in structural biology. *Curr. Opin. Struct. Biol.* 32, 139–146.
- Cronin, C.N., Lim, K.B., and Rogers, J. (2007). Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. *Protein Sci. Publ. Protein Soc.* 16, 2023–2029.
- Cros, J.F., and Palese, P. (2003). Trafficking of viral genomic RNA into and out of the nucleus: influenza, Thogoto and Borna disease viruses. *Virus Res.* 95, 3–12.
- Cros, J.F., García-Sastre, A., and Palese, P. (2005). An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. *Traffic Cph. Den.* 6, 205–213.
- Danieli, T., Pelletier, S.L., Henis, Y.I., and White, J.M. (1996). Membrane fusion mediated by the influenza virus hemagglutinin requires the concerted action of at least three hemagglutinin trimers. *J. Cell Biol.* 133, 559–569.
- Darshan, M.S., Lucchi, J., Harding, E., and Moroianu, J. (2004). The I2 minor capsid protein of human papillomavirus type 16 interacts with a network of nuclear import receptors. *J. Virol.* 78, 12179–12188.
- Datta, K., Wolkerstorfer, A., Szolar, O.H.J., Cusack, S., and Klumpp, K. (2013). Characterization of PA-N terminal domain of Influenza A polymerase reveals sequence specific RNA cleavage. *Nucleic Acids Res.* 41, 8289–8299.
- Davies, W.L., Grunert, R.R., Haff, R.F., MCGahen, J.W., Neumayer, E.M., Paulshock, M., Watts, J.C., Wood, T.R., Hermann, E.C., and Hoffmann, C.E. (1964). ANTIVIRAL ACTIVITY OF 1-ADAMANTANAMINE (AMANTADINE). *Science* 144, 862–863.
- Deng, T., Sharps, J., Fodor, E., and Brownlee, G.G. (2005). In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *J. Virol.* 79, 8669–8674.
- Deng, T., Vreede, F.T., and Brownlee, G.G. (2006a). Different de novo initiation strategies are used by influenza virus RNA polymerase on its cRNA and viral RNA promoters during viral RNA replication. *J. Virol.* 80, 2337–2348.

- Deng, T., Engelhardt, O.G., Thomas, B., Akoulitchev, A.V., Brownlee, G.G., and Fodor, E. (2006b). Role of ran binding protein 5 in nuclear import and assembly of the influenza virus RNA polymerase complex. *J. Virol.* **80**, 11911–11919.
- Dias, A., Bouvier, D., Crépin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S., and Ruigrok, R.W.H. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**, 914–918.
- Digard, P., Blok, V.C., and Inglis, S.C. (1989). Complex formation between influenza virus polymerase proteins expressed in *Xenopus* oocytes. *Virology* **171**, 162–169.
- Dolin, R., Reichman, R.C., Madore, H.P., Maynard, R., Linton, P.N., and Webber-Jones, J. (1982). A controlled trial of amantadine and rimantadine in the prophylaxis of influenza A infection. *N. Engl. J. Med.* **307**, 580–584.
- Dong, A., Xu, X., and Edwards, A.M. (2007). In situ proteolysis for protein crystallization and structure determination. *Nat. Methods* **4**, 1019–1021.
- Drake, J.W. (1993). Rates of spontaneous mutation among RNA viruses. *Proc. Natl. Acad. Sci.* **90**, 4171–4175.
- Du, J., Lü, W., Wu, S., Cheng, Y., and Gouaux, E. (2015). Glycine receptor mechanism elucidated by electron cryo-microscopy. *Nature*.
- DuBois, R.M., Slavish, P.J., Baughman, B.M., Yun, M.-K., Bao, J., Webby, R.J., Webb, T.R., and White, S.W. (2012). Structural and biochemical basis for development of influenza virus inhibitors targeting the PA endonuclease. *PLoS Pathog.* **8**, e1002830.
- Dundon, W.G. (2012). Variability among the neuraminidase, non-structural 1 and PB1-F2 proteins in the influenza A virus genome. *Virus Genes* **44**, 363–373.
- Dupeux, F., Röwer, M., Seroul, G., Blot, D., and Márquez, J.A. (2011). A thermal stability assay can help to estimate the crystallization likelihood of biological samples. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 915–919.
- Dutkowski, R., Thakrar, B., Froehlich, E., Suter, P., Oo, C., and Ward, P. (2003). Safety and pharmacology of oseltamivir in clinical use. *Drug Saf.* **26**, 787–801.
- Eisfeld, A.J., Kawakami, E., Watanabe, T., Neumann, G., and Kawaoka, Y. (2011). RAB11A is essential for transport of the influenza virus genome to the plasma membrane. *J. Virol.* **85**, 6117–6126.
- Elster, C., Fourest, E., Baudin, F., Larsen, K., Cusack, S., and Ruigrok, R.W. (1994). A small percentage of influenza virus M1 protein contains zinc but zinc does not influence in vitro M1-RNA interaction. *J. Gen. Virol.* **75** (Pt 1), 37–42.
- Elton, D., Simpson-Holley, M., Archer, K., Medcalf, L., Hallam, R., McCauley, J., and Digard, P. (2001). Interaction of the influenza virus nucleoprotein with the cellular CRM1-mediated nuclear export pathway. *J. Virol.* **75**, 408–419.
- Enami, M., Sharma, G., Benham, C., and Palese, P. (1991). An influenza virus containing nine different RNA segments. *Virology* **185**, 291–298.
- Engelhardt, O.G., Smith, M., and Fodor, E. (2005). Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* **79**, 5812–5818.
- Evans, G., and Pettifer, R.F. (2001). *CHOOCH*: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Crystallogr.* **34**, 82–86.
- Falcón, A.M., Marión, R.M., Zürcher, T., Gómez, P., Portela, A., Nieto, A., and Ortín, J. (2004). Defective RNA replication and late gene expression in temperature-sensitive influenza viruses expressing deleted forms of the NS1 protein. *J. Virol.* **78**, 3880–3888.
- Fechter, P., Mingay, L., Sharps, J., Chambers, A., Fodor, E., and Brownlee, G.G. (2003). Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. *J. Biol. Chem.* **278**, 20381–20388.

- Fiore, A.E., Bridges, C.B., and Cox, N.J. (2009). Seasonal Influenza Vaccines. In *Vaccines for Pandemic Influenza*, R.W. Compans, and W.A. Orenstein, eds. (Springer Berlin Heidelberg), pp. 43–82.
- Flick, R., Neumann, G., Hoffmann, E., Neumeier, E., and Hobom, G. (1996). Promoter elements in the influenza vRNA terminal structure. *RNA N. Y. N 2*, 1046–1057.
- Floyd, D.L., Ragains, J.R., Skehel, J.J., Harrison, S.C., and van Oijen, A.M. (2008). Single-particle kinetics of influenza virus membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 15382–15387.
- Fodor, E., and Smith, M. (2004). The PA subunit is required for efficient nuclear accumulation of the PB1 subunit of the influenza A virus RNA polymerase complex. *J. Virol.* *78*, 9144–9153.
- Fodor, E., Pritlove, D.C., and Brownlee, G.G. (1994). The influenza virus panhandle is involved in the initiation of transcription. *J. Virol.* *68*, 4092–4096.
- Fodor, E., Crow, M., Mingay, L.J., Deng, T., Sharps, J., Fechter, P., and Brownlee, G.G. (2002). A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J. Virol.* *76*, 8989–9001.
- Fortes, P., Beloso, A., and Ortín, J. (1994). Influenza virus NS1 protein inhibits pre-mRNA splicing and blocks mRNA nucleocytoplasmic transport. *EMBO J.* *13*, 704–712.
- Fournier, E., Moules, V., Essere, B., Paillart, J.-C., Sirbat, J.-D., Isel, C., Cavalier, A., Rolland, J.-P., Thomas, D., Lina, B., et al. (2012). A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* *40*, 2197–2209.
- Fournier, G., Chiang, C., Munier, S., Tomoiu, A., Demeret, C., Vidalain, P.-O., Jacob, Y., and Naffakh, N. (2014). Recruitment of RED-SMU1 Complex by Influenza A Virus RNA Polymerase to Control Viral mRNA Splicing. *PLoS Pathog.* *10*.
- Francis, T. (1940). A New Type of Virus from Epidemic Influenza. *Science* *92*, 405–408.
- Frank, J. (1996). Chapter 2 - Electron Microscopy of Macromolecular Assemblies. In *Three-Dimensional Electron Microscopy of Macromolecular Assemblies*, J. Frank, ed. (Burlington: Academic Press), pp. 12–53.
- Frost, W.H. (2006). The epidemiology of influenza. 1919. *Public Health Rep. Wash. DC* *1974 121 Suppl 1*, 149–159; discussion 148.
- Fu, X., Choi, Y.-K., Qu, D., Yu, Y., Cheung, N.S., and Qi, R.Z. (2006). Identification of nuclear import mechanisms for the neuronal Cdk5 activator. *J. Biol. Chem.* *281*, 39014–39021.
- Fujii, K., Fujii, Y., Noda, T., Muramoto, Y., Watanabe, T., Takada, A., Goto, H., Horimoto, T., and Kawaoka, Y. (2005). Importance of both the Coding and the Segment-Specific Noncoding Regions of the Influenza A Virus NS Segment for Its Efficient Incorporation into Virions. *J. Virol.* *79*, 3766–3774.
- Fujii, Y., Goto, H., Watanabe, T., Yoshida, T., and Kawaoka, Y. (2003). Selective incorporation of influenza virus RNA segments into virions. *Proc. Natl. Acad. Sci.* *100*, 2002–2007.
- Fujiyoshi, Y., Kume, N.P., Sakata, K., and Sato, S.B. (1994). Fine structure of influenza A virus observed by electron cryo-microscopy. *EMBO J.* *13*, 318–326.
- Furuta, Y., Takahashi, K., Kuno-Maekawa, M., Sangawa, H., Uehara, S., Kozaki, K., Nomura, N., Egawa, H., and Shiraki, K. (2005). Mechanism of action of T-705 against influenza virus. *Antimicrob. Agents Chemother.* *49*, 981–986.
- Gack, M.U., Shin, Y.C., Joo, C.-H., Urano, T., Liang, C., Sun, L., Takeuchi, O., Akira, S., Chen, Z., Inoue, S., et al. (2007). TRIM25 RING-finger E3 ubiquitin ligase is essential for RIG-I-mediated antiviral activity. *Nature* *446*, 916–920.

- Gamblin, S.J., and Skehel, J.J. (2010). Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J. Biol. Chem.* *285*, 28403–28409.
- Gao, Q., Lowen, A.C., Wang, T.T., and Palese, P. (2010). A nine-segment influenza A virus carrying subtype H1 and H3 hemagglutinins. *J. Virol.* *84*, 8062–8071.
- Gao, R., Bhatnagar, J., Blau, D.M., Greer, P., Rollin, D.C., Denison, A.M., DeLeon-Carnes, M., Shieh, W.-J., Sambhara, S., Tumpey, T.M., et al. (2013). Cytokine and chemokine profiles in lung tissues from fatal cases of 2009 pandemic influenza A (H1N1): role of the host immune response in pathogenesis. *Am. J. Pathol.* *183*, 1258–1268.
- Garten, R.J., Davis, C.T., Russell, C.A., Shu, B., Lindstrom, S., Balish, A., Sessions, W.M., Xu, X., Skepner, E., Deyde, V., et al. (2009). Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. *Science* *325*, 197–201.
- Gastaminza, P., Perales, B., Falcón, A.M., and Ortín, J. (2003). Mutations in the N-terminal region of influenza virus PB2 protein affect virus RNA replication but not transcription. *J. Virol.* *77*, 5098–5108.
- Gerritz, S.W., Cianci, C., Kim, S., Pearce, B.C., Deminie, C., Discotto, L., McAuliffe, B., Minassian, B.F., Shi, S., Zhu, S., et al. (2011). Inhibition of influenza virus replication via small molecules that induce the formation of higher-order nucleoprotein oligomers. *Proc. Natl. Acad. Sci. U. S. A.* *108*, 15366–15371.
- Giegé, R., and Sauter, C. (2010). Biocrystallography: past, present, future. *HFSP J.* *4*, 109–121.
- Le Goffic, R., Bouguyon, E., Chevalier, C., Vidic, J., Da Costa, B., Leymarie, O., Bourdieu, C., Decamps, L., Dhorne-Pollet, S., and Delmas, B. (2010). Influenza A virus protein PB1-F2 exacerbates IFN-beta expression of human respiratory epithelial cells. *J. Immunol. Baltim. Md 1950* *185*, 4812–4823.
- González, S., and Ortín, J. (1999). Distinct regions of influenza virus PB1 polymerase subunit recognize vRNA and cRNA templates. *EMBO J.* *18*, 3767–3775.
- González, S., Zürcher, T., and Ortín, J. (1996). Identification of two separate domains in the influenza virus PB1 protein involved in the interaction with the PB2 and PA subunits: a model for the viral RNA polymerase structure. *Nucleic Acids Res.* *24*, 4456–4463.
- Gooskens, J., Jonges, M., Claas, E.C.J., Meijer, A., and Kroes, A.C.M. (2009). Prolonged Influenza Virus Infection during Lymphocytopenia and Frequent Detection of Drug-Resistant Viruses. *J. Infect. Dis.* *199*, 1435–1441.
- Gorai, T., Goto, H., Noda, T., Watanabe, T., Kozuka-Hata, H., Oyama, M., Takano, R., Neumann, G., Watanabe, S., and Kawaoka, Y. (2012). F1Fo-ATPase, F-type proton-translocating ATPase, at the plasma membrane is critical for efficient influenza virus budding. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 4615–4620.
- Görlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. *Annu. Rev. Cell Dev. Biol.* *15*, 607–660.
- Gorman, O.T., Donis, R.O., Kawaoka, Y., and Webster, R.G. (1990a). Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. *J. Virol.* *64*, 4893–4902.
- Gorman, O.T., Bean, W.J., Kawaoka, Y., and Webster, R.G. (1990b). Evolution of the nucleoprotein gene of influenza A virus. *J. Virol.* *64*, 1487–1497.
- Gorman, O.T., Bean, W.J., Kawaoka, Y., Donatelli, I., Guo, Y.J., and Webster, R.G. (1991). Evolution of influenza A virus nucleoprotein genes: implications for the origins of H1N1 human and classical swine viruses. *J. Virol.* *65*, 3704–3714.
- Goto, H., Muramoto, Y., Noda, T., and Kawaoka, Y. (2013). The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. *J. Virol.* *87*, 11316–11322.

- Greenspan, D., Palese, P., and Krystal, M. (1988). Two nuclear location signals in the influenza virus NS1 nonstructural protein. *J. Virol.* **62**, 3020–3026.
- GROG Situation de la grippe en France.
- Gui, X., Li, R., Zhang, X., Shen, C., Yu, H., Guo, X., Kang, Y., Chen, J., Chen, H., Chen, Y., et al. (2014). An important amino acid in nucleoprotein contributes to influenza A virus replication by interacting with polymerase PB2. *Virology* **464-465**, 11–20.
- Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W.H., Ortin, J., Hart, D.J., et al. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15**, 500–506.
- Guilligay, D., Chenavas, S., Cusack, S., Ruigrok, R, W, H., and Crepin, T. (2010). L'Arn polymérase du virus de la grippe : de la recherche fondamentale à l'élaboration de nouveaux antiviraux (Biofutur).
- Hale, B.G., Barclay, W.S., Randall, R.E., and Russell, R.J. (2008). Structure of an avian influenza A virus NS1 protein effector domain. *Virology* **378**, 1–5.
- Hale, B.G., Albrecht, R.A., and García-Sastre, A. (2010a). Innate immune evasion strategies of influenza viruses. *Future Microbiol.* **5**, 23–41.
- Hale, B.G., Kerry, P.S., Jackson, D., Precious, B.L., Gray, A., Killip, M.J., Randall, R.E., and Russell, R.J. (2010b). Structural insights into phosphoinositide 3-kinase activation by the influenza A virus NS1 protein. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1954–1959.
- Haller, O., Staeheli, P., and Kochs, G. (2009). Protective role of interferon-induced Mx GTPases against influenza viruses. *Rev. Sci. Tech. Int. Off. Epizoot.* **28**, 219–231.
- Hara, K., Shiota, M., Kido, H., Ohtsu, Y., Kashiwagi, T., Iwahashi, J., Hamada, N., Mizoue, K., Tsumura, N., Kato, H., et al. (2001). Influenza virus RNA polymerase PA subunit is a novel serine protease with Ser624 at the active site. *Genes Cells Devoted Mol. Cell. Mech.* **6**, 87–97.
- Hara, K., Schmidt, F.I., Crow, M., and Brownlee, G.G. (2006). Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80**, 7789–7798.
- Harrison, S.C. (2008). Viral membrane fusion. *Nat. Struct. Mol. Biol.* **15**, 690–698.
- Hause, B.M., Collin, E.A., Liu, R., Huang, B., Sheng, Z., Lu, W., Wang, D., Nelson, E.A., and Li, F. (2014). Characterization of a Novel Influenza Virus in Cattle and Swine: Proposal for a New Genus in the Orthomyxoviridae Family. *mBio* **5**.
- Hay, A.J., Lomniczi, B., Bellamy, A.R., and Skehel, J.J. (1977). Transcription of the influenza virus genome. *Virology* **83**, 337–355.
- Hay, A.J., Skehel, J.J., and McCauley, J. (1982). Characterization of influenza virus RNA complete transcripts. *Virology* **116**, 517–522.
- Hay, A.J., Wolstenholme, A.J., Skehel, J.J., and Smith, M.H. (1985). The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.* **4**, 3021–3024.
- Hayden, F.G., and Hay, A.J. (1992). Emergence and transmission of influenza A viruses resistant to amantadine and rimantadine. *Curr. Top. Microbiol. Immunol.* **176**, 119–130.
- He, X., Zhou, J., Bartlam, M., Zhang, R., Ma, J., Lou, Z., Li, X., Li, J., Joachimiak, A., Zeng, Z., et al. (2008). Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454**, 1123–1126.
- Heggeness, M.H., Smith, P.R., Ulmanen, I., Krug, R.M., and Choppin, P.W. (1982). Studies on the helical nucleocapsid of influenza virus. *Virology* **118**, 466–470.
- Helenius, A. (1992). Unpacking the incoming influenza virus. *Cell* **69**, 577–578.

- Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.* 9, 1665–1672.
- Henkel, M., Mitzner, D., Henklein, P., Meyer-Almes, F.-J., Moroni, A., Difrancesco, M.L., Henkes, L.M., Kreim, M., Kast, S.M., Schubert, U., et al. (2010). The proapoptotic influenza A virus protein PB1-F2 forms a nonselective ion channel. *PLoS One* 5, e11112.
- Herfst, S., Schrauwen, E.J.A., Linster, M., Chutinimitkul, S., de Wit, E., Munster, V.J., Sorrell, E.M., Bestebroer, T.M., Burke, D.F., Smith, D.J., et al. (2012). Airborne transmission of influenza A/H5N1 virus between ferrets. *Science* 336, 1534–1541.
- Herfst, S., Imai, M., Kawaoka, Y., and Fouchier, R. a. M. (2014). Avian influenza virus transmission to mammals. *Curr. Top. Microbiol. Immunol.* 385, 137–155.
- Herrler, G., Compans, R.W., and Meier-Ewert, H. (1979). A precursor glycoprotein in influenza C virus. *Virology* 99, 49–56.
- Herrler, G., Nagele, A., Meier-Ewert, H., Bhowan, A.S., and Compans, R.W. (1981). Isolation and structural analysis of influenza C virion glycoproteins. *Virology* 113, 439–451.
- Holsinger, L.J., and Lamb, R.A. (1991). Influenza virus M2 integral membrane protein is a homotetramer stabilized by formation of disulfide bonds. *Virology* 183, 32–43.
- Honda, A., Mizumoto, K., and Ishihama, A. (2002). Minimum molecular architectures for transcription and replication of the influenza virus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 13166–13171.
- Hsu, M.T., Parvin, J.D., Gupta, S., Krystal, M., and Palese, P. (1987). Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. *Proc. Natl. Acad. Sci. U. S. A.* 84, 8140–8144.
- Huang, T.S., Palese, P., and Krystal, M. (1990). Determination of influenza virus proteins required for genome replication. *J. Virol.* 64, 5669–5673.
- Huet, S., Avilov, S.V., Ferbitz, L., Daigle, N., Cusack, S., and Ellenberg, J. (2010). Nuclear import and assembly of influenza A virus RNA polymerase studied in live cells by fluorescence cross-correlation spectroscopy. *J. Virol.* 84, 1254–1264.
- Hutchinson, E.C., and Fodor, E. (2012). Nuclear import of the influenza A virus transcriptional machinery. *Vaccine* 30, 7353–7358.
- Hutchinson, E.C., Curran, M.D., Read, E.K., Gog, J.R., and Digard, P. (2008). Mutational analysis of cis-acting RNA signals in segment 7 of influenza A virus. *J. Virol.* 82, 11869–11879.
- Hutchinson, E.C., Wise, H.M., Kudryavtseva, K., Curran, M.D., and Digard, P. (2009). Characterisation of influenza A viruses with mutations in segment 5 packaging signals. *Vaccine* 27, 6270–6275.
- Hutchinson, E.C., Kirchbach, J.C. von, Gog, J.R., and Digard, P. (2010). Genome packaging in influenza A virus. *J. Gen. Virol.* 91, 313–328.
- Hutchinson, E.C., Orr, O.E., Liu, S.M., Engelhardt, O.G., and Fodor, E. (2011). Characterization of the interaction between the influenza A virus polymerase subunit PB1 and the host nuclear import factor Ran-binding protein 5. *J. Gen. Virol.* 92, 1859–1869.
- Ilari, A., and Savino, C. (2008). Protein structure determination by x-ray crystallography. *Methods Mol. Biol. Clifton NJ* 452, 63–87.
- Imai, M., Watanabe, T., Hatta, M., Das, S.C., Ozawa, M., Shinya, K., Zhong, G., Hanson, A., Katsura, H., Watanabe, S., et al. (2012). Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature* 486, 420–428.

- INSERM (2010). Grippe A/H1N1 : Bilan et perspectives de la recherche à un an. Available at <http://www.inserm.fr/espace-journalistes/grippe-a-h1n1-bilan-et-perspectives-de-la-recherche-a-un-an>.
- INSERM (2012). Dossier d'information GRIPPE. Available at <http://www.inserm.fr/thematiques/microbiologie-et-maladies-infectieuses/dossiers-d-information/grippe>.
- InVS (2015). Données de surveillance pour la grippe en France. Available at <http://www.invs.sante.fr/Dossiers-thematiques/Maladies-infectieuses/Maladies-a-prevention-vaccinale/Grippe/Grippe-generalites/Donnees-de-surveillance>.
- Jagger, B.W., Wise, H.M., Kash, J.C., Walters, K.-A., Wills, N.M., Xiao, Y.-L., Dunfee, R.L., Schwartzman, L.M., Ozinsky, A., Bell, G.L., et al. (2012). An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 337, 199–204.
- Jäkel, S., and Görlich, D. (1998). Importin beta, transportin, RanBP5 and RanBP7 mediate nuclear import of ribosomal proteins in mammalian cells. *EMBO J.* 17, 4491–4502.
- Jäkel, S., Mingot, J.-M., Schwarzmaier, P., Hartmann, E., and Görlich, D. (2002). Importins fulfil a dual function as nuclear import receptors and cytoplasmic chaperones for exposed basic domains. *EMBO J.* 21, 377–386.
- Jennings, P.A., Finch, J.T., Winter, G., and Robertson, J.S. (1983). Does the higher order structure of the influenza virus ribonucleoprotein guide sequence rearrangements in influenza viral RNA? *Cell* 34, 619–627.
- Johnson, N.P.A.S., and Mueller, J. (2002). Updating the accounts: global mortality of the 1918-1920 “Spanish” influenza pandemic. *Bull. Hist. Med.* 76, 105–115.
- Jones, I.M., Reay, P.A., and Philpott, K.L. (1986). Nuclear location of all three influenza polymerase proteins and a nuclear signal in polymerase PB2. *EMBO J.* 5, 2371–2376.
- de Jong, J.C., Claas, E.C., Osterhaus, A.D., Webster, R.G., and Lim, W.L. (1997). A pandemic warning? *Nature* 389, 554.
- Jorba, N., Coloma, R., and Ortín, J. (2009). Genetic trans-complementation establishes a new model for influenza virus RNA transcription and replication. *PLoS Pathog.* 5, e1000462.
- Junge, C. (2011). Morbidity: A personal response. *Nature* 480, S14–S15.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* 26, 795–800.
- Kao, R.Y., Yang, D., Lau, L.-S., Tsui, W.H.W., Hu, L., Dai, J., Chan, M.-P., Chan, C.-M., Wang, P., Zheng, B.-J., et al. (2010). Identification of influenza A nucleoprotein as an antiviral target. *Nat. Biotechnol.* 28, 600–605.
- Kaplan, B.S., and Webby, R.J. (2013). The avian and mammalian host range of highly pathogenic avian H5N1 influenza. *Virus Res.* 178, 3–11.
- Kawaguchi, A., and Nagata, K. (2007). De novo replication of the influenza virus RNA genome is regulated by DNA replicative helicase, MCM. *EMBO J.* 26, 4566–4575.
- Kawai, T., and Akira, S. (2011). Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* 34, 637–650.
- Keegan, R.M., and Winn, M.D. (2007). Automated search-model discovery and preparation for structure solution by molecular replacement. *Acta Crystallogr. D Biol. Crystallogr.* 63, 447–457.
- Kendal, A.P. (1975). A comparison of “influenza C” with prototype myxoviruses: receptor-destroying activity (neuraminidase) and structural polypeptides. *Virology* 65, 87–99.
- Kerry, P.S., Ayllon, J., Taylor, M.A., Hass, C., Lewis, A., García-Sastre, A., Randall, R.E., Hale, B.G., and Russell, R.J. (2011). A transient homotypic interaction model for the influenza A virus NS1 protein effector domain. *PloS One* 6, e17946.

- Kilbourne, E.D., and Murphy, J.S. (1960). Genetic studies of influenza viruses. I. Viral morphology and growth capacity as exchangeable genetic traits. Rapid in ovo adaptation of early passage Asian strain isolates by combination with PR8. *J. Exp. Med.* *111*, 387–406.
- Kistner, O., Müller, H., Becht, H., and Scholtissek, C. (1985). Phosphopeptide fingerprints of nucleoproteins of various influenza A virus strains grown in different host cells. *J. Gen. Virol.* *66* (Pt 3), 465–472.
- Kistner, O., Müller, K., and Scholtissek, C. (1989). Differential Phosphorylation of the Nucleoprotein of Influenza A Viruses. *J. Gen. Virol.* *70*, 2421–2431.
- Klucevsek, K., Daley, J., Darshan, M.S., Bordeaux, J., and Moroiaru, J. (2006). Nuclear import strategies of high-risk HPV18 L2 minor capsid protein. *Virology* *352*, 200–208.
- Klumpp, K., Ruigrok, R.W., and Baudin, F. (1997). Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.* *16*, 1248–1257.
- Kobayashi, J., and Matsuura, Y. (2013). Structural Basis for Cell-Cycle-Dependent Nuclear Import Mediated by the Karyopherin Kap121p. *J. Mol. Biol.* *425*, 1852–1868.
- Kobayashi, J., Hirano, H., and Matsuura, Y. (2015). Crystal structure of the karyopherin Kap121p bound to the extreme C-terminus of the protein phosphatase Cdc14p. *Biochem. Biophys. Res. Commun.* *463*, 309–314.
- Kowalinski, E., Zubieta, C., Wolkerstorfer, A., Szolar, O.H.J., Ruigrok, R.W.H., and Cusack, S. (2012). Structural analysis of specific metal chelating inhibitor binding to the endonuclease domain of influenza pH1N1 (2009) polymerase. *PLoS Pathog.* *8*, e1002831.
- Koyama, K., Takahashi, M., Oitate, M., Nakai, N., Takakusa, H., Miura, S., and Okazaki, O. (2009). CS-8958, a prodrug of the novel neuraminidase inhibitor R-125489, demonstrates a favorable long-retention profile in the mouse respiratory tract. *Antimicrob. Agents Chemother.* *53*, 4845–4851.
- Krawczyk, E., Hanover, J.A., Schlegel, R., and Supryniewicz, F.A. (2008). Karyopherin beta3: a new cellular target for the HPV-16 E5 oncoprotein. *Biochem. Biophys. Res. Commun.* *371*, 684–688.
- Krug, R.M., Ueda, M., and Palese, P. (1975). Temperature-sensitive mutants of influenza WSN virus defective in virus-specific RNA synthesis. *J. Virol.* *16*, 790–796.
- Krug, R.M., Broni, B.A., and Bouloy, M. (1979). Are the 5' ends of influenza viral mRNAs synthesized in vivo donated by host mRNAs? *Cell* *18*, 329–334.
- Krumbholz, A., Philipps, A., Oehring, H., Schwarzer, K., Eitner, A., Wutzler, P., and Zell, R. (2010). Current knowledge on PB1-F2 of influenza A viruses. *Med. Microbiol. Immunol. (Berl.)* *200*, 69–75.
- Kuzuhara, T., Kise, D., Yoshida, H., Horita, T., Murazaki, Y., Nishimura, A., Echigo, N., Utsunomiya, H., and Tsuge, H. (2009). Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J. Biol. Chem.* *284*, 6855–6860.
- Lamb, R.A., Zebedee, S.L., and Richardson, C.D. (1985). Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* *40*, 627–633.
- Lee, M.T.M., Bishop, K., Medcalf, L., Elton, D., Digard, P., and Tiley, L. (2002). Definition of the minimal viral components required for the initiation of unprimed RNA synthesis by influenza virus RNA polymerase. *Nucleic Acids Res.* *30*, 429–438.
- Lee, S.J., Matsuura, Y., Liu, S.M., and Stewart, M. (2005). Structural basis for nuclear import complex dissociation by RanGTP. *Nature* *435*, 693–696.
- Levy, P.E. (2012). French Economic Evaluations of Influenza and Influenza Vaccination. *PharmacoEconomics* *9*, 62–66.

- Li, L., Chang, S., Xiang, J., Li, Q., Liang, H., Tang, Y., and Liu, Y. (2012). Screen anti-influenza lead compounds that target the PA(C) subunit of H5N1 viral RNA polymerase. *PLoS One* 7, e35234.
- LiCata, V.J., and Wowor, A.J. (2008). Applications of fluorescence anisotropy to the study of protein-DNA interactions. *Methods Cell Biol.* 84, 243–262.
- Liu, J., Lynch, P.A., Chien, C.Y., Montelione, G.T., Krug, R.M., and Berman, H.M. (1997). Crystal structure of the unique RNA-binding domain of the influenza virus NS1 protein. *Nat. Struct. Biol.* 4, 896–899.
- Loo, Y.-M., and Gale, M. (2007). Influenza: Fatal immunity and the 1918 virus. *Nature* 445, 267–268.
- Loucaides, E.M., von Kirchbach, J.C., Foeglein, A., Sharps, J., Fodor, E., and Digard, P. (2009). Nuclear dynamics of influenza A virus ribonucleoproteins revealed by live-cell imaging studies. *Virology* 394, 154–163.
- Luytjes, W., Krystal, M., Enami, M., Parvin, J.D., and Palese, P. (1989). Amplification, expression, and packaging of foreign gene by influenza virus. *Cell* 59, 1107–1113.
- Ma, W., Vincent, A.L., Gramer, M.R., Brockwell, C.B., Lager, K.M., Janke, B.H., Gauger, P.C., Patnayak, D.P., Webby, R.J., and Richt, J.A. (2007). Identification of H2N3 influenza A viruses from swine in the United States. *Proc. Natl. Acad. Sci. U. S. A.* 104, 20949–20954.
- Mahy, B.W.J. (1983). Mutants of Influenza Virus. In *Genetics of Influenza Viruses*, P.D.P. Palese, and P.D.D.W. Kingsbury, eds. (Springer Vienna), pp. 192–254.
- Mänz, B., Brunotte, L., Reuther, P., and Schwemmler, M. (2012). Adaptive mutations in NEP compensate for defective H5N1 RNA replication in cultured human cells. *Nat. Commun.* 3, 802.
- Marfori, M., Mynott, A., Ellis, J.J., Mehdi, A.M., Saunders, N.F.W., Curmi, P.M., Forwood, J.K., Bodén, M., and Kobe, B. (2011). Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim. Biophys. Acta* 1813, 1562–1577.
- Marión, R.M., Zürcher, T., de la Luna, S., and Ortín, J. (1997). Influenza virus NS1 protein interacts with viral transcription-replication complexes in vivo. *J. Gen. Virol.* 78 (Pt 10), 2447–2451.
- Marsh, G.A., Rabadán, R., Levine, A.J., and Palese, P. (2008). Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. *J. Virol.* 82, 2295–2304.
- Martin, K., and Helenius, A. (1991). Transport of incoming influenza virus nucleocapsids into the nucleus. *J. Virol.* 65, 232–244.
- Martín-Benito, J., and Ortín, J. (2013). Chapter Four - Influenza Virus Transcription and Replication. In *Advances in Virus Research*, K.M. and F.A. Murphy, ed. (Academic Press), pp. 113–137.
- Martín-Benito, J., Area, E., Ortega, J., Llorca, O., Valpuesta, J.M., Carrascosa, J.L., and Ortín, J. (2001). Three-dimensional reconstruction of a recombinant influenza virus ribonucleoprotein particle. *EMBO Rep.* 2, 313–317.
- Matlin, K.S., Reggio, H., Helenius, A., and Simons, K. (1981). Infectious entry pathway of influenza virus in a canine kidney cell line. *J. Cell Biol.* 91, 601–613.
- Mayer, D., Molawi, K., Martínez-Sobrido, L., Ghanem, A., Thomas, S., Baginsky, S., Grossmann, J., García-Sastre, A., and Schwemmler, M. (2007). Identification of cellular interaction partners of the influenza virus ribonucleoprotein complex and polymerase complex using proteomic-based approaches. *J. Proteome Res.* 6, 672–682.
- Mazur, I., Anhlan, D., Mitzner, D., Wixler, L., Schubert, U., and Ludwig, S. (2008). The proapoptotic influenza A virus protein PB1-F2 regulates viral polymerase activity by interaction with the PB1 protein. *Cell. Microbiol.* 10, 1140–1152.

- McAuley, J.L., Chipuk, J.E., Boyd, K.L., Van De Velde, N., Green, D.R., and McCullers, J.A. (2010). PB1-F2 proteins from H5N1 and 20 century pandemic influenza viruses cause immunopathology. *PLoS Pathog.* *6*, e1001014.
- McCown, M.F., and Pekosz, A. (2006). Distinct domains of the influenza A virus M2 protein cytoplasmic tail mediate binding to the M1 protein and facilitate infectious virus production. *J. Virol.* *80*, 8178–8189.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). *Phaser* crystallographic software. *J. Appl. Crystallogr.* *40*, 658–674.
- McDevitt, J., Rudnick, S., First, M., and Spengler, J. (2010). Role of Absolute Humidity in the Inactivation of Influenza Viruses on Stainless Steel Surfaces at Elevated Temperatures. *Appl. Environ. Microbiol.* *76*, 3943–3947.
- Mena, I., Jambrina, E., Albo, C., Perales, B., Ortín, J., Arrese, M., Vallejo, D., and Portela, A. (1999). Mutational analysis of influenza A virus nucleoprotein: identification of mutations that affect RNA replication. *J. Virol.* *73*, 1186–1194.
- Micheal Clodfelter (2002). *Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1500-2000* (McFarland & Company).
- Min, J.-Y., and Krug, R.M. (2006). The primary function of RNA binding by the influenza A virus NS1 protein in infected cells: Inhibiting the 2'-5' oligo (A) synthetase/RNase L pathway. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 7100–7105.
- Min, J.-Y., Li, S., Sen, G.C., and Krug, R.M. (2007). A site on the influenza A virus NS1 protein mediates both inhibition of PKR activation and temporal regulation of viral RNA synthesis. *Virology* *363*, 236–243.
- Mitamura, K., and Sugaya, N. (2006). [Diagnosis and Treatment of influenza--clinical investigation on viral shedding in children with influenza]. *Uirusu* *56*, 109–116.
- Moeller, A., Kirchdoerfer, R.N., Potter, C.S., Carragher, B., and Wilson, I.A. (2012). Organization of the influenza virus replication machinery. *Science* *338*, 1631–1634.
- Mondal, A., Potts, G.K., Dawson, A.R., Coon, J.J., and Mehle, A. (2015). Phosphorylation at the Homotypic Interface Regulates Nucleoprotein Oligomerization and Assembly of the Influenza Virus Replication Machinery. *PLoS Pathog* *11*, e1004826.
- Monod, A., Swale, C., Tarus, B., Tissot, A., Delmas, B., Ruigrok, R.W., Crépin, T., and Slama-Schwok, A. (2015). Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza. *Expert Opin. Drug Discov.* *10*, 345–371.
- Morens, D.M., Taubenberger, J.K., Folkers, G.K., and Fauci, A.S. (2010). Pandemic Influenza's 500th Anniversary. *Clin. Infect. Dis.* *51*, 1442–1444.
- Mukaigawa, J., and Nayak, D.P. (1991). Two signals mediate nuclear localization of influenza virus (A/WSN/33) polymerase basic protein 2. *J. Virol.* *65*, 245–253.
- Müller, R., Poch, O., Delarue, M., Bishop, D.H., and Bouloy, M. (1994). Rift Valley fever virus L segment: correction of the sequence and possible functional role of newly identified regions conserved in RNA-dependent polymerases. *J. Gen. Virol.* *75* (Pt 6), 1345–1352.
- Munier, S., Moisy, D., Marc, D., and Naffakh, N. (2010). [Interspecies transmission, adaptation to humans and pathogenicity of animal influenza viruses]. *Pathol. Biol. (Paris)* *58*, e59–e68.
- Muramoto, Y., Takada, A., Fujii, K., Noda, T., Iwatsuki-Horimoto, K., Watanabe, S., Horimoto, T., Kida, H., and Kawaoka, Y. (2006). Hierarchy among viral RNA (vRNA) segments in their role in vRNA incorporation into influenza A virions. *J. Virol.* *80*, 2318–2325.
- Muramoto, Y., Noda, T., Kawakami, E., Akkina, R., and Kawaoka, Y. (2013). Identification of novel influenza A virus proteins translated from PA mRNA. *J. Virol.* *87*, 2455–2462.

- Muratore, G., Goracci, L., Mercorelli, B., Foeglein, Á., Digard, P., Cruciani, G., Palù, G., and Loregian, A. (2012). Small molecule inhibitors of influenza A and B viruses that act by disrupting subunit interactions of the viral polymerase. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 6247–6252.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* *53*, 240–255.
- Murti, K.G., and Webster, R.G. (1986). Distribution of hemagglutinin and neuraminidase on influenza virions as revealed by immunoelectron microscopy. *Virology* *149*, 36–43.
- Nabel, G.J., Wei, C.-J., and Ledgerwood, J.E. (2011). Vaccinate for the next H2N2 pandemic now. *Nature* *471*, 157–158.
- Naffakh, N., Tomoiu, A., Rameix-Welti, M.-A., and van der Werf, S. (2008). Host restriction of avian influenza viruses at the level of the ribonucleoproteins. *Annu. Rev. Microbiol.* *62*, 403–424.
- Naito, T., Momose, F., Kawaguchi, A., and Nagata, K. (2007). Involvement of Hsp90 in assembly and nuclear import of influenza virus RNA polymerase subunits. *J. Virol.* *81*, 1339–1349.
- Nakajima, K., Desselberger, U., and Palese, P. (1978). Recent human influenza A (H1N1) viruses are closely related genetically to strains isolated in 1950. *Nature* *274*, 334–339.
- Nath, S.T., and Nayak, D.P. (1990). Function of two discrete regions is required for nuclear localization of polymerase basic protein 1 of A/WSN/33 influenza virus (H1 N1). *Mol. Cell. Biol.* *10*, 4139–4145.
- Nayak, D.P., and Hui, E.K.W. (2004). The role of lipid microdomains in virus biology. *Subcell. Biochem.* *37*, 443–491.
- Nayak, D.P., Hui, E.K.-W., and Barman, S. (2004). Assembly and budding of influenza virus. *Virus Res.* *106*, 147–165.
- Nayak, D.P., Balogun, R.A., Yamada, H., Zhou, Z.H., and Barman, S. (2009). Influenza virus morphogenesis and budding. *Virus Res.* *143*, 147–161.
- Nelson, L.M., Rose, R.C., and Moroianu, J. (2003). The L1 major capsid protein of human papillomavirus type 11 interacts with Kap beta2 and Kap beta3 nuclear import receptors. *Virology* *306*, 162–169.
- Nemeroff, M.E., Qian, X.Y., and Krug, R.M. (1995). The influenza virus NS1 protein forms multimers in vitro and in vivo. *Virology* *212*, 422–428.
- Nemeroff, M.E., Barabino, S.M., Li, Y., Keller, W., and Krug, R.M. (1998). Influenza virus NS1 protein interacts with the cellular 30 kDa subunit of CPSF and inhibits 3' end formation of cellular pre-mRNAs. *Mol. Cell* *1*, 991–1000.
- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat. Genet.* *20*, 46–50.
- Neumann, G., Hughes, M.T., and Kawaoka, Y. (2000). Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J.* *19*, 6751–6758.
- Ng, A.K.-L., Zhang, H., Tan, K., Li, Z., Liu, J., Chan, P.K.-S., Li, S.-M., Chan, W.-Y., Au, S.W.-N., Joachimiak, A., et al. (2008). Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* *22*, 3638–3647.
- Ng, A.K.-L., Lam, M.K.-H., Zhang, H., Liu, J., Au, S.W.-N., Chan, P.K.-S., Wang, J., and Shaw, P.-C. (2012). Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J. Virol.* *86*, 6758–6767.

- Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C., and Berger, I. (2014). Multiprotein complex production in insect cells by using polyproteins. *Methods Mol. Biol. Clifton NJ* 1091, 131–141.
- Nieto, A., de la Luna, S., Bárcena, J., Portela, A., Valcárcel, J., Melero, J.A., and Ortín, J. (1992). Nuclear transport of influenza virus polymerase PA protein. *Virus Res.* 24, 65–75.
- Nieto, A., de la Luna, S., Bárcena, J., Portela, A., and Ortín, J. (1994). Complex structure of the nuclear translocation signal of influenza virus polymerase PA subunit. *J. Gen. Virol.* 75 (Pt 1), 29–36.
- Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R.H., and Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* 439, 490–492.
- Noda, T., Sugita, Y., Aoyama, K., Hirase, A., Kawakami, E., Miyazawa, A., Sagara, H., and Kawaoka, Y. (2012). Three-dimensional analysis of ribonucleoprotein complexes in influenza A virus. *Nat. Commun.* 3, 639.
- Normile, D. (2006). Avian influenza. Human transmission but no pandemic in Indonesia. *Science* 312, 1855.
- Noti, J.D., Blachere, F.M., McMillen, C.M., Lindsley, W.G., Kashon, M.L., Slaughter, D.R., and Beezhold, D.H. (2013). High Humidity Leads to Loss of Infectious Influenza Virus from Simulated Coughs. *PLoS ONE* 8, e57485.
- Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team, Dawood, F.S., Jain, S., Finelli, L., Shaw, M.W., Lindstrom, S., Garten, R.J., Gubareva, L.V., Xu, X., Bridges, C.B., et al. (2009). Emergence of a novel swine-origin influenza A (H1N1) virus in humans. *N. Engl. J. Med.* 360, 2605–2615.
- Obayashi, E., Yoshida, H., Kawai, F., Shibayama, N., Kawaguchi, A., Nagata, K., Tame, J.R.H., and Park, S.-Y. (2008). The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* 454, 1127–1131.
- Oestereich, L., Lüdtke, A., Wurr, S., Rieger, T., Muñoz-Fontela, C., and Günther, S. (2014). Successful treatment of advanced Ebola virus infection with T-705 (favipiravir) in a small animal model. *Antiviral Res.* 105, 17–21.
- O'Neill, R.E., Jaskunas, R., Blobel, G., Palese, P., and Moroianu, J. (1995). Nuclear import of influenza virus RNA can be mediated by viral nucleoprotein and transport factors required for protein import. *J. Biol. Chem.* 270, 22701–22704.
- O'Neill, R.E., Talon, J., and Palese, P. (1998). The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO J.* 17, 288–296.
- Ortega, J., Martín-Benito, J., Zürcher, T., Valpuesta, J.M., Carrascosa, J.L., and Ortín, J. (2000). Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *J. Virol.* 74, 156–163.
- Ozawa, M., Fujii, K., Muramoto, Y., Yamada, S., Yamayoshi, S., Takada, A., Goto, H., Horimoto, T., and Kawaoka, Y. (2007). Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. *J. Virol.* 81, 30–41.
- Palese, P., and Compans, R.W. (1976). Inhibition of influenza virus replication in tissue culture by 2-deoxy-2,3-dehydro-N-trifluoroacetylneuraminic acid (FANA): mechanism of action. *J. Gen. Virol.* 33, 159–163.
- Palese, P., and Schulman, J.L. (1976). Mapping of the influenza virus genome: identification of the hemagglutinin and the neuraminidase genes. *Proc. Natl. Acad. Sci. U. S. A.* 73, 2142–2146.

- Palese, P., Tobita, K., Ueda, M., and Compans, R.W. (1974). Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* 61, 397–410.
- Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., et al. (2001). High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* 6, 429–440.
- Pape, T., and Schneider, T.R. (2004). *HKL2MAP*: a graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Crystallogr.* 37, 843–844.
- Paterson, D., and Fodor, E. (2012). Emerging roles for the influenza A virus nuclear export protein (NEP). *PLoS Pathog.* 8, e1003019.
- Patterson, K.D., and Pyle, G.F. (1991). The geography and mortality of the 1918 influenza pandemic. *Bull. Hist. Med.* 65, 4–21.
- Peiris, J.S.M., Yu, W.C., Leung, C.W., Cheung, C.Y., Ng, W.F., Nicholls, J.M., Ng, T.K., Chan, K.H., Lai, S.T., Lim, W.L., et al. (2004). Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* 363, 617–619.
- Perales, B., Sanz-Ezquerro, J.J., Gastaminza, P., Ortega, J., Santarén, J.F., Ortín, J., and Nieto, A. (2000). The replication activity of influenza virus polymerase is linked to the capacity of the PA subunit to induce proteolysis. *J. Virol.* 74, 1307–1312.
- Pérez, D.R., and Donis, R.O. (1995). A 48-amino-acid region of influenza A virus PB1 protein is sufficient for complex formation with PA. *J. Virol.* 69, 6932–6939.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Pflug, A., Guilligay, D., Reich, S., and Cusack, S. (2014). Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 516, 355–360.
- Pielak, R.M., and Chou, J.J. (2011). Influenza M2 proton channels. *Biochim. Biophys. Acta BBA - Biomembr.* 1808, 522–529.
- Plotch, S.J., Bouloy, M., and Krug, R.M. (1979). Transfer of 5'-terminal cap of globin mRNA to influenza viral complementary RNA during transcription in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 76, 1618–1622.
- Plotch, S.J., Bouloy, M., Ulmanen, I., and Krug, R.M. (1981). A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* 23, 847–858.
- Plotkin, J.B., and Dushoff, J. (2003). Codon bias and frequency-dependent selection on the hemagglutinin epitopes of influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* 100, 7152–7157.
- Poch, O., Sauvaget, I., Delarue, M., and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* 8, 3867–3874.
- Pons, M.W., Schulze, I.T., Hirst, G.K., and Hauser, R. (1969). Isolation and characterization of the ribonucleoprotein of influenza virus. *Virology* 39, 250–259.
- Poole, E., Elton, D., Medcalf, L., and Digard, P. (2004). Functional domains of the influenza A virus PB2 protein: identification of NP- and PB1-binding sites. *Virology* 321, 120–133.
- Poon, L.L., Pritlove, D.C., Fodor, E., and Brownlee, G.G. (1999). Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J. Virol.* 73, 3473–3476.
- Portela, A., and Digard, P. (2002). The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J. Gen. Virol.* 83, 723–734.

- Pritlove, D.C., Poon, L.L., Devenish, L.J., Leahy, M.B., and Brownlee, G.G. (1999). A hairpin loop at the 5' end of influenza A virus virion RNA is required for synthesis of poly(A)⁺ mRNA in vitro. *J. Virol.* *73*, 2109–2114.
- Pumroy, R.A., Nardozi, J.D., Hart, D.J., Root, M.J., and Cingolani, G. (2012). Nucleoporin Nup50 stabilizes closed conformation of armadillo repeat 10 in importin α 5. *J. Biol. Chem.* *287*, 2022–2031.
- Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* *40*, 191–285.
- Qiu, Y., Nemeroff, M., and Krug, R.M. (1995). The influenza virus NS1 protein binds to a specific region in human U6 snRNA and inhibits U6-U2 and U6-U4 snRNA interactions during splicing. *RNA N. Y. N* *1*, 304–316.
- Rambo, R.P., and Tainer, J.A. (2013). Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* *496*, 477–481.
- Randall, R.E., and Goodbourn, S. (2008). Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J. Gen. Virol.* *89*, 1–47.
- Regan, J.F., Liang, Y., and Parslow, T.G. (2006). Defective assembly of influenza A virus due to a mutation in the polymerase subunit PA. *J. Virol.* *80*, 252–261.
- Reguera, J., Malet, H., Weber, F., and Cusack, S. (2013). Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 7246–7251.
- Reich, S., Guilligay, D., Pflug, A., Malet, H., Berger, I., Crépin, T., Hart, D., Lunardi, T., Nanao, M., Ruigrok, R.W.H., et al. (2014). Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* *516*, 361–366.
- Reperant, L.A., Rimmelzwaan, G.F., and Kuiken, T. (2009). Avian influenza viruses in mammals. *Rev. Sci. Tech. Int. Off. Epizoot.* *28*, 137–159.
- Resa-Infante, P., Jorba, N., Zamarreño, N., Fernández, Y., Juárez, S., and Ortín, J. (2008). The host-dependent interaction of alpha-importins with influenza PB2 polymerase subunit is required for virus RNA replication. *PLoS One* *3*, e3904.
- Resa-Infante, P., Recuero-Checa, M.A., Zamarreño, N., Llorca, O., and Ortín, J. (2010). Structural and functional characterization of an influenza virus RNA polymerase-genomic RNA complex. *J. Virol.* *84*, 10477–10487.
- Resa-Infante, P., Jorba, N., Coloma, R., and Ortín, J. (2011). The influenza RNA synthesis machine. *RNA Biol.* *8*, 207–215.
- Resa-Infante, P., Thieme, R., Ernst, T., Arck, P.C., Ittrich, H., Reimer, R., and Gabriel, G. (2014). Importin- α 7 is required for enhanced influenza A virus replication in the alveolar epithelium and severe lung damage in mice. *J. Virol.* *88*, 8166–8179.
- Richardson, J.C., and Akkina, R.K. (1991). NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. *Arch. Virol.* *116*, 69–80.
- Ritchey, M.B., Palese, P., and Kilbourne, E.D. (1976a). RNAs of influenza A, B, and C viruses. *J. Virol.* *18*, 738–744.
- Ritchey, M.B., Palese, P., and Schulman, J.L. (1976b). Mapping of the influenza virus genome. III. Identification of genes coding for nucleoprotein, membrane protein, and nonstructural protein. *J. Virol.* *20*, 307–313.
- Robb, N.C., Smith, M., Vreede, F.T., and Fodor, E. (2009). NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. *J. Gen. Virol.* *90*, 1398–1407.

- Rodriguez, A., Pérez-González, A., and Nieto, A. (2007). Influenza virus infection causes specific degradation of the largest subunit of cellular RNA polymerase II. *J. Virol.* *81*, 5315–5324.
- Ross, A.E., Vuica, M., and Desiderio, S. (2003). Overlapping signals for protein degradation and nuclear localization define a role for intrinsic RAG-2 nuclear uptake in dividing cells. *Mol. Cell. Biol.* *23*, 5308–5319.
- Rossman, J.S., and Lamb, R.A. (2011). Influenza Virus Assembly and Budding. *Virology* *411*, 229–236.
- Rossmann, M.G., and Blow, D.M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* *15*, 24–31.
- Rudolph, M.G., Kraus, I., Dickmanns, A., Eickmann, M., Garten, W., and Ficner, R. (2003). Crystal structure of the borna disease virus nucleoprotein. *Struct. Lond. Engl.* *11*, 1219–1226.
- Ruigrok, R.W., and Baudin, F. (1995). Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. *J. Gen. Virol.* *76* (Pt 4), 1009–1014.
- Ruigrok, R.W., Calder, L.J., and Wharton, S.A. (1989). Electron microscopy of the influenza virus submembranal structure. *Virology* *173*, 311–316.
- Ruigrok, R.W., Barge, A., Durrer, P., Brunner, J., Ma, K., and Whittaker, G.R. (2000). Membrane interaction of influenza virus M1 protein. *Virology* *267*, 289–298.
- Ruigrok, R.W.H., Crépin, T., and Kolakofsky, D. (2011). Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr. Opin. Microbiol.* *14*, 504–510.
- Russell, C.A., Jones, T.C., Barr, I.G., Cox, N.J., Garten, R.J., Gregory, V., Gust, I.D., Hampson, A.W., Hay, A.J., Hurt, A.C., et al. (2008). The Global Circulation of Seasonal Influenza A (H3N2) Viruses. *Science* *320*, 340–346.
- Russell, C.A., Fonville, J.M., Brown, A.E.X., Burke, D.F., Smith, D.L., James, S.L., Herfst, S., van Boheemen, S., Linster, M., Schrauwen, E.J., et al. (2012). The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science* *336*, 1541–1547.
- Salk, J.E., Menke, W.J., and Francis, T. (1945). A Clinical, Epidemiological and Immunological Evaluation of Vaccination Against Epidemic Influenza. *Am. J. Epidemiol.* *42*, 57–93.
- Samson, M., Pizzorno, A., Abed, Y., and Boivin, G. (2013). Influenza virus resistance to neuraminidase inhibitors. *Antiviral Res.* *98*, 174–185.
- Sanz-Ezquerro, J.J., Zürcher, T., de la Luna, S., Ortín, J., and Nieto, A. (1996). The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *J. Virol.* *70*, 1905–1911.
- Satterly, N., Tsai, P.-L., van Deursen, J., Nussenzveig, D.R., Wang, Y., Faria, P.A., Levay, A., Levy, D.E., and Fontoura, B.M.A. (2007). Influenza virus targets the mRNA export machinery and the nuclear pore complex. *Proc. Natl. Acad. Sci. U. S. A.* *104*, 1853–1858.
- Scheiffele, P., Rietveld, A., Wilk, T., and Simons, K. (1999). Influenza Viruses Select Ordered Lipid Domains during Budding from the Plasma Membrane. *J. Biol. Chem.* *274*, 2038–2044.
- Schmitt, A.P., and Lamb, R.A. (2005). Influenza Virus Assembly and Budding at the Viral Budozone. In *Advances in Virus Research*, P. Roy, ed. (Academic Press), pp. 383–416.
- Schnell, J.R., and Chou, J.J. (2008). Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* *451*, 591–595.

- Scholtissek, C., von Hoyningen, V., and Rott, R. (1978). Genetic relatedness between the new 1977 epidemic strains (H1N1) of influenza and human influenza strains isolated between 1947 and 1957 (H1N1). *Virology* 89, 613–617.
- Schrauwen, E.J.A., de Graaf, M., Herfst, S., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2014). Determinants of virulence of influenza A virus. *Eur. J. Clin. Microbiol. Infect. Dis. Off. Publ. Eur. Soc. Clin. Microbiol.* 33, 479–490.
- Schulman, J.L., and Palese, P. (1975). Susceptibility of different strains of influenza A virus to the inhibitory effects of 2-deoxy-2,3-dehydro-n-trifluoroacetylneuraminic acid (FANA). *Virology* 63, 98–104.
- Selman, M., Dankar, S.K., Forbes, N.E., Jia, J.-J., and Brown, E.G. (2012). Adaptive mutation in influenza A virus non-structural gene is linked to host switching and induces a novel protein by alternative splicing. *Emerg. Microbes Infect.* 1, e42.
- Semenyuk, A.V., and Svergun, D.I. (1991). GNOM – a program package for small-angle scattering data processing. *J. Appl. Crystallogr.* 24, 537–540.
- Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas', A.F., and Gilmanshin, R.I. (1991). Study of the “molten globule” intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* 31, 119–128.
- Sénat Français (2011). Rapport d'information. Au nom de la commission des affaires sociales (1) sur l'étude de la Cour des comptes relative à l'utilisation des fonds mobilisés pour la lutte contre la pandémie grippale A(H1N1).
- Sha, B., and Luo, M. (1997). Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nat. Struct. Biol.* 4, 239–244.
- Shapiro, G.I., and Krug, R.M. (1988). Influenza virus RNA replication in vitro: synthesis of viral template RNAs and virion RNAs in the absence of an added primer. *J. Virol.* 62, 2285–2290.
- Shapiro, G.I., Gurney, T., and Krug, R.M. (1987). Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *J. Virol.* 61, 764–773.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. A* 64, 112–122.
- Sheldrick, G.M. (2010). Experimental phasing with *SHELXC / D / E*: combining chain tracing with density modification. *Acta Crystallogr. D Biol. Crystallogr.* 66, 479–485.
- Sheng, Z.-M., Chertow, D.S., Ambroggio, X., McCall, S., Przygodzki, R.M., Cunningham, R.E., Maximova, O.A., Kash, J.C., Morens, D.M., and Taubenberger, J.K. (2011). Autopsy series of 68 cases dying before and during the 1918 influenza pandemic peak. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16416–16421.
- Sheu, T.G., Fry, A.M., Garten, R.J., Deyde, V.M., Shwe, T., Bullion, L., Peebles, P.J., Li, Y., Klimov, A.I., and Gubareva, L.V. (2011). Dual resistance to adamantanes and oseltamivir among seasonal influenza A(H1N1) viruses: 2008-2010. *J. Infect. Dis.* 203, 13–17.
- Shi, M., Jagger, B.W., Wise, H.M., Digard, P., Holmes, E.C., and Taubenberger, J.K. (2012). Evolutionary conservation of the PA-X open reading frame in segment 3 of influenza A virus. *J. Virol.* 86, 12411–12413.
- Shinya, K., Watanabe, S., Ito, T., Kasai, N., and Kawaoka, Y. (2007). Adaptation of an H7N7 equine influenza A virus in mice. *J. Gen. Virol.* 88, 547–553.
- Shope, R.E. (1931). SWINE INFLUENZA. *J. Exp. Med.* 54, 349–359.
- Shortridge, K.F. (1999). Poultry and the influenza H5N1 outbreak in Hong Kong, 1997: abridged chronology and virus isolation. *Vaccine* 17 *Suppl* 1, S26–S29.

- Shortridge, K.F., Zhou, N.N., Guan, Y., Gao, P., Ito, T., Kawaoka, Y., Kodihalli, S., Krauss, S., Markwell, D., Murti, K.G., et al. (1998). Characterization of avian H5N1 influenza viruses from poultry in Hong Kong. *Virology* 252, 331–342.
- Sikora, D., Rocheleau, L., Brown, E.G., and Pelchat, M. (2014). Deep sequencing reveals the eight facets of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. *Sci. Rep.* 4, 6181.
- Simonsen, L., Clarke, M.J., Schonberger, L.B., Arden, N.H., Cox, N.J., and Fukuda, K. (1998). Pandemic versus Epidemic Influenza Mortality: A Pattern of Changing Age Distribution. *J. Infect. Dis.* 178, 53–60.
- Skorko, R., Summers, D.F., and Galarza, J.M. (1991). Influenza A virus in vitro transcription: roles of NS1 and NP proteins in regulating RNA synthesis. *Virology* 180, 668–677.
- Smith, D.J., Lapedes, A.S., de Jong, J.C., Bestebroer, T.M., Rimmelzwaan, G.F., Osterhaus, A.D.M.E., and Fouchier, R.A.M. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science* 305, 371–376.
- Smith, W., Andrewes, C.H., and Laidlaw, P.P. (1933). A virus obtained from influenza patients. *The Lancet* 222, 66–68.
- Song, L. (2014). It is Unlikely That Influenza Viruses Will Cause a Pandemic Again Like What Happened in 1918 and 1919. *Front. Public Health* 2.
- Soniat, M., Sampathkumar, P., Collett, G., Gizzi, A.S., Banu, R.N., Bhosle, R.C., Chamala, S., Chowdhury, S., Fiser, A., Glenn, A.S., et al. (2013). Crystal Structure of human Karyopherin β 2 bound to the PY-NLS of *Saccharomyces cerevisiae* Nab2. *J. Struct. Funct. Genomics* 14, 31–35.
- Souza, L.S. de F., Ramos, E.A.G., Carvalho, F.M., Guedes, V.M.C.R., Souza, L.S., Rocha, C.M., Soares, A.B., Velloso, L. de F., Macedo, I.S., Moura, F.E.A., et al. (2003). Viral respiratory infections in young children attending day care in urban Northeast Brazil. *Pediatr. Pulmonol.* 35, 184–191.
- Stokes, J., Chenoweth, A.D., Waltz, A.D., Gladen, R.G., and Shaw, D. (1937). Results of immunization by means of active virus of human influenza 1. *J. Clin. Invest.* 16, 237–243.
- Stray, S.J., and Air, G.M. (2001). Apoptosis by influenza viruses correlates with efficiency of viral mRNA synthesis. *Virus Res.* 77, 3–17.
- Suarez, D.L., and Perdue, M.L. (1998). Multiple alignment comparison of the non-structural genes of influenza A viruses. *Virus Res.* 54, 59–69.
- Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C., Huang, J., et al. (1998). Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* 279, 393–396.
- Sugiyama, K., Obayashi, E., Kawaguchi, A., Suzuki, Y., Tame, J.R.H., Nagata, K., and Park, S.-Y. (2009). Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *EMBO J.* 28, 1803–1811.
- Sugrue, R.J., and Hay, A.J. (1991). Structural characteristics of the M2 protein of influenza A viruses: evidence that it forms a tetrameric channel. *Virology* 180, 617–624.
- Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* 76, 2879–2886.
- Svergun, D., Barberato, C., and Koch, M.H.J. (1995). *CRY SOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* 28, 768–773.
- Svergun, D.I., Volkov, V.V., Kozin, M.B., and Stuhrmann, H.B. (1996). New Developments in Direct Shape Determination from Small-Angle Scattering. 2. Uniqueness. *Acta Crystallogr. A* 52, 419–426.

- Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* *80*, 2946–2953.
- Sylte, M.J., and Suarez, D.L. (2009). Influenza neuraminidase as a vaccine antigen. *Curr. Top. Microbiol. Immunol.* *333*, 227–241.
- Takizawa, N., Kumakura, M., Takeuchi, K., Kobayashi, N., and Nagata, K. (2010). Sorting of influenza A virus RNA genome segments after nuclear export. *Virology* *401*, 248–256.
- Tang, Y., Zaitseva, F., Lamb, R.A., and Pinto, L.H. (2002). The gate of the influenza virus M2 proton channel is formed by a single tryptophan residue. *J. Biol. Chem.* *277*, 39880–39886.
- Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W.H., Daigle, N., Ellenberg, J., et al. (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat. Struct. Mol. Biol.* *14*, 229–233.
- Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W.H., Cusack, S., and Hart, D.J. (2008). Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* *4*, e1000136.
- Tarus, B., Bakowicz, O., Chenavas, S., Duchemin, L., Estrozi, L.F., Bourdieu, C., Lejal, N., Bernard, J., Moudjou, M., Chevalier, C., et al. (2012). Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie* *94*, 776–785.
- Taubenberger, J.K., and Morens, D.M. (2006). 1918 Influenza: the Mother of All Pandemics. *Emerg. Infect. Dis.* *12*, 15–22.
- Tauber, S., Ligertwood, Y., Quigg-Nicol, M., Dutia, B.M., and Elliott, R.M. (2012). Behaviour of influenza A viruses differentially expressing segment 2 gene products in vitro and in vivo. *J. Gen. Virol.* *93*, 840–849.
- Tawar, R.G., Duquerroy, S., Vornrhein, C., Varela, P.F., Damier-Piolle, L., Castagné, N., MacLellan, K., Bedouelle, H., Bricogne, G., Bhella, D., et al. (2009). Crystal structure of a nucleocapsid-like nucleoprotein-RNA complex of respiratory syncytial virus. *Science* *326*, 1279–1283.
- Taylor, R.M. (1949). Studies on Survival of Influenza Virus Between Epidemics and Antigenic Variants of the Virus. *Am. J. Public Health Nations Health* *39*, 171–178.
- Teissier, E., Zandomenighi, G., Loquet, A., Lavillette, D., Lavergne, J.-P., Montserret, R., Cosset, F.-L., Böckmann, A., Meier, B.H., Penin, F., et al. (2011). Mechanism of inhibition of enveloped virus membrane fusion by the antiviral drug arbidol. *PloS One* *6*, e15874.
- Thierry, F., and Danos, O. (1982). Use of specific single stranded DNA probes cloned in M13 to study the RNA synthesis of four temperature-sensitive mutants of HK/68 influenza virus. *Nucleic Acids Res.* *10*, 2925–2938.
- Thomas, F., and Magill, T.P. (1936). Vaccination of Human Subjects with Virus of Human Influenza. *Exp. Biol. Med.* *33*, 604–606.
- Tiley, L.S., Hagen, M., Matthews, J.T., and Krystal, M. (1994). Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J. Virol.* *68*, 5108–5116.
- Tomescu, A.I., Robb, N.C., Hengrung, N., Fodor, E., and Kapanidis, A.N. (2014). Single-molecule FRET reveals a corkscrew RNA structure for the polymerase-bound influenza virus promoter. *Proc. Natl. Acad. Sci.* *111*, E3335–E3342.
- Tong, S., Zhu, X., Li, Y., Shi, M., Zhang, J., Bourgeois, M., Yang, H., Chen, X., Recuenco, S., Gomez, J., et al. (2013). New World Bats Harbor Diverse Influenza A Viruses. *PLoS Pathog.* *9*.

- Torreira, E., Schoehn, G., Fernández, Y., Jorba, N., Ruigrok, R.W.H., Cusack, S., Ortín, J., and Llorca, O. (2007). Three-dimensional model for the isolated recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res.* **35**, 3774–3783.
- Toyoda, T., Adyshev, D.M., Kobayashi, M., Iwata, A., and Ishihama, A. (1996). Molecular assembly of the influenza virus RNA polymerase: determination of the subunit-subunit contact sites. *J. Gen. Virol.* **77** (Pt 9), 2149–2157.
- Treanor, J. (2004). Influenza Vaccine — Outmaneuvering Antigenic Shift and Drift. *N. Engl. J. Med.* **350**, 218–220.
- Trifonov, V., Khiabani, H., Greenbaum, B., and Rabadan, R. (2009). The origin of the recent swine influenza A(H1N1) virus infecting humans. *Euro Surveill. Bull. Eur. Sur Mal. Transm. Eur. Commun. Dis. Bull.* **14**.
- Trowitzsch, S., Bieniossek, C., Nie, Y., Garzoni, F., and Berger, I. (2010). New baculovirus expression tools for recombinant protein complex production. *J. Struct. Biol.* **172**, 45–54.
- Turan, K., Mibayashi, M., Sugiyama, K., Saito, S., Numajiri, A., and Nagata, K. (2004). Nuclear MxA proteins form a complex with influenza virus NP and inhibit the transcription of the engineered influenza virus genome. *Nucleic Acids Res.* **32**, 643–652.
- Ulmanen, I., Broni, B.A., and Krug, R.M. (1981). Role of two of the influenza virus core P proteins in recognizing cap 1 structures (m7GpppNm) on RNAs and in initiating viral RNA transcription. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 7355–7359.
- Ungchusak, K., Auewarakul, P., Dowell, S.F., Kitphati, R., Auwanit, W., Puthavathana, P., Uiprasertkul, M., Boonnak, K., Pittayawonganon, C., Cox, N.J., et al. (2005). Probable Person-to-Person Transmission of Avian Influenza A (H5N1). *N. Engl. J. Med.* **352**, 333–340.
- Vagin, A., and Teplyakov, A. (1997). *MOLREP*: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* **30**, 1022–1025.
- Vahlenkamp, T.W., and Harder, T.C. (2006). Influenza virus infections in mammals. *Berl. Münch. Tierärztl. Wochenschr.* **119**, 123–131.
- Valleron, A.-J., Cori, A., Valtat, S., Meurisse, S., Carrat, F., and Boëlle, P.-Y. (2010). Transmissibility and geographic spread of the 1889 influenza pandemic. *Proc. Natl. Acad. Sci.* **107**, 8778–8781.
- Varga, Z.T., Grant, A., Manicassamy, B., and Palese, P. (2012). Influenza virus protein PB1-F2 inhibits the induction of type I interferon by binding to MAVS and decreasing mitochondrial membrane potential. *J. Virol.* **86**, 8359–8366.
- Vasin, A.V., Temkina, O.A., Egorov, V.V., Klotchenko, S.A., Plotnikova, M.A., and Kiselev, O.I. (2014). Molecular mechanisms enhancing the proteome of influenza A viruses: An overview of recently discovered proteins. *Virus Res.* **185**, 53–63.
- Viboud, C., Alonso, W.J., and Simonsen, L. (2006). Influenza in Tropical Regions. *PLoS Med* **3**, e89.
- Vijayachandran, L.S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., et al. (2011). Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. *J. Struct. Biol.* **175**, 198–208.
- Viswanathan, K., Chandrasekaran, A., Srinivasan, A., Raman, R., Sasisekharan, V., and Sasisekharan, R. (2010). Glycans as receptors for influenza pathogenesis. *Glycoconj. J.* **27**, 561–570.
- Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.

- Wan, Z., Ye, J., Xu, L., Shao, H., Jin, W., Qian, K., Wan, H., and Qin, A. (2014). Antigenic mapping of the hemagglutinin of an H9N2 avian influenza virus reveals novel critical amino acid positions in antigenic sites. *J. Virol.* *88*, 3898–3901.
- Wang, C., Lamb, R.A., and Pinto, L.H. (1995). Activation of the M2 ion channel of influenza virus: a role for the transmembrane domain histidine residue. *Biophys. J.* *69*, 1363–1371.
- Ward, A.C., Castelli, L.A., Lucantoni, A.C., White, J.F., Azad, A.A., and Macreadie, I.G. (1995). Expression and analysis of the NS2 protein of influenza A virus. *Arch. Virol.* *140*, 2067–2073.
- Watanabe, K., Handa, H., Mizumoto, K., and Nagata, K. (1996). Mechanism for inhibition of influenza virus RNA polymerase activity by matrix protein. *J. Virol.* *70*, 241–247.
- Webster, R.G., Bean, W.J., Gorman, O.T., Chambers, T.M., and Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol. Rev.* *56*, 152–179.
- Weissenhorn, W., Hinz, A., and Gaudin, Y. (2007). Virus membrane fusion. *FEBS Lett.* *581*, 2150–2155.
- Wernimont, A., and Edwards, A. (2009). In Situ Proteolysis to Generate Crystals for Structure Determination: An Update. *PLoS ONE* *4*, e5094.
- Whittaker, G., Kemler, I., and Helenius, A. (1995). Hyperphosphorylation of mutant influenza virus matrix protein, M1, causes its retention in the nucleus. *J. Virol.* *69*, 439–445.
- WHO (1980). A revision of the system of nomenclature for influenza viruses: a WHO Memorandum. *Bull. World Health Organ.* *58*, 585–591.
- WHO (2015). Influenza (Seasonal). Available at <http://www.who.int/mediacentre/factsheets/fs211/en/>.
- WHO Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003 - 2015. Available at http://www.who.int/influenza/human_animal_interface/EN_GIP_20150126CumulativeNumberH5N1cases.pdf?ua=1.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP 4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 235–242.
- Wise, H.M., Foeglein, A., Sun, J., Dalton, R.M., Patel, S., Howard, W., Anderson, E.C., Barclay, W.S., and Digard, P. (2009). A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. *J. Virol.* *83*, 8021–8031.
- Wise, H.M., Hutchinson, E.C., Jagger, B.W., Stuart, A.D., Kang, Z.H., Robb, N., Schwartzman, L.M., Kash, J.C., Fodor, E., Firth, A.E., et al. (2012). Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. *PLoS Pathog.* *8*, e1002998.
- Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2013). Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* *280*, 5705–5736.
- Wolstenholme, A.J., Barrett, T., Nichol, S.T., and Mahy, B.W. (1980). Influenza virus-specific RNA and protein syntheses in cells infected with temperature-sensitive mutants defective in the genome segment encoding nonstructural proteins. *J. Virol.* *35*, 1–7.
- Wong, I., and Lohman, T.M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc. Natl. Acad. Sci.* *90*, 5428–5432.
- World Health Organisation (2015). Influenza.
- Worobey, M., Han, G.-Z., and Rambaut, A. (2014). Genesis and pathogenesis of the 1918 pandemic H1N1 influenza A virus. *Proc. Natl. Acad. Sci. U. S. A.* *111*, 8107–8112.
- Wright et al (2013). Orthomyxoviruses. In *In Fields Virology 6th Edition*, D.M. Knipe, and P. Howley, ed. pp. 1186–1243.

- Wright, P.F. et al (2013). General virology-Principle of viral structures. In *In Fields Virology 6th Edition*, D.M. Knipe, and P. Howley, ed. pp. 1186–1243.
- Wyatt, P.J. (1991). Combined differential light scattering with various liquid chromatography separation techniques. *Biochem. Soc. Trans.* *19*, 485.
- Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. *Anal. Chim. Acta* *272*, 1–40.
- Xu, D., Farmer, A., and Chook, Y.M. (2010). Recognition of nuclear targeting signals by Karyopherin- β proteins. *Curr. Opin. Struct. Biol.* *20*, 782–790.
- Yasuda, J., Nakada, S., Kato, A., Toyoda, T., and Ishihama, A. (1993). Molecular assembly of influenza virus: association of the NS2 protein with virion matrix. *Virology* *196*, 249–255.
- Ye, Q., Krug, R.M., and Tao, Y.J. (2006). The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* *444*, 1078–1082.
- Ye, Q., Guu, T.S.Y., Mata, D.A., Kuo, R.-L., Smith, B., Krug, R.M., and Tao, Y.J. (2012). Biochemical and structural evidence in support of a coherent model for the formation of the double-helical influenza A virus ribonucleoprotein. *mBio* *4*, e00467–00412.
- Yen, H.-L., Hoffmann, E., Taylor, G., Scholtissek, C., Monto, A.S., Webster, R.G., and Govorkova, E.A. (2006). Importance of neuraminidase active-site residues to the neuraminidase inhibitor resistance of influenza viruses. *J. Virol.* *80*, 8787–8795.
- Yoneyama, M., Kikuchi, M., Natsukawa, T., Shinobu, N., Imaizumi, T., Miyagishi, M., Taira, K., Akira, S., and Fujita, T. (2004). The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat. Immunol.* *5*, 730–737.
- Yoon, S.-W., Webby, R.J., and Webster, R.G. (2014). Evolution and ecology of influenza A viruses. *Curr. Top. Microbiol. Immunol.* *385*, 359–375.
- Yuan, P., Bartlam, M., Lou, Z., Chen, S., Zhou, J., He, X., Lv, Z., Ge, R., Li, X., Deng, T., et al. (2009). Crystal structure of an avian influenza polymerase PAN reveals an endonuclease active site. *Nature* *458*, 909–913.
- Zell, R., Krumbholz, A., Eitner, A., Krieg, R., Halbhuber, K.-J., and Wutzler, P. (2007). Prevalence of PB1-F2 of influenza A viruses. *J. Gen. Virol.* *88*, 536–546.
- Zhang, S., Wang, J., Wang, Q., and Toyoda, T. (2010). Internal initiation of influenza virus replication of viral RNA and complementary RNA in vitro. *J. Biol. Chem.* *285*, 41194–41201.
- Zheng, W., Li, J., Wang, S., Cao, S., Jiang, J., Chen, C., Ding, C., Qin, C., Ye, X., Gao, G.F., et al. (2015). Phosphorylation controls the nuclear-cytoplasmic shuttling of influenza a virus nucleoprotein. *J. Virol.* *89*, 5822–5834.
- Zhou, Z.H. (2008). Towards atomic resolution structural determination by single-particle cryo-electron microscopy. *Curr. Opin. Struct. Biol.* *18*, 218–228.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* *419*, 182–185.
- Zürcher, T., de la Luna, S., Sanz-Ezquerro, J.J., Nieto, A., and Ortín, J. (1996). Mutational analysis of the influenza virus A/Victoria/3/75 PA protein: studies of interaction with PB1 protein and identification of a dominant negative mutant. *J. Gen. Virol.* *77* (Pt 8), 1745–1749.

Publications

- Crépin T., Swale C., Monod A., Garzoni F., Chaillet M. and Berger I. (2015). **Polyproteins in structural biology**. Current Opinion in Structural Biology, 32:139–146
- Monod A., Swale C., Tarus B., Tissot A., Delmas B., Ruigrok R., Crépin C., and Slama-Schwok A. (2015). **Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza**. Expert Opin. Drug Discov. 10(4):345-371
- Crépin T., Dias A., Palencia A., Swale C., Cusack C., And Ruigrok. (2010). **Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit**. Journal of Virology, p. 9096–9104

Publication n°1:

Thibaut Crépin, Christopher Swale, Alexandre Monod, Frederic Garzoni, Maxime Chaillet and Imre Berger

Polyproteins in structural biology

Current Opinion in Structural Biology 2015, 32:139–146



ELSEVIER

Polyproteins in structural biology

Thibaut Crépin¹, Christopher Swale¹, Alexandre Monod¹,
Frederic Garzoni^{1,2}, Maxime Chaillet^{1,2} and Imre Berger^{1,2,3}

Polyproteins are chains of covalently conjoined smaller proteins that occur in nature as versatile means to organize the proteome of viruses including HIV. During maturation, viral polyproteins are typically cleaved into the constituent proteins with different biological functions by highly specific proteases, and structural analyses at defined stages of this maturation process can provide clues for antiviral intervention strategies. Recombinant polyproteins that use similar mechanisms are emerging as powerful tools for producing hitherto inaccessible protein targets such as the influenza polymerase, for high-resolution structure determination by X-ray crystallography. Conversely, covalent linking of individual protein subunits into single polypeptide chains are exploited to overcome sample preparation bottlenecks. Moreover, synthetic polyproteins provide a promising tool to dissect dynamic folding of polypeptide chains into three-dimensional architectures in single-molecule structure analysis by atomic force microscopy (AFM). The recent use of natural and synthetic polyproteins in structural biology and major achievements are highlighted in this contribution.

Addresses

¹ Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, UMI 3265, 71 Avenue des Martyrs, 38042 Grenoble Cedex 9, France

² The European Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, BP181, 38042 Grenoble Cedex 9, France

³ The School of Biochemistry, University of Bristol, Bristol BS8 1TD, United Kingdom

Corresponding authors: Crépin, Thibaut (crepin@embl.fr) and Berger, Imre (iberger@embl.fr)

Current Opinion in Structural Biology 2015, 32:139–146

This review comes from a themed issue on **New constructs and expressions of proteins**

Edited by Imre Berger and Roslyn M Bill

<http://dx.doi.org/10.1016/j.sbi.2015.04.007>

0959-440X/Crown Copyright © 2015 Published by Elsevier Ltd. All rights reserved.

Introduction

Polyproteins composed of covalently linked individual proteins with different biological functions are prevalent in nature. For instance SARS coronavirus, the agent that causes severe acute respiratory syndrome, realizes its entire proteome from two large polyproteins, each encoded by a

long single open reading frame (ORF) [1]. The expressed SARS polyproteins are then processed into the individual functional protein subunits by the action of highly specific proteases also encoded by the ORFs [1,2]. A further example is human immune deficiency virus (HIV) which causes acquired immune deficiency syndrome (AIDS). The RNA genome of this retrovirus is organized in three major genes *gag*, *pol* and *env*, which encode for polyproteins and undergo proteolytic processing at defined stages during maturation [3]. Recombinant polyprotein approaches mimicking viral polyproteins have recently emerged as a powerful means to express high-value protein complexes for structure determination [4,5]. Recently, the long elusive influenza polymerase has been produced successfully from a self-processing synthetic polyprotein, enabling high-resolution structure determination [6,7]. Polyprotein fusions which are not processed by protease, but remain covalently conjoined by engineered linkers have been instrumental to obtain important insight into numerous essential physiological processes including multidrug efflux, co-translational protein targeting or enzymatic processing of chromatin, among others [8,9,10,11,12]. Moreover, single-chain engineering approaches linking protein domains into novel, artificial polyproteins have resulted in new classes of high-affinity binder molecules as potential protein therapeutics [13] and accelerated elucidation of mechanisms governing protein folding by single-molecule techniques [14]. Thus, polyprotein technologies have recently gained prominence as particularly useful tools for unlocking previously often inaccessible protein samples to detailed structural and mechanistic studies, as illustrated in the following. In this contribution, the use of polyproteins in structural biology is discussed, by highlighting recently determined structures of naturally occurring polyproteins on one hand and by reviewing recent structural studies where recombinant polyprotein constructions were utilized on the other.

Natural polyproteins

Polyproteins are used in nature by many viruses to structure their proteome. As a consequence, viral polyproteins are an intense focus of research efforts for numerous reasons [15–17]. For instance, inhibition by small molecules of proteases that process polyproteins during viral maturation can provide a powerful handle to combat viral disease [2,18–20]. In HIV, the physical infrastructure of the virus is provided by the *gag* gene which gives rise to the precursor Gag (group specific antigen) polyprotein [21]. Viral protease processes Gag during maturation into several proteins and spacer peptides dictating immature

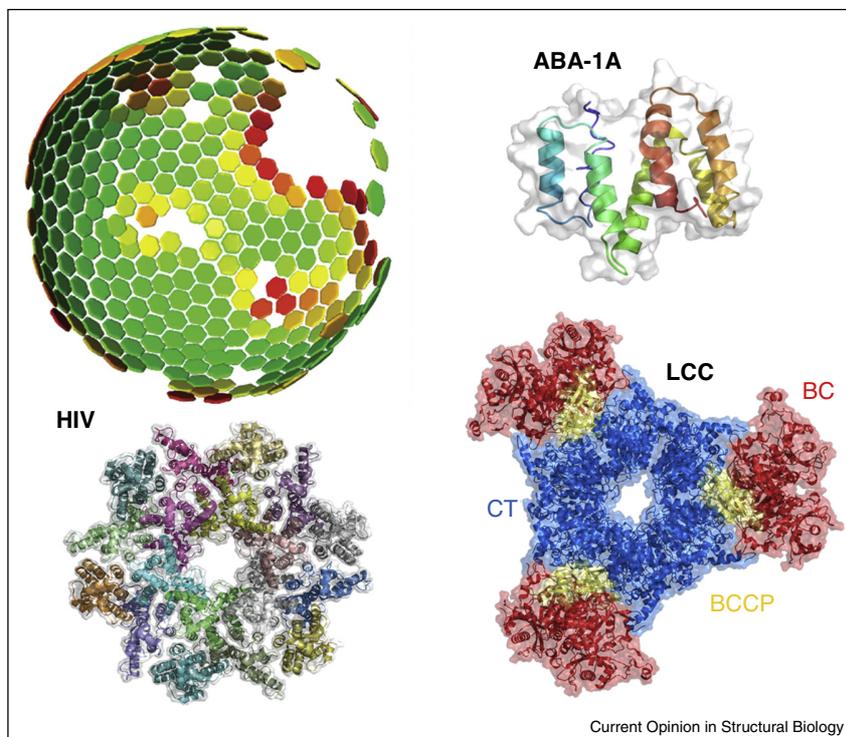
and mature viral capsid structure [22]. Maturation is a two-stage process. Precursor Gag polyprotein first forms a hexameric lattice at the plasma membrane of an infected cell. This induces budding and release of immature viral particles. Proteolytic processing of Gag then rearranges the viral structure to the mature form [23]. Inhibition of Gag-processing protease enabled preparation of immature retroviral capsids suitable for structure analysis by electron cryo-tomography and subtomogram averaging methods (Figure 1) [24–26,27**]. Comparison of the structures of HIV Gag protein in reconstituted tubular arrays on one hand and in intact virus particles on the other, provided unprecedented insights into the conformational plasticity of this precursor polyprotein [26,27**]. The studies also revealed that retroviral capsid proteins can adopt different quaternary arrangements during virus assembly, notwithstanding conserved tertiary structures [27**].

A different, less frequently encountered type of natural, non-viral polyproteins are so-called tandemly repetitive polyproteins (TRPs). TRPs are also produced as large precursor proteins and then processed by proteases into several copies of proteins with similar function. TRPs are

made of consecutively arranged repeats of amino-acid stretches. Examples of known TRPs include polyprotein lipid binding proteins of nematodes [28] and filaggrins, which are keratinocyte produced TRPs crucial to health and appearance of skin [29,30]. The solution structure of a mature, post-translationally processed repeat unit of a TRP, ABA-1A from the nematode polyprotein allergen of *Ascaris*, was determined, representing the first structure of this class of proteins [31]. ABA-1A adopts a novel fold comprising two juxtaposed four-helical bundles that share a long central alpha-helix (Figure 1). Nematode polyprotein allergens have no known counterpart in humans. The *Ascaris* ABA-1A structure therefore may serve as a starting point for the development of new drugs and therapeutic intervention strategies against disease states caused by these intestinal parasites.

An inverse ‘polyprotein’ concept of covalently linking functional protein units into long modular polypeptide chains characterizes mega-enzymes that functionally arrange multiple domains into ordered assembly lines for the production of a wide variety of bioactive molecules. Modular polyketide synthases (PKS) and their metazoan homologs, fatty acid synthases (FAS) belong to this class

Figure 1



Natural polyprotein structures. Polyproteins, prevalent in biology, are illustrated here by the structures of the human immune deficiency virus (HIV) immature capsid determined by electron cryo-tomography revealing molecular details of hexameric Gag [24,26,27**], the structure of a repeated unit of the ABA-1 nematode polyprotein allergen derived from nuclear magnetic resonance (NMR) spectroscopy [31] and the crystal structure of a single-chain, multi-domain long-chain acyl-CoA carboxylase, LCC [34]. CT stands for carboxyltransferase, BCCP for biotin carboxyl carrier protein, and BC for biotin carboxylase components.

of catalysts [32,33]. Recently, the structure and function of a modular multi-domain long-chain acyl-CoA carboxylase from *Mycobacterium avium* was elucidated [34]. The crystal structure revealed extensive swapping of functional domains in the holo-enzyme which is a homo-hexamer (Figure 1). Thus, four intertwined protomers are involved in completing one catalytic reaction cycle [34].

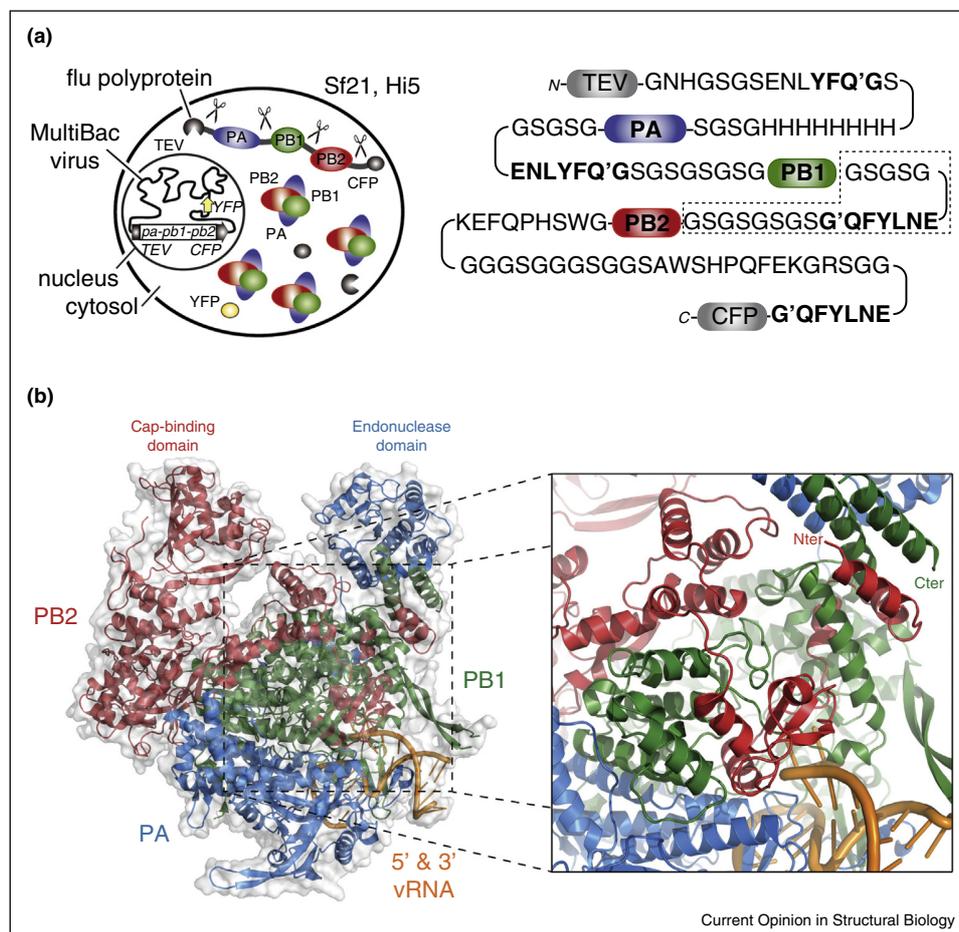
Synthetic polyproteins

Influenza polymerase produced from a recombinant polyprotein

The highly successful strategy of viruses to utilize polyproteins to their advantage has found its equivalent in recombinant technology. A synthetic polyprotein,

expressed recombinantly in baculovirus-infected insect cells, has enabled the structure determination of influenza polymerase. Despite its common appearance, a detailed understanding of the molecular mechanisms of the virus that causes influenza has remained elusive. More than 40 years ago, the influenza polymerase was discovered, a key protein complex that replicates the genetic material of the virus [35,36]. Atomic resolution information on the structure and function of this protein machine is essential, as it may open up important avenues for drug discovery. However, the influenza polymerase remained inaccessible for decades — to produce this valuable protein complex for detailed analysis, proved to be a seemingly insurmountable technical challenge. This has now changed

Figure 2



Influenza polymerase. **(a)** A self-processing recombinant polyprotein was used to express influenza polymerase complex in high quality and quantity in insect cells by using the MultiBac system [37] as illustrated (left). The polymerase was expressed from a single open reading frame encoding for tobacco etch virus N1a protease (TEV), the polymerase subunits PA, PB1 and PB2 and a fluorescent protein (CFP). A second, yellow fluorescent protein (YFP) was inserted elsewhere into the MultiBac baculoviral genome as an expression performance marker. The resulting polyprotein is processed into the individual subunits by highly specific proteolysis mediated by TEV. The polyprotein construct is shown schematically (right) illustrating tag placement and design of spacers in between the subunits (adapted from [6**]). The C-terminal part of PB1 and the N-terminal part of PB2 co-fold into a helical bundle necessitating careful linker design (boxed). **(b)** Crystal structure of influenza polymerase bound to cognate viral RNA (vRNA) is shown in a ribbon representation. PA is colored in blue, PB1 in green and PB2 in red. RNA substrate is colored in orange. Cap-binding domain and endonuclease domain are indicated. The structure motif formed by PB1 C-terminal domain and PB2 N-terminal domain is shown in a magnification (right). PB2 N-terminus and PB1 C-terminus are marked.

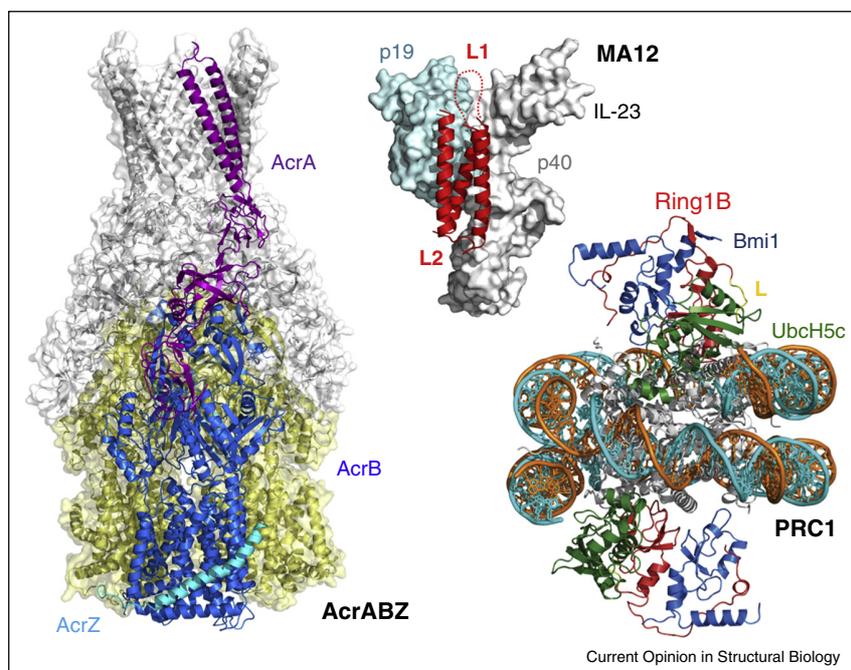
dramatically with structures of influenza polymerase complex determined by X-ray crystallography [6**,7]. These break-through studies provide unprecedented insight into the inner workings of this viral protein machine. This revolution in understanding influenza was brought about by applying a polyprotein strategy to produce influenza polymerase recombinantly, in the quality and quantity required for high-resolution structural and functional analysis (Figure 2). The strategy applied recapitulates the mechanism adopted by SARS coronavirus. A single ORF was constructed encoding for a highly specific protease, NIa, from tobacco etch virus (TEV), fused in frame with consecutively arranged PA, PB1 and PB2 subunits of the polymerase. At the far end of the polyprotein, a reporter protein was inserted to track protein production ‘in real-time’ during expression, by recording fluorescence [5*]. All protein units within the polyprotein were spaced apart by customized linkers containing the specific site for cleavage by the TEV protease, purification tag sequences and spacer residues (Figure 2). The polyproteins encoding for influenza polymerases were successfully expressed with the Multi-Bac baculovirus insect cell expression system developed for producing multiprotein complexes [37].

Sample preparation bottlenecks resolved by synthetic polyproteins that remain uncleaved

The polyprotein encoding for influenza polymerase was proteolytically cleaved into the constituent protein subunits to yield the sample that crystallized. Conversely, engineering of individual proteins into covalently linked polypeptide chains can also accelerate structure determination considerably. These ‘polyproteins’ remain conjoined as single-chains during sample preparation and structure determination. Particularly prominent examples for such single-chain engineering include the insertion of T4 lysozyme into the primary sequence of G-protein coupled receptors (GPCRs) to facilitate crystallogenesis [38]. In a variation of this approach, the catalytic domain of *Pyrococcus abyssi* glycogen synthase was recently used to stabilize an intracellular loop of the human OX₂ orexin receptor to determine its crystal structure bound to an insomnia drug [39].

Elaborate single-chain engineering into polyproteins was applied to determine the architecture of the bacterial multidrug efflux pump AcrABZ-TolC [8**]. In this study, the efflux pump was assembled by preparing two single-chain polypeptide fusions, AcrB-AcrA-AcrB and

Figure 3



Linking individual proteins into polypeptide chains. Proteins engineered into single polypeptide chains were used to obtain suitable sample for structure determination of AcrABZ complex [8**], an alphabody (MA12) to neutralize human interleukin IL-23 (p19 and p40) [13] and the first structure of a histone-modifying enzyme, the Polycomb Repressive Complex (PRC) 1 ubiquitylation module, bound to a nucleosome [12**]. Linker amino acid segments are marked (L, L1, L2).

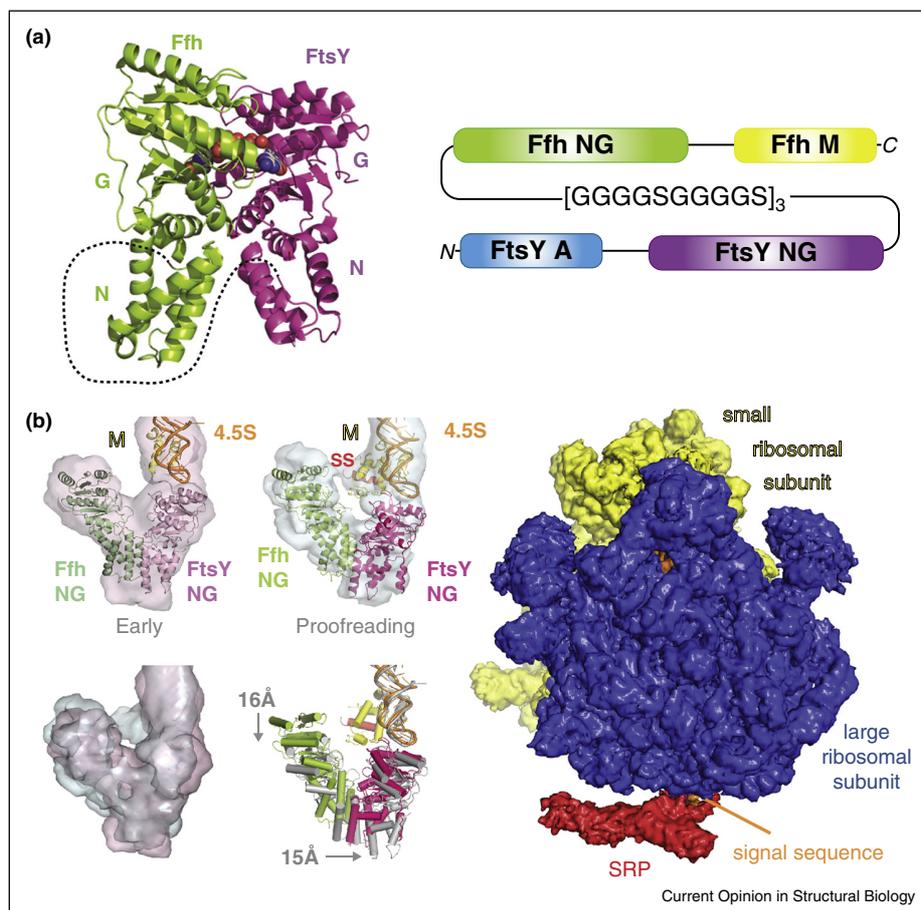
AcrA-AcrZ, respectively, each conjoined by extended glycine/serine rich linkers. TolC was co-expressed with the AcrA-AcrZ fusion. This strategy resulted in a reconstituted complex with the correct stoichiometry enabling structure determination by hybrid methods. Fitting of crystal coordinates into EM densities allowed to model AcrBZ/AcrA interactions (Figure 3) as well as the holo-complex containing TolC.

Single-chain engineering likewise enabled the development of alphabodies, a novel scaffold representing a promising alternative to antibodies for various biomedical and biotechnological applications [13]. Alphabodies comprise *in silico* designed short individual alpha-helical protein segments that are then conjoined by covalent

linkers into single-chain antiparallel coiled-coils that are highly stable and suitable for affinity maturation. The crystal structure of a complex with human interleukin (IL-23) revealed the structural basis of IL-23 antagonism by the alphabody, MA12 (Figure 3).

Covalent linking of subunits of a protein complex was required to obtain the first crystal structure of a histone-modifying enzyme complex bound to a nucleosome core particle [12**]. Polycomb repressor complex (PRC) 1, an essential regulator of cell fate, comprises an activity to ubiquitylate nucleosomal histone H2A at residue K119. PRC1 uses its E3 ubiquitin ligase subunits, Ring1B and Bmi1, together with an E2 ubiquitin-conjugating enzyme, UbcH5c for this purpose. The E2–E3 complex was

Figure 4



SRP pathway revealed by single-chain protein engineering. (a) *E. coli* signal recognition particle (SRP) protein subunit Ffh (green) and the SRP receptor FtsY (purple) were covalently linked into a single polypeptide chain (left). NG domains are marked. The dashed line indicates the extended polypeptide linker. Bound nucleotide is shown (spheres). The single polypeptide construct is illustrated on the right, detailing the arrangement of FtsY A (blue) and NG (purple) domains, the 30 amino acid glycine/serine rich linker connecting FtsY with Ffh, and the Ffh NG (green) and M (yellow) domains. N and C-termini are marked. (b) Structures of SRP/FtsY bound to a translating ribosome were elucidated by electron cryo-microscopy (color coding as in panel a). 4.5S RNA is colored in orange. EM density and fitted models of the 'early' and 'proofreading' stages are shown (left, top). Substantial rearrangements are observed as illustrated by the overlay of the EM densities and the model coordinates (left, bottom) (adapted from [10*]). The cryo-EM structure of the ribosome-SRP-FtsY co-translational targeting complex in the closed state is shown on the right (adapted from [11*]).

produced using an engineered single-chain fusion of Ring1B to UbcH5c to overcome the low affinity and salt-sensitive E2–E3 interaction. The resulting single-chain construct was co-expressed with Bmi1 to produce the trimeric complex which was purified and reconstituted with nucleosome core particles, and crystallized [12^{••}]. The structure showed two copies of PRC1 E2–E3 complex bound to one nucleosome, revealing intricate interactions (Figure 3).

SRP pathway revealed by single-chain protein engineering

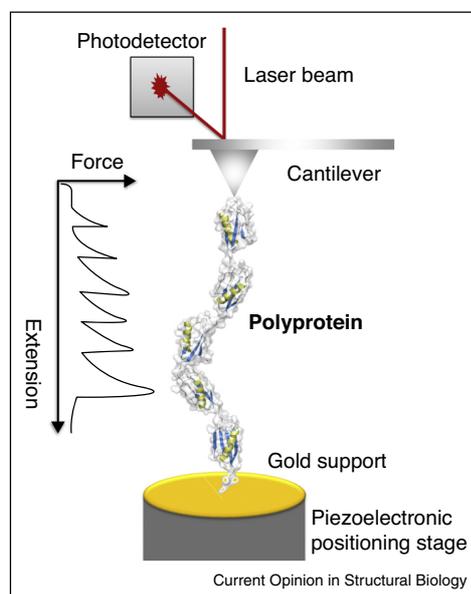
Around one third of the proteins in living cells are delivered to the plasma membrane. This is carried out by a universally conserved, complex mechanism involving ribosomes that are translating mRNA into membrane-bound nascent polypeptide chains, the signal recognition particle, SRP and the SRP receptor, FtsY [9,40,41]. Snapshots of this elaborate process were obtained by single particle electron cryo-microscopy and biochemical analysis [9,10[•],11[•],42–44]. These studies revealed SRP binding to ribosome nascent chain complexes [42], followed by the ‘early’ [44], ‘proofreading’ [10[•]] and ‘closed’ [11[•]] states upon FtsY binding and GTP hydrolysis. Successful structure determination critically relied on stabilizing SRP binding to FtsY. This was achieved by covalently linking the SRP subunit Ffh with FtsY into a single polypeptide chain with virtually wild-type activity (Figure 4).

Polyprotein single-molecule structural biology

Transient unfolding and refolding of proteins can be an essential feature of protein structure space in living organisms, for example in the translocation of proteins into and across cellular membranes or the muscle stroke. The use of atomic force microscopy (AFM) has emerged as a powerful technique to probe protein structure, enabling analysis of the mechanical stability and folding pathway of protein specimens at the single-molecule level. By means of a stretching force, applied through a microscopic cantilever to a biological target fixed to a support and recorded by the deflection of a laser beam, the analyzed protein is unfolded to an extended state (Figure 5).

Historically, polyproteins were used in single-molecule AFM to measure unique mechanical fingerprint profiles of a protein response to the applied mechanical force [14,45[•]]. The muscle protein titin is a prominent example [46–48]. Titin consists of several hundred repeated protein domains including fibronectin and immunoglobulin-like folds. The use of polyproteins in single-molecule AFM results in considerable improvement of the statistical evaluation of singular domains within the polyprotein chain [14]. This can be exploited in the analysis of homomeric polyproteins that are constructed genetically or by chemical fusion reaction from identical copies of the same protein species of choice. Moreover, in addition to

Figure 5



Polyprotein single-molecule structural biology. The setup of atomic force microscopy of polyproteins is shown in a schematic representation (adapted from [14]). The polyprotein is tethered to the gold support resting on the piezoelectric positioning stage (bottom) on one end, and the tip of a cantilever made of silicon nitride on the other. A laser beam is focused on the back of the cantilever (top). The cantilever is displaced by the force that acts on the polyprotein chain resulting in change of the deflection of the laser beam, recorded by a photodetector. Increasing force causes each domain of the polyprotein to unfold, resulting in characteristic spikes in a force/extension diagram (inset).

providing a clear fingerprint, polyproteins also have the advantage that a larger number of events can be recorded per experiment as compared to only one event if monomeric proteins are used.

For these reasons polyproteins emerge as work-horses of single-molecule structural biology by AFM. Numerous polyproteins have been analyzed for their mechanical properties by using this technique, including poly-I27, derived from the I-band region of Titin, oligo-calmodulin, poly-ubiquitin, polyproteins made of the virulence factor GB1 of *Peptostreptococcus magnus* [46–54] and others, providing unique insights into biological folding/unfolding mechanisms. The availability of a large and growing number of well characterized homomeric specimens furthermore enables now the construction of chimeric polyproteins as a tool to study mechanically uncharacterized proteins, by using the unique fingerprints of the known protein unit as a reference [14,45[•]].

Conclusions

Natural and synthetic polyproteins are at the core of contemporary structural biology. Analysis of viral polyproteins

and the architectural consequences of their processing during maturation not only furthers our understanding of viral mechanisms but provides important clues for drug design to combat viral diseases. Synthetic polyproteins are emerging as invaluable tools to accelerate research by unlocking hitherto inaccessible proteins for high resolution structure determination. Artificial polyproteins obtained by single-chain protein engineering approaches are instrumental to overcome sample production bottlenecks and provide novel means to illuminate biological mechanisms, including folding and unfolding properties at the single-molecule level. We anticipate a major increase in the use of polyproteins in structural biology as valuable tools to tackle large and complex biological systems in the future.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

We thank all members of our laboratories for their assistance. We thank Timothy J. Richmond, Darren Hart, Christiane Schaffitzel, Alexander Pflug and Stefan Reich for their contributions and especially Rob Ruijgrok for discussions and helpful suggestions. We are grateful to the Alliance Grenobloise pour la Biologie Structurale et Cellulaire Intégrées (GRAL) (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB) for generous support. Ben Luisi (Cambridge, UK) and John Briggs kindly provided the AcrABZ model coordinates and the HIV immature capsid image, respectively. TC, CS and AM are supported by the Centre National pour la Recherche Scientifique (CNRS), the Agence Nationale de Recherche (ANR) (contract nr. ANR-14-CE09-0017) and the European Commission (EC) Framework Program 7 (FP7) project FluPHARM (contract nr. 259751). CS acknowledges support from the Hofmann-La Roche pRED external collaboration program. IB, MC and FG are supported by the Agence Nationale de Recherche (ANR) through the Investissement d'Avenir program, and the EC through FP7 projects BioSTRUCT-X (contract nr. 283570) and ComplexINC (contract nr. 279039).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Al-Tawfiq JA, Zumla A, Memish ZA: **Coronaviruses: severe acute respiratory syndrome coronavirus and Middle East respiratory syndrome coronavirus in travelers.** *Emerg Infect Dis* 2014, **20**:1562-1564.
 2. Hilgenfeld R: **From SARS to MERS: crystallographic studies on coronavirus proteases enable antiviral drug design.** *FEBS J* 2014, **281**:4085-4096.
 3. Lucas S, Nelson AM: **HIV and the spectrum of human disease.** *J Pathol* 2015, **235**:229-241.
 4. Vijayachandran LS, Viola C, Garzoni F, Trowitzsch S, Bieniossek C, Chaillet M, Schaffitzel C, Busso D, Romier C, Poterszman A *et al.*: **Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells.** *J Struct Biol* 2011, **175**:198-208.
 5. Nie Y, Bellon-Echeverria I, Trowitzsch S, Bieniossek C, Berger I: **Multiprotein complex production in insect cells by using polyproteins.** *Methods Mol Biol* 2014, **1091**:131-141.
- Detailed description of the methodology and protocols for expressing protein complexes from synthetic polyproteins in the MultiBac baculovirus insect cell expression system.
6. Reich S, Guilligay D, Pflug A, Malet H, Berger I, Crépin T, Hart D, Lunardi T, Nanao M, Ruijgrok RW, Cusack S: **Structural insight into cap-snatching and RNA synthesis by influenza polymerase.** *Nature* 2014, **516**:361-366.
- First crystal structure of the influenza polymerase bound to cognate viral RNA substrate, providing unprecedented insights into cap-snatching and RNA synthesis. Influenza polymerase, inaccessible to structural analysis for decades since its discovery, was produced here for the first time successfully from a synthetic polyprotein construct.
7. Pflug A, Guilligay D, Reich S, Cusack S: **Structure of influenza A polymerase bound to the viral RNA promoter.** *Nature* 2014, **516**:355-360.
 8. Du D, Wang Z, James NR, Voss JE, Klimont E, Ohene-Agyei T, Venter H, Chiu W, Luisi BF: **Structure of the AcrAB-TolC multidrug efflux pump.** *Nature* 2014, **509**:512-515.
- Structure determination of the bacterial AcrABZ-TolC multidrug efflux pump by hybrid methods combining X-ray crystallography and electron cryo-microscopy. Elaborate single-chain protein engineering was a prerequisite for successful sample preparation of the complete efflux pump.
9. Knoops K, Schoehn G, Schaffitzel C: **Cryo-electron microscopy of ribosomal complexes in cotranslational folding, targeting, and translocation.** *Wiley Interdiscip Rev RNA* 2012, **3**:429-441.
 10. von Loeffelholz O, Knoops K, Ariosa A, Zhang X, Karupphasamy M, Huard K, Schoehn G, Berger I, Shan SO, Schaffitzel C: **Structural basis of signal sequence surveillance and selection by the SRP-FtsY complex.** *Nat Struct Mol Biol* 2013, **20**:604-610.
- Structure of the 'proof-reading' step in co-translational targeting, revealed by electron cryo-microscopy and biochemical analysis of a reconstituted assembly comprising ribosome nascent chain complex, SRP and SRP receptor, FtsY. The SRP protein subunit Ffh and FtsY were provided as an engineered protein fusion conjoined by an extended glycine/serine rich linker. The resulting protein fusion displayed wild-type activity.
11. von Loeffelholz O, Jiang Q, Ariosa A, Karupphasamy M, Huard K, Berger I, Shan SO, Schaffitzel C: **Ribosome-SRP-FtsY cotranslational targeting complex in the closed state.** *Proc Natl Acad Sci U S A* 2015, **112**:3943-3948.
- High-resolution electron cryo-microscopy study revealing the structure and mechanism of the 'closed' step in bacterial co-translational targeting. The emerging signal sequence bound to the Ffh M-domain was observed in near-atomic detail.
12. McGinty RK, Henrici RC, Tan S: **Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome.** *Nature* 2014, **514**:591-596.
- 3.0 Å crystal structure determination of the first histone-modifying enzyme complex, the PRC1 deubiquitylation module, bound to a nucleosome. Single-chain engineering was required to stabilize the enzyme complex for crystallization.
13. Desmet J, Verstraete K, Bloch Y, Lorent E, Wen Y, Devreese B, Vandenbroucke K, Loverix S, Hettmann T, Deroo S *et al.*: **Structural basis of IL-23 antagonism by an Alphabody protein scaffold.** *Nat Commun* 2014, **5**:5237.
 14. Hoffmann T, Dougan L: **Single molecule force spectroscopy using polyproteins.** *Chem Soc Rev* 2012, **41**:4781-4796.
 15. Shin G, Yost SA, Miller MT, Elrod EJ, Grakoui A, Marcotrigiano J: **Structural and functional insights into alphavirus polyprotein processing and pathogenesis.** *Proc Natl Acad Sci U S A* 2012, **109**:16534-16539.
 16. Yost SA, Marcotrigiano J: **Viral precursor polyproteins: keys of regulation from replication to maturation.** *Curr Opin Virol* 2013, **3**:137-142.
 17. Kostyuchenko VA, Zhang Q, Tan JL, Ng TS, Lok SM: **Immature and mature dengue serotype 1 virus structures provide insight into the maturation process.** *J Virol* 2013, **87**:7700-7707.
 18. Miller M: **The early years of retroviral protease crystal structures.** *Biopolymers* 2010, **94**:521-529.
 19. Noble CG, Shi PY: **Structural biology of dengue virus enzymes: towards rational design of therapeutics.** *Antiviral Res* 2012, **96**:115-126.
 20. Konvalinka J, Kräusslich HG, Müller B: **Retroviral proteases and their roles in virion maturation.** *Virology* 2015. pii: S0042-6822(15)00157-9.
 21. Bell NM, Lever AM: **HIV Gag polyprotein: processing and early viral particle assembly.** *Trends Microbiol* 2013, **21**:136-144.

22. Briggs JA, Kräusslich HG: **The molecular architecture of HIV-1**. *J Mol Biol* 2011, **410**:491-500.
23. Sundquist WI, Kräusslich HG: **HIV-1 assembly, budding, and maturation**. *Cold Spring Harb Perspect Med* 2012, **2**:a006924.
24. Bharat TA, Davey NE, Ulbrich P, Riches JD, de Marco A, Rumlova M, Sachse C, Ruml T, Briggs JA: **Structure of the immature retroviral capsid at 8 Å resolution by cryo-electron microscopy**. *Nature* 2012, **487**:385-389.
25. de Marco A, Heuser AM, Glass B, Kräusslich HG, Müller B, Briggs JA: **Role of the SP2 domain and its proteolytic cleavage in HIV-1 structural maturation and infectivity**. *J Virol* 2012, **86**:13708-13716.
26. Bharat TA, Castillo Menendez LR, Hagen WJ, Lux V, Igonet S, Schorb M, Schur FK, Kräusslich HG, Briggs JA: **Cryo-electron microscopy of tubular arrays of HIV-1 Gag resolves structures essential for immature virus assembly**. *Proc Natl Acad Sci U S A* 2014, **111**:8233-8238.
27. Schur FK, Hagen WJ, Rumlová M, Ruml T, Müller B, Kräusslich HG, Briggs JA: **Structure of the immature HIV-1 capsid in intact virus particles at 8.8 Å resolution**. *Nature* 2015, **517**:505-508.
- The structure of the immature HIV-1 capsid in intact virus particles formed by HIV Gag precursor polyprotein was elucidated by electron cryotomography and sub-tomogram averaging methods, revealing unexpected structural plasticity.
28. Kennedy MW: **The polyprotein lipid binding proteins of nematodes**. *Biochim Biophys Acta* 2000, **1476**:149-164.
29. McAleer MA, Irvine AD: **The multifunctional role of filaggrin in allergic skin disease**. *J Allergy Clin Immunol* 2013, **131**:280-291.
30. Harding CR, Aho S, Bosko CA: **Filaggrin — revisited**. *Int J Cosmet Sci* 2013, **35**:412-423.
31. Meenan NA, Ball G, Bromek K, Uhrin D, Cooper A, Kennedy MW, Smith BO: **Solution structure of a repeated unit of the ABA-1 nematode polyprotein allergen of *Ascaris* reveals a novel fold and two discrete lipid-binding sites**. *PLoS Negl Trop Dis* 2011, **5**:e1040.
32. Smith JL, Skiniotis G, Sherman DH: **Architecture of the polyketide synthase module: surprises from electron cryo-microscopy**. *Curr Opin Struct Biol* 2015, **31**:9-19.
33. Bukhari HS, Jakob RP, Maier T: **Evolutionary origins of the multienzyme architecture of giant fungal fatty acid synthase**. *Structure* 2014, **22**:1775-1785.
34. Tran TH, Hsiao YS, Jo J, Chou CY, Dietrich LE, Walz T, Tong L: **Structure and function of a single-chain, multi-domain long-chain acyl-CoA carboxylase**. *Nature* 2015, **518**:120-124.
35. Eisfeld AJ, Neumann G, Kawaoka Y: **At the centre: influenza A virus ribonucleoproteins**. *Nat Rev Microbiol* 2015, **13**:28-41.
36. Ortín J, Martín-Benito J: **The RNA synthesis machinery of negative-stranded RNA viruses**. *Virology* 2015. pii: S0042-6822(15)00154-3.
37. Bieniossek C, Imasaki T, Takagi Y, Berger I: **MultiBac: expanding the research toolbox for multiprotein complexes**. *Trends Biochem Sci* 2012, **37**:49-57.
38. Thorsen TS, Matt R, Weis WI, Kobilka BK: **Modified T4 lysozyme fusion proteins facilitate G protein-coupled receptor crystallogenesis**. *Structure* 2014, **22**:1657-1664.
39. Yin J, Mobarec JC, Kolb P, Rosenbaum DM: **Crystal structure of the human OX2 orexin receptor bound to the insomnia drug suvorexant**. *Nature* 2015, **519**:247-250.
40. Saraogi I, Shan SO: **Co-translational protein targeting to the bacterial membrane**. *Biochim Biophys Acta* 2014, **1843**:1433-1441.
41. Zhang X, Shan SO: **Fidelity of cotranslational protein targeting by the signal recognition particle**. *Annu Rev Biophys* 2014, **43**:381-408.
42. Schaffitzel C, Oswald M, Berger I, Ishikawa T, Abrahams JP, Koerten HK, Koning RI, Ban N: **Structure of the *E. coli* signal recognition particle bound to a translating ribosome**. *Nature* 2006, **444**:503-506.
43. Zhang X, Schaffitzel C, Ban N, Shan SO: **Multiple conformational switches in a GTPase complex control co-translational protein targeting**. *Proc Natl Acad Sci U S A* 2009, **106**:1754-1759.
44. Estrozi LF, Boehringer D, Shan SO, Ban N, Schaffitzel C: **Cryo-EM structure of the *E. coli* translating ribosome in complex with SRP and its receptor**. *Nat Struct Mol Biol* 2011, **18**:88-90.
45. Scholl ZN, Li Q, Marszalek PE: **Single molecule mechanical manipulation for studying biological properties of proteins, DNA, and sugars**. *Wiley Interdiscip Rev Nanomed Biotechnol* 2014, **6**:211-229.
- Authoritative overview article presenting the state-of-the-art of single-molecule structure analysis of biological samples (proteins, nucleic acid, sugars) by using atomic force microscopy.
46. Krüger M, Linke WA: **The giant protein titin: a regulatory node that integrates myocyte signaling pathways**. *J Biol Chem* 2011, **286**:9905-9912.
47. Bang ML, Centner T, Fornoff F, Geach AJ, Gotthardt M, McNabb M, Witt CC, Labeit D, Gregorio CC, Granzier H, Labeit S: **The complete gene sequence of titin, expression of an unusual approximately 700-kDa titin isoform, and its interaction with obscurin identify a novel Z-line to I-band linking system**. *Circ Res* 2001, **89**:1065-1072.
48. Williams PM, Fowler SB, Best RB, Toca-Herrera JL, Scott KA, Steward A, Clarke J: **Hidden complexity in the mechanical properties of titin**. *Nature* 2003, **422**:446-449.
49. Valbuena A, Oroz J, Vera AM, Gimeno A, Gómez-Herrero J, Carrión-Vázquez M: **Quasi-simultaneous imaging/pulling analysis of single polyprotein molecules by atomic force microscopy**. *Rev Sci Instrum* 2007, **78**:113707.
50. Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM: **Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering**. *Prog Biophys Mol Biol* 2000, **74**:63-91.
51. Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM: **The mechanical stability of ubiquitin is linkage dependent**. *Nat Struct Biol* 2003, **10**:738-743.
52. He C, Genchev GZ, Lu H, Li H: **Mechanically untying a protein slipknot: multiple pathways revealed by force spectroscopy and steered molecular dynamics simulations**. *J Am Chem Soc* 2012, **134**:10428-10435.
53. Cao Y, Li H: **Polyprotein of GB1 is an ideal artificial elastomeric protein**. *Nat Mater* 2007, **6**:109-114.
54. Li YD, Lamour G, Gsponer J, Zheng P, Li H: **The molecular mechanism underlying mechanical anisotropy of the protein GB1**. *Biophys J* 2012, **103**:2361-2368.

Publication n°2:

Alexandre Monod, Christopher Swale, Bogdan Tarus, Alice Tissot, Bernard Delmas, Rob WH Ruigrok, Thibaut Crépin & Anny Slama-Schwok

Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza

Expert Opin. Drug Discov. (2015) 10(4):345-371

EXPERT OPINION

1. Introduction
2. Novel antiviral candidates targeting external viral proteins
3. Structural basis for small-molecules inhibition of individual proteins of the RNP
4. Targeting the non-structural protein 1
5. Conclusion
6. Expert opinion

Learning from structure-based drug design and new antivirals targeting the ribonucleoprotein complex for the treatment of influenza

Alexandre Monod, Christopher Swale, Bogdan Tarus, Alice Tissot, Bernard Delmas, Rob WH Ruigrok, Thibaut Crépin & Anny Slama-Schwok[†]

[†]*Virologie et Immunologie Moléculaires, INRA UR, Jouy en Josas, France*

Introduction: Influenza viruses are a threat to human health. There are presently only two methods for treating influenza: vaccines, which require yearly updates, and two classes of antivirals that suffer with the problem of resistance by current human influenza viruses; this is especially the case with amantadine and rimantadine. Consequently, there is an urgent need for the development of new antivirals with new mechanisms of action.

Areas covered: In this review, the authors focus on viral protein domains, their associated activity and their inhibition by small molecules defined by a structure-based design with a special emphasis on the ribonucleoprotein complex and its inhibitors. Several new classes of antiviral candidates targeting viral replication through individual domains of the polymerase and the nucleoprotein (NP) have been developed through structure-based design.

Expert opinion: To date, the antivirals targeting neuraminidase are by far the most developed and potent. Antiviral candidates targeting the NP and polymerase domains are in the pipeline but their pharmacokinetics needs further studies. The recently published structures of the polymerase expand the possibilities for development of new antivirals. Combination therapies targeting conserved viral targets and new cellular proteins or exploiting drug promiscuity hold promises to fight against the emergence of resistance.

Keywords: antivirals, cap-binding domain, endonuclease domain, neuraminidase and matrix 2 inhibitors, nucleoprotein inhibitors, replication and transcription, resistance, ribonucleoprotein complex

Expert Opin. Drug Discov. (2015) 10(4):345-371

1. Introduction

Influenza viruses are important pathogens of humans and animals. In humans, they cause seasonal epidemics, resulting in about 3 – 5 million yearly cases of severe illness and 500,000 deaths [1]. Highly transmissible influenza A viruses can lead to pandemics as it occurred in 1918, 1957, 1968 and 2009. Influenza viruses have a segmented RNA genome, and exchange of segments between different viruses can lead to viruses with new biological characteristics. In particular, when a human virus mixes its genomic segments with those of an animal virus, a new virus may emerge with novel hemagglutinin (HA) and neuraminidase (NA). Such an occasional event called antigenic shift is associated with pandemics, as previously observed in 1957 and 1968, with the emergence of H2N2 and H3N2 lineages in the human

informa
healthcare

Article highlights.

- There is a need for new antivirals to protect the populations against resistant viral strains to approved antivirals targeting neuraminidase and matrix 2 (M2).
- To overcome resistance, new antiviral candidates targeting the conserved stem region of hemagglutinin and inhibiting its pH-induced structural change are being developed. New structure-based inhibitors of NA and active inhibitors against resistant strains of M2 were reported.
- Within the ribonucleoprotein, inhibitors of the nucleoprotein, modulating its oligomerization and its RNA binding, were developed. Two functional domains of the polymerase were targeted by new antiviral candidates: the PA endonuclease and the PB2 cap-binding domain. PB1–PA interactions were disrupted by novel inhibitors.
- Non-structural protein 1 is a conserved protein that is targeted both at the level of its RNA-binding ability and probably at its effector domain through its interface with cellular factors.
- The new structures of the polymerase hold a lot of possibility for developing new antivirals. Antiviral combinations may probably be the necessary solution for blocking resistance to antivirals.

This box summarizes key points contained in the article.

population, respectively. Antigenic drift results from gradual accumulation of mutations in the surface glycoproteins and leads to annual flu epidemics.

Due to the high mutation rate of the virus, influenza vaccines that promote antibody production against the HA confer protection for no more than a few years. Every year, the World Health Organization estimates which strains of the virus are expected to circulate in the next year, allowing production of vaccines that will provide a strong immunity against the emerging strains. The vaccine is reformulated once a year before the winter periods in the northern and southern hemispheres.

Influenza viruses belong to the Orthomyxoviridae family. Wild aquatic birds constitute the natural host reservoir for a large variety of influenza A viruses, but some influenza viruses infect humans, pigs and other mammals. The type A viruses are the most virulent among the three influenza types. The influenza A viruses can be divided into different serotypes, according to the antibody response against their HAs (H1 – H18) and the NA (N1 to N10), but only H1 – H3 and N1 and N2 are commonly found in humans (H2 was present only from 1957 to 1968). Influenza B almost exclusively infects humans and thus pandemics of influenza B do not occur. Influenza C viruses infect humans, dogs and pigs. Less common than other types, it only causes mild disease in children.

The influenza A genome contains eight RNA segments, each segment encoding one to two main proteins. Whereas three segments encode the HA, the NA and the

matrix 2 (M2) proteins, the five others encode the replicative machinery and proteins controlling the host response and the traffic of the genome in the cell compartments. Entry of the particles into the cell is mediated by endocytosis, and fusion between the viral and the endosomal membranes occurs after endosome acidification. Using microtubule and actin-associated motors, viral ribonucleoproteins (RNPs) migrate toward the nucleus to ensure the primary transcription (Figure 1). The genomic segments are of negative sense, meaning that they are of opposite sense than mRNA, and transcription is the first activity of the virus after entry in the host cell. Transcription and replication are ensured by the viral RNA-dependent RNA-polymerase (RdRp), a multiprotein complex composed of three subunits, PA, PB1 and PB2 which interacts with both nucleoprotein (NP) and the viral genome [2,3]. While replication and transcription occur in the nucleus of the cell, several viral proteins (NEP and M1) are associated with the export of the RNP complexes from the nucleus toward the viral budding site in rafts at the apical side of the plasma membrane. The non-structural protein 1 (NS1) has multiple roles during the virus cycle, the best recognized role being its ability to downregulate the IFN induction in the infected cell.

What are the targets for available antivirals and their role in the viral cycle? Current antiviral drugs used against influenza are inhibitors of detachment of the viral particle from the host cell, NA inhibitors (oseltamivir and zanamivir) and M2 channel protein inhibitors (adamantane derivatives) [4,5]. Their efficiency remains questionable: American H3N2 circulating strains in 2005 were generally resistant to amantadine and rimantadine.

In this review, focus is made on viral protein domains, their associated activity and their inhibition by small molecules defined by a structure-based design with a special emphasis on the RNP complex and its inhibitors. Influenza antivirals were previously actively reviewed [1,2,4-7]. Nucleic acid-based and vaccine approaches are out of the scope of this review. We will first detail the external proteins HA, NA and M2 and their new inhibitors.

2. Novel antiviral candidates targeting external viral proteins

2.1 Hemagglutinin

The HA is a surface glycoprotein of influenza virus required for viral particle entry. HA is a homotrimeric protein formed by a globular head positioned over a stem domain. Each HA monomer contains two disulfide-linked polypeptides HA1 and HA2 that are derived from the proteolytic cleavage of a precursor by host protease. HA binds to sialic acid receptors of the host cell (Figure 1 and Table 1). Seasonal vaccines provide neutralizing antibodies targeting the highly variable region of HA. Carbohydrate-binding agents that recognize glycosylation sites on HA prevent virus adsorption to the cell [8]. Besides its activity as binding to sialic acid receptors,

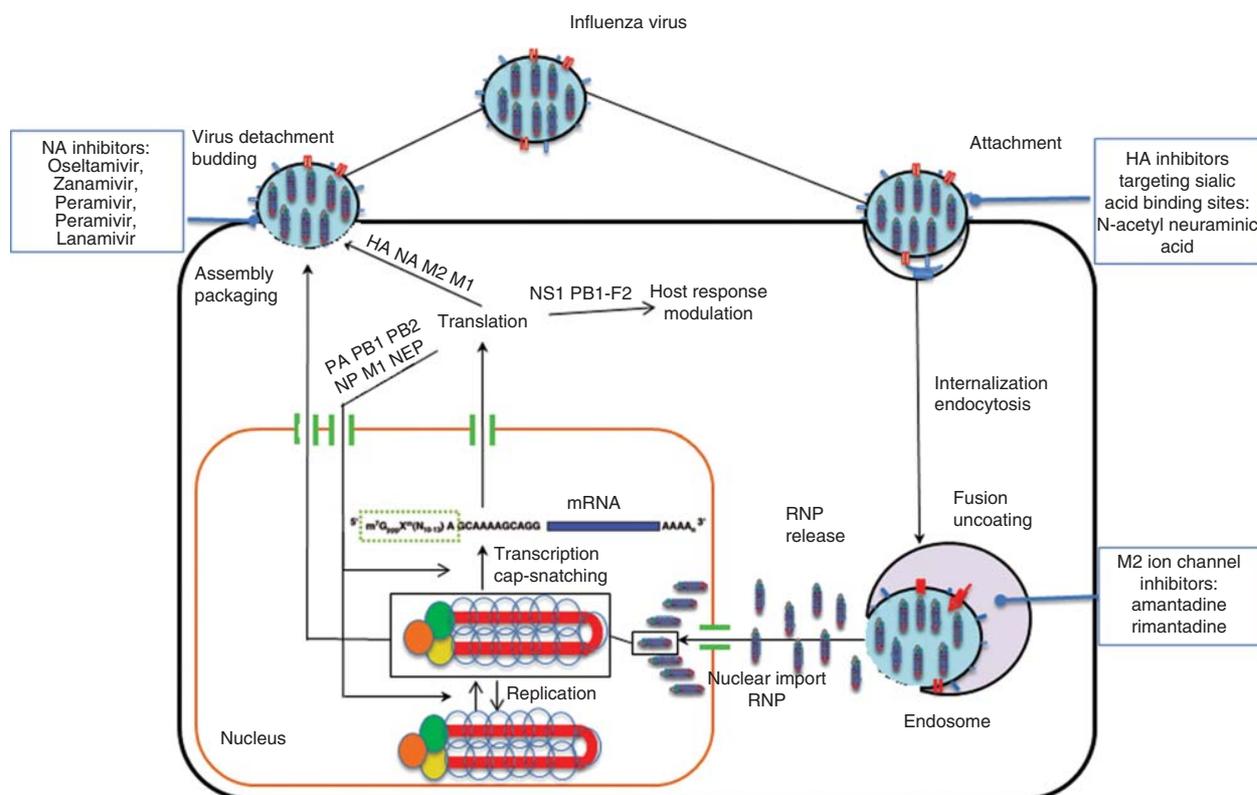
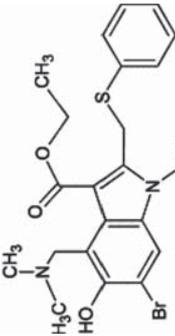
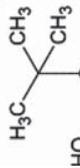
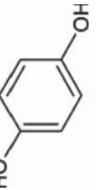
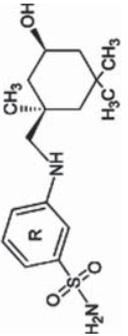
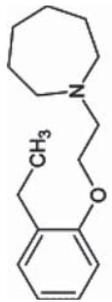


Figure 1. Illustration showing influenza virus' lifecycle: influenza viral particle attaches to the host cell (represented by a black rectangle) by binding to sialic acid receptors mediated by the hemagglutinin. Internalization of the particle is mediated by endocytosis. Following internalization, endosomal acidification initiates conformational changes in HA, driving fusion of virus and cell membranes. The rupture of the membrane of the endosome releases the viral RNP complexes of each of the eight segments of genomic RNA, whereas acidification of the virion interior by the M2 proton channel drives the dissociation of M1 from the viral genome. Nuclear import of the RNP mediated by the cellular transport machinery allows transcription to begin in the nucleus involving a specific cap-snatching mechanism and the formation of a capped messenger RNA. The mRNAs are exported to the cytoplasm and translated into proteins by cellular ribosomes. Newly translated proteins are transported to the nucleus (PA, PB1, PB2, NP, M1, NEP). Formation of new progeny requires replication to take place in the nucleus by the RNP followed by translation. Progeny viral RNPs are then exported to the cytoplasm with the assistance of M1 and NEP and assembly of novel viral particles containing HA, NA, M2 and M1 and their budding at the cell surface that requires the NA to detach. The figures also shows the presently available antivirals targeting M2, NA or HA, the latter being targeted by vaccines. HA: Hemagglutinin; M1: Matrix protein 1; M2: Matrix 2; NA: Neuraminidase; NEP: Nuclear export protein; RNP: Ribonucleoprotein.

HA, together with the M2 proton channel, participates in virus entry by endocytosis through fusion of virus and cell membranes and release of the viral RNP complexes from the acidified endocytic particle. This activity is inhibited by a variety of small molecules that prevent the fusogenic activity by blocking the irreversible conformational change of HA induced by acidic pH. The drug arbidol is representative of such a class of antivirals which inhibits HA-mediated membrane fusion by increasing the stability of the HA to low pH [9,10]. The advantage of arbidol also resides in its broad spectrum activity not only against influenza A and B but also against other viruses such as respiratory syncytial virus (RSV), parainfluenza virus, Coxsackievirus, rhinovirus, hepatitis B

virus and hepatitis C virus. This broad spectrum activity may also be enhanced by arbidol interactions with additional targets as phospholipids of the cell membrane. Binding of small molecules as *t*-butyl benzene-1,4 diol at the interface between two HA monomers likely inhibit HA by avoiding structural changes required for membrane fusion [11]. Structural optimization of salicylamide-based HA inhibitor led to the design of benzenesulfonamide derivatives that stabilized the pre-fusion of HA structure [12]. Inhibition of conserved sites in the stem region of HA holds promise for broad spectrum protective effects [13]. Small-molecule inhibitors derived from aminoalkyl phenol or sulfonamide scaffolds targeting pockets in the stem region were recently reported [14].

Table 1. External proteins of influenza virus and their inhibitors binding sites*.

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
HA	Surface glycoprotein associated with membrane fusion		Interface between 2 HA in trimer	Arbidol/ approved in Russia and China	IC ₅₀ = 7 – 21 μM H2N2, H7N7 CC ₅₀ ≥ 120 μM	Leneva et al. (2009) [9,10]
			3EYK	Inhibitor of membrane fusion- t-butyl benzene-1,4 diol	nd	Russel et al. (2008) [11]
			HA2 in HA trimer	Benzenesulfonamide derivatives†/preclinical	IC ₅₀ = 86 – 210 nM H1N1 CC ₅₀ = 71 – > 100 μM Kd = 200 nM	Tang et al. (2011) [12]
			Stem region 3R2X	Protein		
			Stem region	MBX2329 [§]	IC ₅₀ = 0.3 – 5 μM (H1N1-resistant, H5N1); no activity against H3N2	Fleishman et al. (2011) [13] Basu et al. (2014) [14]

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[(5-(thiophen-2-yl)-1,2-oxazol-3-yl)methyl]tricyclo[3.3.1.1 ~ 3.7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

†The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

‡Both compounds exhibit strong synergy with oseltamivir.

§These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.

#Laninamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-(diaminomethyleneamino)-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

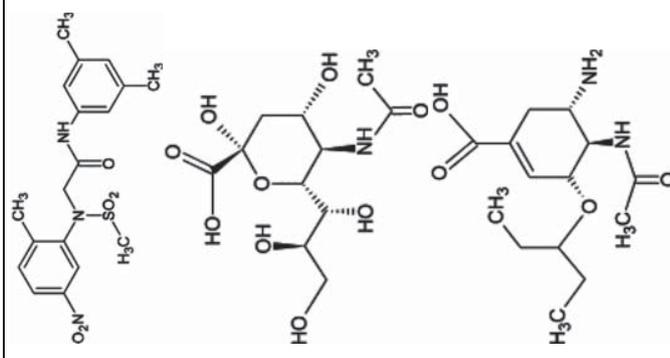
††The pharmacokinetics properties of the 2,3-difluoroisalic acid (DFSA) derivative FaxGudFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

‡‡S31N-resistant strain arising from wsn in Ref. 3.

§§The full chemical names of the compounds represented in the third column are listed above.

¶¶HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

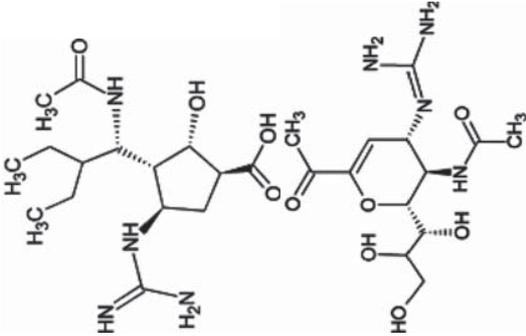
Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
NA	Surface glycoprotein releasing progeny viral particles		MBX2346/preclinical	Osetamivir [†] , /approved	CC ₅₀ > 100 μM IC ₅₀ = 0.8 – 0.5 nM N5 p09N1; CC ₅₀ > 35 μM	Russell et al. (2006) [16]

Active site '150 cavity' ZHT7, ZHTU

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(penta-3-yloxy)-1-thia-3-azaspiro[4.5]deca-6-ene-7-carboxylic acid.
 B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(penta-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.
 C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(penta-3-yloxy)cyclohex-1-ene-1-carboxylic acid.
 D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-2-oxazol-3-yl]methyltricyclo[3.3.1.1 ~ 3.7 ~]decan-1-aminium.
 *Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.
[†]The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.
[§]Both compounds exhibit strong synergy with oseltamivir.
[¶]These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.
[#]Lanamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.
^{**}AV580: (3R,4R,5S)-5-(diaminomethyleneamino)-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.
^{††}The pharmacokinetics properties of the 2,3-difluoroisalic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].
^{‡‡}S31N-resistant strain arising from wsn in Ref. 3.
^{¶¶}The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
			active site 2HTU	Peramivir/approved Japan	IC ₅₀ = 3.4 nM	
			Active site 3CKZ	Zanamivir/approved	IC ₅₀ = 0.6 – 1.1 nM N5 p09N1; CC ₅₀ > 30 μM	

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]dec-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-1,2-oxazol-3-yl]methyltricyclo[3.3.1.1 ~ 3.7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

†The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

‡Both compounds exhibit strong synergy with oseltamivir.

§These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1; pandemic H1N1 2009, N5; H12N5.

#Lanamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5; H12N5.

**AV580: (3R,4R,5S)-5-[(diaminomethylene)amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

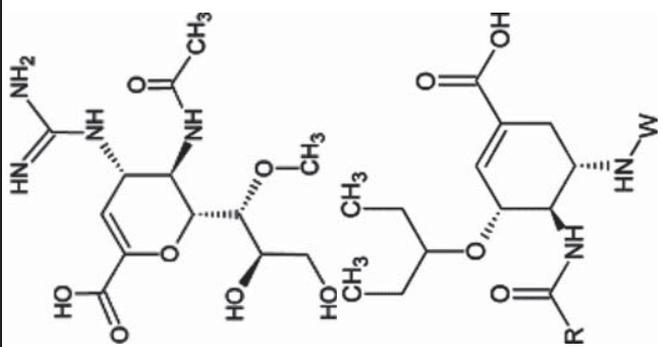
††The pharmacokinetics properties of the 2,3-difluoroacetic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

‡‡S31N-resistant strain arising from wsn in Ref. 3.

§§The full chemical names of the compounds represented in the third column are listed above.

¶¶HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
			Active site 3T18	Laninamivir [#] /approved in Japan	IC ₅₀ = 0.9 – 1.8 nM N5 p09N1	Vavricka (2011) [18]
			Active site	Tight inhibitor AV5080***/preclinical	IC ₅₀ = 0.03 – 0.06 nM (H5N1 p09 N1); CC ₅₀ > 690µM IC ₅₀ = 0.7 – 1 µM p09N1 H5N1	Ivachtchenko et al. (2014) [19]

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(penta-3-yloxy)-1-thia-3-azaspiro[4.5]dec-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(penta-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(penta-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[(5-(thiophen-2-yl)-1,2-oxazol-3-yl)methyl]tricyclo[3.3.1.1 ~ 3.7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

†The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

‡Both compounds exhibit strong synergy with oseltamivir.

§These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.

#Laninamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-[(diaminomethylene)amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino]cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

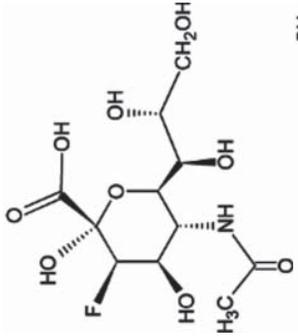
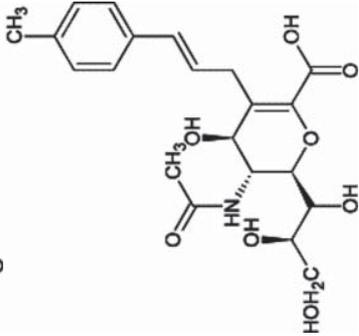
‡‡The pharmacokinetics properties of the 2,3-difluoroacetic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

§§S31N-resistant strain arising from wsn in Ref. 3.

¶¶The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or K _d strain/CC ₅₀	Ref.
			4H52	Covalent inhibitor ^{††} /preclinical	IC ₅₀ = 1.7 – 7.7 μM; p09N1 H5N1	Vavricka et al. (2013) [20] Kim et al. (2013) [21]
			3O9K	C3-substituted sialic acid derivatives/preclinical	K _i = 0.46 nM H1N1, H3N2	Rudrawar et al. (2010) [22]

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]dec-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-1,2-oxazol-3-yl]methyltricyclo[3.3.1.1 ~ 3,7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

[†]The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

^{††}Both compounds exhibit strong synergy with oseltamivir.

[‡]These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.

[#]Lanamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-[(diaminomethylene)amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino]cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

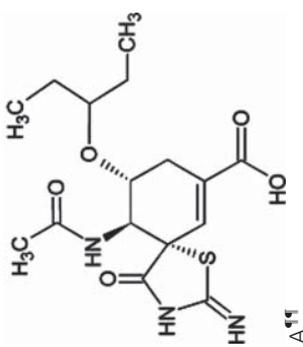
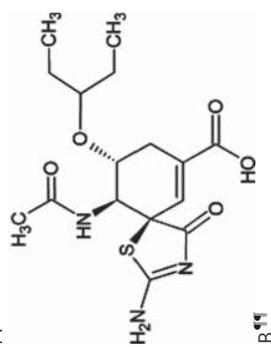
^{†††}The pharmacokinetics properties of the 2,3-difluorosialic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

^{§§}S31N-resistant strain arising from wsn in Ref. 3.

^{¶¶}The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
		 <p>A^{¶¶}</p>	4MJV	Spirolactam derivatives (preclinical)	K _i = 1.5 nM, H1N1	Mohan <i>et al.</i> (2014) [23]
		 <p>B^{¶¶}</p>	4MJU			

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]dec-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-1,2-oxazol-3-yl]methyl]tricyclo[3.1.1 ~ 3,7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

†The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

‡Both compounds exhibit strong synergy with oseltamivir.

¶These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.

#Laninamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-[(diaminomethylene)amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic

properties and a favorable pharmacokinetics profile in animals.

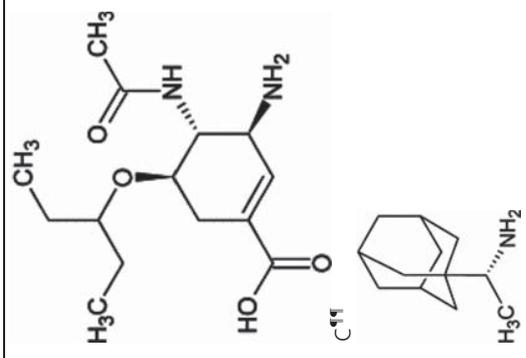
††The pharmacokinetics properties of the 2,3-difluoroacetic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

‡‡S31N-resistant strain arising from wsn in Ref. 3.

¶¶The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
M2	Proton channel		4K51	Triazole-substituted sialic acid derivatives /preclinical		Kerry et al. (2013) [24]
			Pore site 2RLF	Rimantadine/approved	0.1 μM wsn, 200 μM S31N ^{§§}	Schnell and Chou (2008) [26]

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-1,2-oxazol-3-yl]methyltricyclo[3.3.1.1 ~ 3,7 ~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

[†]The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

[‡]Both compounds exhibit strong synergy with oseltamivir.

[§]These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009. N5: H12N5.

[¶]Lanamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-[[diaminomethylene]amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

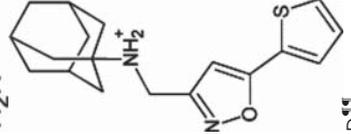
^{‡‡}The pharmacokinetics properties of the 2,3-difluorosialic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

^{§§}S31N-resistant strain arising from wsn in Ref. 3.

^{¶¶}The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 1. External proteins of influenza virus and their inhibitors binding sites* (continued).

Protein	Function	Structure	Binding site/ PDB	Inhibitor/status	IC ₅₀ or Kd strain/CC ₅₀	Ref.
			Pore site 2KQT	Amantadine/approved	100 μM wsn, > 500 μM S31N	Cady <i>et al.</i> (2010) [27]
			Pore site 2LYO	M2-WJ332/preclinical	26 μM S31N	Wang <i>et al.</i> (2013) [3,29]

A = (2E,5S,9R,10S)-10-(acetylamino)-2-imino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-6-ene-7-carboxylic acid.

B = (5R,9R,10S)-10-(acetylamino)-2-amino-4-oxo-9-(pentan-3-yloxy)-1-thia-3-azaspiro[4.5]deca-2,6-diene-7-carboxylic acid.

C = (3S,4R,5R)-4-(acetylamino)-3-amino-5-(pentan-3-yloxy)cyclohex-1-ene-1-carboxylic acid.

D = (3S,5S,7S)-N-[5-(thiophen-2-yl)-1,2-oxazol-3-yl]methyltricyclo[3.3.1.1^{0,2}-3,7-~]decan-1-aminium.

*Only part of the inhibitor structures that are found in the PDB and those involving a structure-based design combined with experimental validation are summarized here.

[†]The derivatives exhibit good water solubility and good stability in human liver microsomes (6 – 8 ml/min/kg) and no inhibition of major CYP450 isoforms.

[‡]Both compounds exhibit strong synergy with oseltamivir.

[§]These NA inhibitors currently approved in the US are only FDA-approved for uncomplicated influenza infections [72]. Peramivir is in clinical trials in US. p09N1: pandemic H1N1 2009, N5: H12N5.

[#]Laninamivir exists as an octanoate prodrug that is slowly processed in adult patients; the prodrug binds as oseltamivir and its efficacy is not improved as compared to oseltamivir against resistant strain. N5: H12N5.

**AV580: (3R,4R,5S)-5-[[diaminomethyl(ene)amino]-3-(1-ethylpropoxy)-4-[(fluoroacetyl)amino] cyclohex-1-ene-1-carboxylic acid R = CF₃ W = 4,5-dihydro-1H-imidaz-2-yl. AV5080 has good solubility and metabolic properties and a favorable pharmacokinetics profile in animals.

^{††}The pharmacokinetics properties of the 2,3-difluoroisalic acid (DFSA) derivative FaxGuDFSA were comparable to those of zanamivir, a 100% efficacy at 10 mg/kg/day with no side effects in infected mice (H3N2) [21].

^{‡‡}S31N-resistant strain arising from wsn in Ref. 3.

^{§§}The full chemical names of the compounds represented in the third column are listed above.

HA: Hemagglutinin; M2: Matrix 2; NA: Neuraminidase; nd: Not determined; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

2.2 Neuraminidase

The NA is a second glycoprotein expressed at the surface of influenza virus which releases the newly formed viral particle by cleaving the terminal sialic acid from HA receptors on cell membranes (Figure 1 and Table 1). NA is a tetrameric enzyme. NA forms two phylogenetic groups with distinctive structural features in the active site, including a protein loop (150-loop) in an open conformation in group 1 (N1, N4, N5 and N8), whereas in group 2 NAs (N2, N3, N6, N7 and N9) this loop is in a closed conformation. The currently used antivirals are oseltamivir (Tamiflu) and zanamivir (Relenza) that target the active site of NA and represent successful early structure-based designed compounds (Table 1) [15,16]. Zanamivir has a poor solubility and is administered as aerosol, whereas Tamiflu is formulated as a prodrug. Although zanamivir is well tolerated and has few side effects, the route of administration of this drug can represent a problem especially for children and elderly patients who might not be able to inhale zanamivir suitably [3]. To circumvent these issues, an intravenous formulation of zanamivir has been developed and is currently under Phase III clinical trial. Peramivir is a cyclopentane compound, approved in Japan for use in adult and pediatric patients with influenza A virus and influenza B virus infection [17]; in addition, it is currently undergoing clinical trials in the US and in other countries. This compound is administered only as an intravenous formulation, due to its very low bioavailability. Laninamivir targets more specifically class 2 NA and is potent, in particular, against the 2009 pandemic strain [18]. The octanoate prodrug of laninamivir has binding and inhibitory properties that make it like Tamiflu (Table 1). Many mutations of NA have been identified in patients, of which some confer resistance to treatment [1]. In 2009, the swine flu pandemic (p09N1) was sensitive to oseltamivir but resistance develops in this strain as observed earlier in the H274Y H1N1-resistant strains.

Based on crystallographic structures, novel anti-NA compounds have been developed, some of them being very tight binders as AV5080, whereas others act as 'suicide inhibitors' forming a covalent bond with NA active site [19-21]. Exploitation of the cavity formed by the open-loop conformation (the 150-cavity) led to the development of novel sialic acid derivatives, such as oseltamivir derivatives modified with a spirolactam ring or carrying various side-chains [22-24]. In the C4 position, the OH group of sialic acid was replaced by a guanidinium in peramivir, zanamivir and laninamivir and by an amine in oseltamivir and AV580. The amide group in the C5 position and carboxylic acid in the second position are shared by most compounds and sialic acid, the latter being recognized by an arginine (R292). Conjugates of polymers (biodegradable polyglutamine) with multiple copies of zanamivir interfered more efficiently than zanamivir with intracellular trafficking of the endocytosed viruses and the subsequent virus-endosome fusion [25]. Their advantage resides in a

higher antiviral activity by 1000- to 10,000-folds and *in vivo* longer-lasting effects than the monomeric drug.

2.3 Matrix 2

M2 is a homotetramer composed of four parallel transmembrane α -helices positioned across the viral lipid envelope, forming the ion channel where proton conduction takes place. In response to low pH in the endosome, M2 creates a proton flux from the endosome to the virion core (Figure 1). M2 is also responsible for the delayed acidification of the Golgi before viral assembly to insure correct structures of the other viral proteins. The N-terminal end of the channel is gated by the V27 residue, and the channel width (pore) is controlled by four additional pore-facing residues such as A30, S31 (N31 in resistant strains), G34, H37 and W41 that constitute the functional core of the channel (Figure 2). H37 is the pH sensor, whereas W41 acts as a C-terminal gate and is adjacent to amphipathic helices stabilizing the channel. The above pore-controlling residues constitute the high-affinity binding site for amantadine and rimantadine that inhibit proton conduction by pore occlusion [26-28]. Adamantanes were the first anti-influenza drugs approved, although presenting CNS side effects. Resistance to adamantanes as adamantane is the main problem of most present circulating or threatening influenza virus, including avian H5N1. The hydrophobic adamantane moiety interacts with the channel gate V27, while their positively charged group interacts with H37. The channel adapts to saturated polycyclic compounds of various sizes [29] and amantadine substituted with aromatic side chain as thienyl group in M2WJ332 [3] were potent inhibitors of drug-resistant S31N mutant of M2 (Figure 2 and Table 1).

Resistance to antiviral treatment addressed above led to the targeting of the RNP complex, the viral minimal machinery for all transcription and replication [2,30]. Because these proteins are usually less prone to mutations, and some of these proteins as the NP are unique to virus and have no equivalent in the host, targeting the RNP by small molecules could tend to an ultimate goal in antiviral treatment: i) to be broad spectrum and effective against various strains and ii) to generate minimal escape from treatment and if possible be effective on existing resistant strains.

3. Structural basis for small-molecules inhibition of individual proteins of the RNP

3.1 Nucleoprotein

NP is a highly conserved protein that binds to and protects the viral genomic or anti-genomic RNA [30-32]. NP has a structural chaperone-like role in the assembly of the RNP; however, much of the structural information on how RNA binds to NP is still incomplete, and the assembly of NP-RNA oligomers into RNP mainly relies on exciting cryo-electron microscopy (EM) studies at 20 Å resolution [33-35]. The reconstruction showed a double helical stem region of NP oligomers on viral single-stranded RNA in which NP

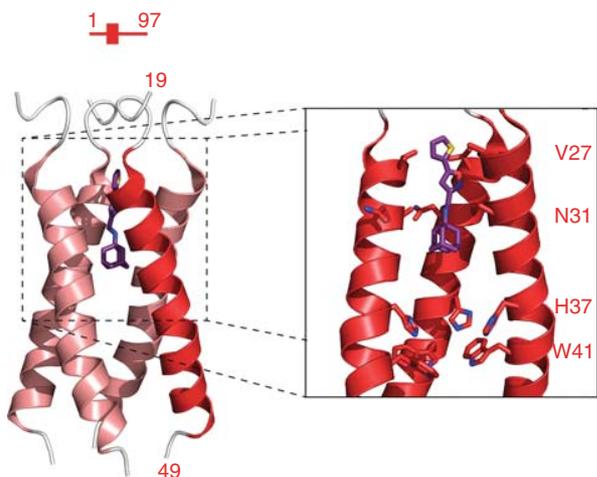


Figure 2. Illustration showing part of the M2 proton channel constituted by the assembly of four parallel transmembrane α -helices across the viral coat. The core of the channel involves V27 as the gate to the channel, and important residues N31, H37 and W41 are represented. The channel lumen is blocked by the inhibitor M2-WJ332 that is active against the resistant S31N protein [3,29].

mediated the assembly of the two strands. Interactions of NP with the polymerase subunits through other types of interactions were also observed, consistent with NP–PB2 interactions characterized by other methods [36–38]. How NP discriminates these viral RNAs (viral RNA of negative polarity [vRNA] and complementary RNA intermediate of positive polarity [cRNA]) from viral mRNA and cellular RNAs is still unknown [39]. NP is hypothesized to contribute to switching between transcription and replication that take place in the nucleus of infected cells; the exact role of NP in these processes is an active field of research [40–42]. NP binds single-stranded viral RNA as oligomers that are essential for RNP complex activity on long templates [43]. NP self-association is mediated by a flexible tail-loop that protrudes into a pocket in adjacent subunit on trimer formation; a dimeric interface was also defined [44–46]. The tail loop position (residues 402 – 428 of NP) differs within the influenza A H1N1 and H5N1 NP trimer and influenza B tetramer [47], thus suggesting that tail loop may support different oligomers [48,49]. NP oligomerization is a dynamic process and NP can also adopt a monomeric structure [30], which is stabilized by phosphorylation and possibly through the interaction with other (cellular) proteins as importin $\alpha 5$ [30,50–53]. NP structure is characterized by a head and body domain that are separated by a highly basic groove involved in RNA-binding. Figure 3 and Table 2 recapitulate the present antiviral candidates targeting NP for which structure is known or estimated by molecular dynamic simulations and confirmed by experimental evidence. The ability of NP to self-associate and form NP–RNA oligomers required for RNP function was targeted by

two opposite approaches (Figure 3A). In the first one, nucleozin stabilized the interface between NP monomers inducing the formation of nonfunctional higher order oligomers that modified the RNP trafficking to the nucleus (Figure 3A) [54–56]. Nucleozin was the first molecule identified by high-throughput screening using the nuclear transport of NP as readout. The nucleozin-binding sites were identified by structural and modeling studies and were confirmed by escape mutations to nucleozin analog treatment at the interface between two NP molecules in a hydrophobic pocket formed by Y289, N309 and Y52 and Y313 (Figure 3C) [57]. Nucleozin and its derivatives bear the arylpiperazine scaffold associated with a relatively low solubility and yet their pharmacokinetics was not reported. These compounds are presently the most potent compounds targeting NP. Shen *et al.* used an opposite approach that avoided oligomer formation by blocking the insertion of the oligomerization loop targeting the E339–R416 salt bridge with peptides or compound 3 bearing a thiazole carboxamide motif (Figure 3A, center and Table 2) [58]. Competing with NP binding to RNA was another approach that identified naproxen that bound in a small hydrophobic site formed by Y148, F489 and R361 in the RNA-binding groove of NP based on molecular modeling and confirmed by surface plasmon resonance (Figure 3A, right and 3B) [59]. Naproxen combined an antiviral activity with an anti-inflammatory effect; being also a generic drug inhibiting the inducible COX, naproxen is an approved drug (anti-COX activity) and as such has favorable solubility and stability properties. Naproxen and its designed derivatives were shown to stabilize monomeric NP by modifying the cavity in which the oligomerization loop inserts, close to their binding site [60].

Within the context of the RNP, the RdRp, composed by the three subunits PA, PB1 and PB2, mediates replication and transcription of the segmented viral RNA. The transcription process involves a unique cap-snatching mechanism in which the RdRp binds to host cellular pre-mRNAs via their 5'-cap structure and then cleaves them to generate capped oligonucleotides, which are used as primers to initiate the transcription of viral mRNAs. Owing to the cap, viral mRNAs are translated into proteins by the cell machinery. The cap-snatching mechanism is an essential and conserved process in all members of the Orthomyxoviridae family, including influenza A, B and C viruses. Because the host cell has no analogous activity, cap-snatching inhibitors could be specific against all influenza subtypes and strains. Cap-snatching processes can be inhibited by targeting either the PB2 cap-binding or PA endonuclease domains. Influenza polymerase domains such as PA endonuclease [61,62], PA C-terminal domain associated with PB1 N-terminal peptide [63,64], PB1 C-terminal associated with PB2 N-terminal [65], PB2 Cap-binding domain [66] as well large parts of PB2 C-terminal [67] were solved and these structures provided the basis for structure-based design of specific inhibitors as detailed below. During the course of writing this review, the structures of the

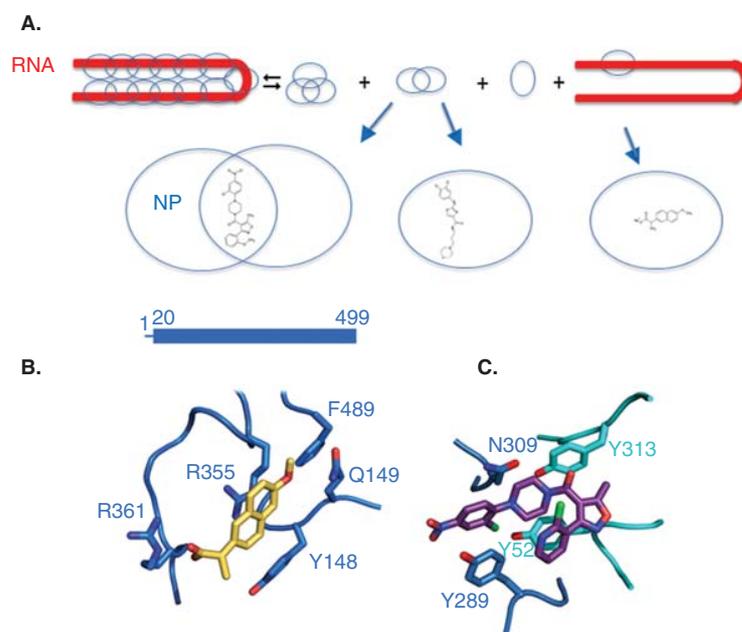


Figure 3. Illustration showing inhibitors of the nucleoprotein NP: two opposite strategies have been developed for modulating NP oligomerization loop as explained by the cartoon shown in panel (A). NP is found in various assemblies with and without single-stranded RNA that may interconvert depending on the conditions. Attempts to decrease NP self-association through binding to the oligomerization [58] (panel A, center) or to impede RNA binding to the binding groove of monomeric NP (panel (A), right and yellow naproxen in panel (B) [59,60] is presented. The nucleozin derivative, shown in purple, stabilizes the interface between two adjacent NP monomers and favors nonfunctional oligomers of NP (panel (A), left and panel (C) the residues of the two NP are shown in cyan and blue) [55,57].

whole polymerase of influenza A and B were published [68,69]. Figure 4 shows the overall structure of Influenza A polymerase; the previously solved domains of PA endonuclease (green) and PB2 Cap-binding (orange) with bound inhibitors are highlighted.

3.2 PB2 cap-binding

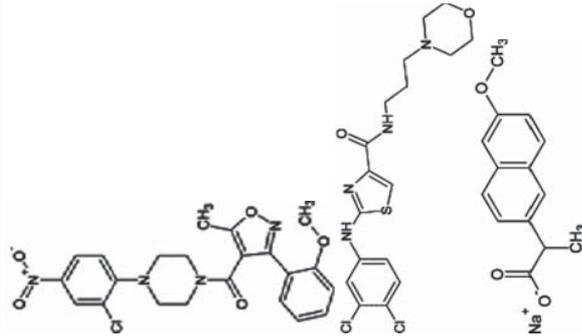
Transcription by the cap-snatching mechanism requires that host-cell pre-mRNAs are bound via their 5'-cap structure to the PB2 subunit. Thus, the PB2 cap-binding site is a potential target for new drugs to inhibit viral replication. The X-ray structure of the cap-binding site with bound 7-methylguanosine triphosphate (m^7GTP) has been obtained [66] and revealed that the m^7GTP is sandwiched between two aromatic residues (Phe404 and His357), forming strong π -stacking interactions. The guanine base is also recognized by hydrogen bonds to key PB2 residues (Lys376 and Glu361) and several residues are involved in interactions with phosphate groups. Docking studies using the structure of the PB2 cap-binding domain suggested that different 7-methylguanine derivatives are good candidates [70]. Crystal structures of cap-binding domain-inhibitor complexes have been obtained with three active m^7GTP derivatives and confirmed a unique binding mode of the m^7GTP scaffold in the PB2 recognition site for these compounds and the m^7GTP [70]. More recently, a series

of azaindole derivatives targeting PB2 has been identified [71]. The X-ray crystal structure of a pyrimidine azaindole derivative denoted VX787 (described in Table 2) bound to PB2 [71] confirmed that VX787 occupies the m^7GTP -binding site where the azaindole ring makes interactions with the key residues, Lys376 and Glu361, as well as stacking of the three aromatic rings of VX787 between the side chains of His357, Phe323 Phe404 and Phe325 (Figure 4, orange inset). Orally bioavailable, this compound shows a strong potency for a large number of influenza A strains, including pandemic 2009 H1N1 and avian H5N1 flu strains in a mouse influenza model. VX787 is the one of the two compounds targeting the polymerase actually under clinical trial. The other compound is T705, a ribonucleotide analog (6-fluoro-3-hydroxy-2-pyrazinecarboxamide) also known as favipiravir that inhibits many viral RdRps, thus being effective against a variety of viruses [72].

3.3 PA endonuclease

The endonuclease activity necessary for the cleavage of the host cellular pre-mRNAs is carried out by the N-terminal domain of PA subunit (Nter-PA, ~ 25 kDa, residues 1 – 196). The structure of Nter-PA has been solved in several crystal forms with (or without) various ligands [61,62,73-79]. Two years after the first structure of the PA catalytic domain

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites.

Protein	Function	Structure*	Binding site/PDB	Inhibitor*/ Status†	IC ₅₀ or K _d strain/ CC ₅₀	Ref.
NP	RNA-binding		Hydrophobic pocket 3RO5	Nucleozine Nucleozine derivatives	IC ₅₀ = 69 nM (WSN) 160 nM (H3N2) 330 nM (H5N1); CC ₅₀ > 250 μM IC ₅₀ = 520 nM (H1N1 2009), 2 μM (H3N2) 690 nM (H5N1); CC ₅₀ > 100 μM	Kao et al. (2010) [55] Gerritz et al. (2011) [57]
			Oligomerization loop	Compound 3	IC ₅₀ = 2.7 μM (WSN); CC ₅₀ = 35.5 μM	Shen et al. (2011) [58]
			RNA-binding groove	Naproxen and derivatives	IC ₅₀ = 16 μM (WSN) 12 μM (H3N2), CC ₅₀ = 1605 μM	Lejal et al. (2013); Tarus et al. (2014) [59,60]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

†All compounds are in preclinical stage except VX787.

Nucleozine derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3,4-dichlorophenylamino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl]-2-(5-fluoranyl)-1H-pyridin-3-yl]pyrimidin-4-yl]amino] bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydroxyphenyl) ketomethyl-7-N-methylguanine.

Hydroxypyrimidine derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxypyrimidin-4-(3H)-one.

Hydroxyquinoline derivative: 7-(4-(fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxypyridine derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetrazol-5-yl)phenyl]-1,2-Dihydroxypyridin-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polyphenol EGCG: epigallocatechin gallate.

Quinoline derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

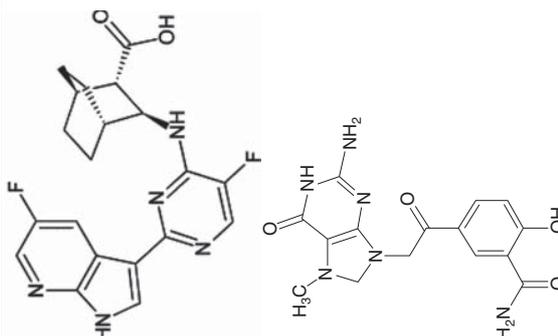
DPBA: 2-(4-Dioxo-4-phenylbutanoic acid).

Pyrimidine derivative: 2-[3-(acetylaminophenyl)-5-hydroxy-6-oxo-1,6-dihydroxypyrimidine-4-carboxylic acid].

Tintori 2014: ethyl 4-[[2-[3-cyano-4-(3-ethoxy-4-phenylmethoxyphenyl)-6-phenylpyridin-2-yl]sulfonyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis(2-furyl)quinoxaline-6-[[1-(2-furyl)methyl]aminocarbonyl]-Nd: Not determined; NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites (continued).

Protein	Function	Structure*	Binding site/PDB	Inhibitor* / Status [‡]	IC ₅₀ or K _d strain/ CC ₅₀	Ref.
PB2	Cap-binding		4P1U Cap-binding site 4NCE 4NCM 4CB5 4CB4,4CB6, 4CB7	Azaindole derivative (VX787) clinical trial Anthracene derivative Pyrimidine derivative 7-methylguanaine derivatives	EC ₅₀ = 0.0026 μM (H1N1)	Clark et al. (2014) [71] Pautus et al. (2013) [70]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

†All compounds are in preclinical stage except VX787.

Nucleosine derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3,4-dichlorophenyl)amino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl-2-(5-fluoranyl-1H-pyrrolo[2,3-b]pyridin-3-yl)pyrimidin-4-yl]amino] bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydroxyphenyl) ketomethyl-7-N-methylguanaine.

Hydroxypyrimidine derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxypyrimidin-4(3H)-one.

Hydroxyquinoline derivative: 7-(4-fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxypyridine derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetrazol-5-yl)phenyl]-1,2-Dihydropyridin-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polyphenol EGCG: epigallocatechin gallate.

Quinoline derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

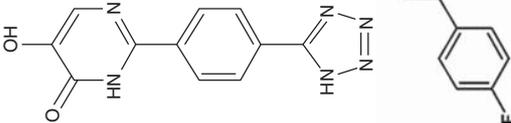
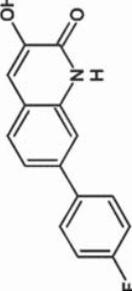
DPBA: 2-(4-Dioxo-4-phenylbutanoic acid.

Pyrimidine derivative: 2-[3-(acetylamino)phenyl]-5-hydroxy-6-oxo-1,6-dihydropyrimidin-4-carboxylic acid.

Tintori 2014: ethyl 4-[2-(3-cyano-4-(3-ethoxy-4-phenylmethoxyphenyl)-6-phenylpyridin-2-yl)sulfonyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis(2-furyl)quinoxaline-6-(1-[(2-furyl)methyl]aminocarbonyl). Nd: Not determined; NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites (continued).

Protein	Function	Structure*	Binding site/PDB	Inhibitor*/ Status [†]	IC ₅₀ or K _d strain/ CC ₅₀	Ref.
PA	Endonuclease		Endonuclease active site 4W9S	Hydroxyquinoline derivative	IC ₅₀ = 0.15 μM (endo. inhibition; H3N2)	Sagong <i>et al.</i> (2014) [77]
			Endonuclease active site 4KIL	Hydroxyquinoline derivative	IC ₅₀ = 0.5 μM (endo. inhibition; H3N2)	Sagong <i>et al.</i> (2013) [78]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

[†]All compounds are in preclinical stage except VX787.

Nucleosine derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3,4-dichlorophenyl)amino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl-2-(5-fluoranyl-1H-pyridin-3-yl)pyridin-3-yl]pyrimidin-4-yl]amino bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydroxyphenyl) ketomethyl-7-N-methylguanine.

Hydroxyquinoline derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxyquinolin-2(1H)-one.

Hydroxyquinoline derivative: 7-(4-fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxyquinoline derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetraol-5-yl)phenyl]-1,2-Dihydropyridin-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polyphenol EGCG: epigallocatechin gallate.

Quinoline derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

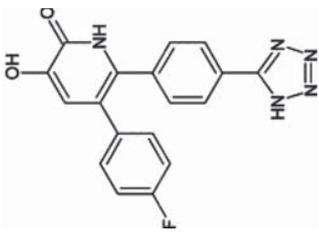
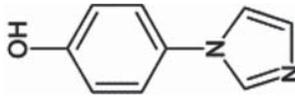
DPBA: 2,4-Dioxo-4-phenylbutanoic acid.

Pyrimidine derivative: 2-[3-(acetylamino)phenyl]-5-hydroxy-6-oxo-1,6-dihydropyrimidin-4-carboxylic acid.

Tintori 2014: ethyl 4-[[2-[3-cyano-4-(3-ethoxy-4-phenylmethoxyphenyl)-6-phenylpyridin-2-yl]sulfonyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis[(2-furyl)quinoxaline-6-(1-[(2-furyl)methyl]aminocarbonyl)Nd: Not determined; NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites (continued).

Protein	Function	Structure*	Binding site/PDB	Inhibitor* / Status [‡]	IC ₅₀ or K _d strain/ CC ₅₀	Ref.
			Endonuclease active site 4M5U	Hydroxypyridine derivative	IC ₅₀ = 0.023 μM (endo. inhibition; H1N1)	Parhi et al. (2013) [73]
			Endonuclease active site 4LN7, 4MK1, 4MK2, 4MK5, 4M5O, 4M4Q, 4M5R	Pyridine derivatives Phenol derivative	IC ₅₀ = 0.041 μM (endo. inhibition; H1N1) IC ₅₀ = 1000 μM (endo. inhibition; H1N1)	Bauman et al. (2013) [73] Kowalinski et al. (2012) [74]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

[‡]All compounds are in preclinical stage except VX787.

Nucleosine derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3-(4-chlorophenyl)amino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl-2-(5-fluoranyl-1H-pyrrol-2,3-b)pyridin-3-yl]pyrimidin-4-yl]amino] bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydroxyphenyl) ketomethyl-7-N-methylguanane.

Hydroxypyrimidine derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxypyrimidin-4(3H)-one.

Hydroxyquinoline derivative: 7-(4-fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxypyridine derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetrazol-5-yl)phenyl]-1,2-Dihydroxy-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polypheol ECGG: epigallocatechin gallate.

Quinoline derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

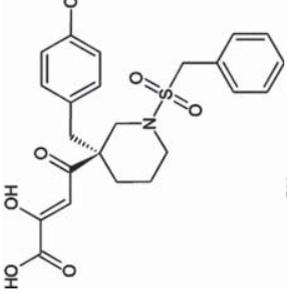
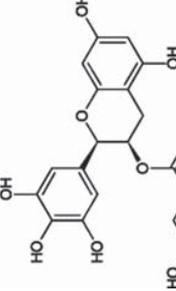
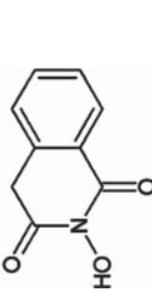
DPBA: 2-(4-Dioxo-4-phenylbutanoic acid.

Pyrimidine derivative: 2-[3-(acetylamino)phenyl]-5-hydroxy-6-oxo-1,6-dihydroxy-4-carboxylic acid.

Tintori 2014: ethyl 4-[2-[3-cyano-4-(3-ethoxy-4-phenylmethoxyphenyl)-6-phenylpyridin-2-yl]sulfonyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis(2-furyl)quinoxaline-6-(1-[(2-furyl)methyl]aminocarbonyl)indole. Not determined. NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites (continued).

Protein	Function	Structure*	Binding site/PDB	Inhibitor*/ Status†	IC ₅₀ or K _d strain/ CC ₅₀	Ref.
			Endonuclease active site 4AWK 4AVG, 4AWG	Diketo derivatives	IC ₅₀ = 0.13 μM (endo. inhibition); IC ₅₀ = 20 μM (antiviral inhibition; H3N2); CC ₅₀ > 100 μM	
			Endonuclease active site 4AWF	DPBA	IC ₅₀ = 2.7 μM (endo. inhibition); IC ₅₀ = inactive (antiviral inhibition; H3N2); CC ₅₀ > 200 μM	
			4AWM	EGCG from green tea	IC ₅₀ = 1.9 μM (endo. inhibition; H3N2); IC ₅₀ = 1.1 μM (antiviral inhibition); CC ₅₀ > 54 μM	
			4EF 4E5G 4E5H	Hydroxyquinoline derivative DPBA Benzylpiperidine derivative	IC ₅₀ = 15 μM (endo. inhibition; H1N1); IC ₅₀ = 20 μM (endo. inhibition; H1N1); IC ₅₀ = 0.43 μM (endo. inhibition; H1N1);	DuBois et al. (2012) [75]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

†All compounds are in preclinical stage except VX787.

Nucleoside derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3,4-dichlorophenyl)amino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl]-2-(5-fluoranyl)-1H-pyridin-3-yl]pyrimidin-4-yl]amino] bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydroxyphenyl) ketomethyl-7-N-methylguanine.

Hydroxypyrimidine derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxypyrimidin-4(3H)-one.

Hydroxyquinoline derivative: 7-(4-(4-fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxyquinoline derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetrazol-5-yl)phenyl]-1,2-Dihydropyridin-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polyphenol EGCG: epigallocatechin gallate.

Quinolone derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

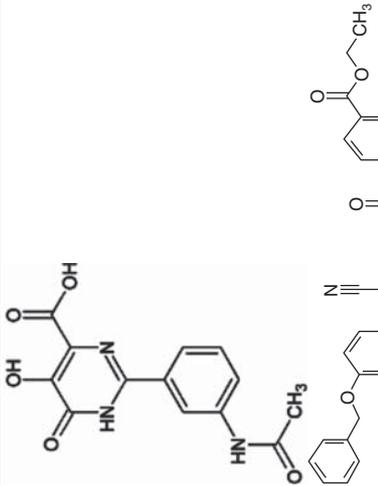
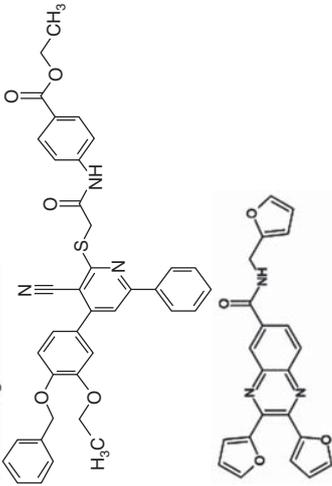
DPBA: 2-(4-Dioxo-4-phenylbutanoic acid).

Pyrimidine derivative: 2-[3-(acetilamino)phenyl]-5-hydroxy-6-oxo-1,6-dihydropyrimidine-4-carboxylic acid.

Tintori 2014: ethyl 4-[[2-[3-cyano-4-(3-ethoxy-4-phenylmethoxyphenyl)-6-phenylpyridin-2-yl]sulfonyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis(2-furyl)quinoxaline-6-[[1-(2-furyl)methyl]aminocarbonyl]Nd: Not determined; NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

Table 2. Internal proteins involved in RNP function and in antagonizing host antiviral response against influenza A virus and their inhibitor binding sites (continued).

Protein	Function	Structure*	Binding site/PDB	Inhibitor*/ Status†	IC ₅₀ or Kd strain/ CC ₅₀	Ref.
PA-PB1 peptide			4E5J 4E5I, 4E5L	Pyrimidine derivatives	IC ₅₀ = 5 μM (endo. inhibition); IC ₅₀ > 50 μM (antiviral inhibition); CC ₅₀ > 60 μM (H1N1)	Tintori et al. (2014) [84] Pagano et al. (2014) [83]
NS1	Antagonist of the cellular antiviral IFN response		RNA-binding site RNA-binding site	Benzoic acid (phenol) derivative Benzofurazan derivatives Quinoxaline derivatives Polyphenol E	30 < IC ₅₀ < 180 μM (H1N1) IC ₅₀ = 45 μM (endo. inhibition); IC ₅₀ = 1 μM (anti-viral inhib.); CC ₅₀ = 20 μM (WSN,H1N1) IC ₅₀ = 0.29 – 9 μM CC ₅₀ > 60 μM	You et al. (2011) [95] Cho et al. (2012) [94]

Adapted in part from Ref. [17] with permission of Future Medicine Ltd.

*The full chemical names of the compounds represented in the third column are listed below.

†All compounds are in preclinical stage except VX787.

Nucleosine derivative: [4-(2-chloro-4-nitrophenyl)piperazin-1-yl] [3-(2-methoxyphenyl)-5-methyl-1,2-oxazol-4-yl]methanone.

Compound 3: 2-((3,4-dichlorophenyl)amino)-N-(3-morpholinopropyl)thiazole-5-carboxamide.

VX787: (2S,3S)-3-[[5-fluoranyl]-2-(5-fluoranyl)-1H-pyridin-3-yl]pyrimidin-4-yl]amino] bicyclo[2.2.2]octane-2-carboxylic acid.

Cap analogue: 9-N-(3-Carboxy-4-hydrophenyl) ketomethyl-7-N-methylguanine.

Hydroxypyrimidine derivative: 2-(4-(1H-tetrazol-5-yl)phenyl)-5-hydroxypyrimidin-4(3H)-one.

Hydroxyquinoline derivative: 7-(4-fluorophenyl)-3-hydroxyquinolin-2(1H)-one.

Hydroxypyridine derivative: 5-(4-Fluorophenyl)-3-hydroxy-6-[4-(1H-1,2,3,4-tetrazol-5-yl)phenyl]-1,2-Dihydropyridin-2-one.

Phenol derivative: 4-(1H-imidazol-1-yl)phenol.

Diketo derivative: (2Z)-4-[(3S)-1-(benzylsulfonyl)-3-(4-chlorobenzyl)piperidin-3-yl]-2-hydroxy-4-oxobut-2-enoic acid.

Polyphenol EGCG: epigallocatechin gallate.

Quinoline derivative: 2-hydroxyisoquinoline-1,3(2H,4H)-dione.

DPBA: 2-(4-Dioxo-4-phenylbutanoic acid).

Pyrimidine derivative: 2-[3-(acetylamino)phenyl]-5-hydroxy-6-oxo-1,6-dihydropyrimidine-4-carboxylic acid.

Tintori 2014: ethyl 4-[[2-[3-cyano-4-(3-ethoxy-4-phenyl)methoxyphenyl]-6-phenylpyridin-2-yl]sulfanylacetyl]amino]benzoate.

Quinoxaline derivative: 2,3-Bis(2-furyl)quinoxaline-6-[1-[(2-furyl)methyl]amino]carboxylate]. Nd: Not determined; NP: Nucleoprotein; NS1: Non-structural protein 1; PDB: Protein Data Bank; WSN: Influenza A/WSN/33 H1N1.

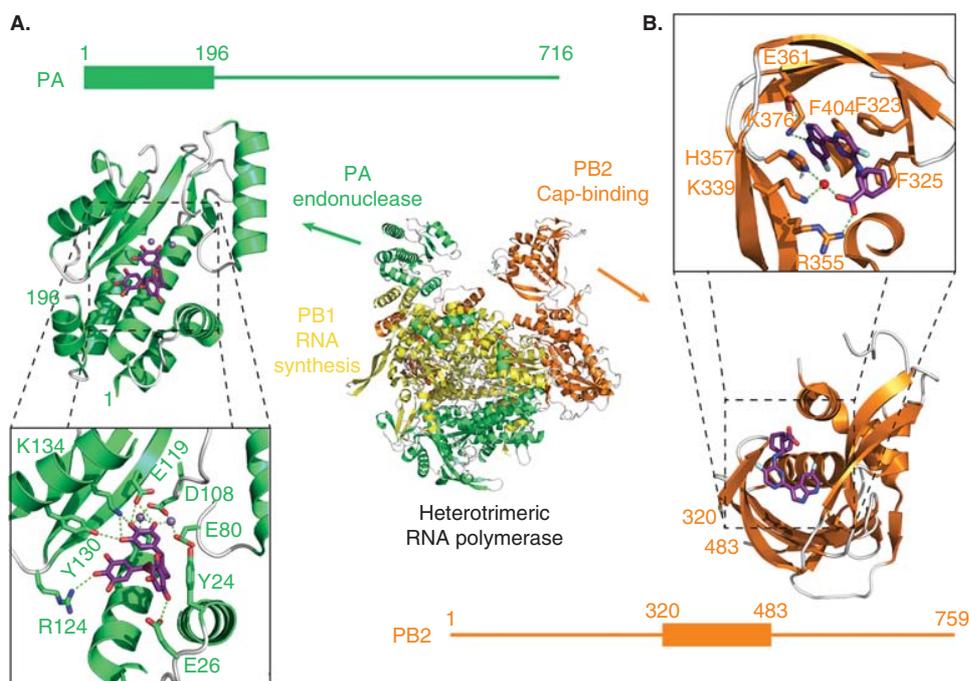


Figure 4. Illustration showing the binding of endonuclease and cap-binding inhibitors. The recent full heterotrimeric RNA-polymerase X-ray structure is shown as a reference ([68] PDB code: 4WSB), PA is represented in green, PB1 in yellow and PB2 in orange. (A) Bound green tea catechin to PA endonuclease active site is shown. Divalent cations (Mn^{2+} ions, purple spheres) and key active site residues that interact with the compound or are close to it are represented. Putative hydrogen bonds ($< 3.2 \text{ \AA}$) and additional possible interactions ($< 3.6 \text{ \AA}$) are shown as green dotted lines. Green inset: green tea catechin coordinates the two manganese ions and the active site residues are engaged in the interaction ([74] PDB code: 4AWM). (B) Bound VX787 to PB2 cap-binding domain is shown. Orange inset: the azaindole ring system is shown interacting with the key protein residues K376 and E361 (green dotted lines). The three aromatic rings are stacked between the side chains of H357, F404, F323 and F325. The carboxylic group of the ligand makes two water-mediated interactions (red spheres and green dotted lines) with the nitrogen of H357 and K339 ([71] PDB code: 4P1U).

was solved [61,62], two laboratories described high-resolution X-ray cocrystal structures of PA endonuclease domain with a series of specific compounds known or predicted to block the endonuclease activity [74,75]. This domain presents a high conservation among influenza A, B and C subtypes and identical residues are found in particular in the metal-binding residues His41, Glu80, Asp108, Glu119 and the catalytic residue Lys134. The cocrystal structure of Nter-PA with the well-known inhibitor 2,4 dioxo-4-phenylbutanoic acid (Table 2) [80] have been obtained by the two groups. Furthermore, whereas Dubois *et al.* showed crystal structures with different compounds like pyrimidine and hydroxyquinoline derivatives (Table 2), Kowalinski *et al.* characterized crystal structures with diketo compounds and a green tea catechin (Figure 4). All these inhibitors coordinate the two-metal endonuclease active site and engage the active site residues (Figure 4, green inset). The structures showed a structure-activity relationship of several known influenza inhibitors, opening the way for drug design optimization. In particular, residues away from the two-metal center may increase potency and new binding pockets bearing aromatic residues may deserve

further investigation. Additionally, the large water molecule network found in the structures may be targeted by inhibitors for an entropic gain while keeping the metal center recognition moiety. Recently, hydroxypyridine derivative has been identified as a bimetal chelating ligand at the endonuclease active site through an X-ray crystallographic screening campaign of fragments library [73]. Thus, two new series of endonuclease inhibitors hydroxypyridine and hydroxyquinoline derivatives have been reported [76,78] as well as their azaindole analogs [77].

3.4 Targeting PA-peptide PB1 interface

Efficient assembly of the influenza virus heterotrimeric RdRp is critical for virus replication and pathogenicity. Therefore, interfering with the assembly of the RdRp complex could be a way to fight against influenza infection. A large library of drugs, in particular benzofurazan derivatives, was found to act by inhibiting the viral RdRp complex through disruption of the complex formed by PA and PB1 subunits [81-84]. Since the association of these subunits is essential for replication [85-88] and the sequence of this domain is highly conserved [89,90],

this interaction represents an attractive target for inhibitors. Due to the crystallographic structures of the PA-peptide PB1 complex [63,64] a short peptide derived from the C terminus of PB1 has been shown to disrupt the interaction between the C-terminal part of PB1 and the N-terminal part of PB2 and to inhibit viral polymerase activity and replication [91]. Docking studies were performed to explore the structural features responsible for the biological activity of benzofurazan derivatives (Table 2) [83,84]. Docking suggested that their most reliable binding mode involves a π -stacking interactions with PA Trp706 of the benzofurazan scaffold, a hydrogen bond with Gln408 and electrostatic interactions with PA Lys643.

Last, most of the compounds targeting the RNP are still in a preclinical stage in terms of therapeutics for humans and many of drug-related aspects still require adequate studies. Likewise, no drug targeting NS1 as described below has been approved and the following are mostly preclinical studies.

4. Targeting the non-structural protein 1

NS1 is a highly conserved multifunctional protein. NS1 protects the virus against the antiviral responses of the host mediated by IFN- α/β pathways in a way that seems to be strain-specific [92]. NS1 antagonizes IFN by a number of ways, as binding to double-stranded (ds) viral RNA, thus protecting it from cellular factors, by blocking retinoic acid inducible gene-I and NF- κ B activation. NS1 is active as a dimer and is divided into two functional domains: an N-terminal domain that binds dsRNA (RNA-binding domain [RBD]) and a C-terminal effector domain that binds multiple cellular proteins. The structures of the isolated RBD alone and with RNA bound provided the structural basis for small molecule inhibition of the RNA-protein association or destabilization of the homodimer interface required for dsRNA binding [93]. In particular, a deep cavity was identified in the conserved tracks of basic and hydrophilic residues complementary to the phosphate backbone of A-form dsRNA of NS1 as a site for small molecule binding as a polyphenol and quinoxaline derivatives (Table 2) [94,95]. The recent structure of the C-terminal domain of NS1 in complex with a chromatin-associated factor may open a new way to inhibit NS1 through associated cellular factor-NS1 interface(s) [96].

5. Conclusion

Many of the antiviral candidates described in this review still require preclinical and clinical studies to assess their safety, toxicity and efficacy profiles required for the drug to be used on humans. The goal will be to treat in priority more susceptible populations as immunocompromised, young infants and elderly people in the search of antivirals directed against influenza A virus and probably adapt to their specific immune conditions. This goal is, nevertheless, complicated to reach,

owing to the difficulties of enrolling patients in appropriate clinical studies, complicated by comorbidities and placebo-controlled conditions which may be ethically unacceptable. FDA approval is thus challenging in these conditions. Future perspectives for new classes of antivirals targeting the RNP complex are very promising but these await pharmacokinetics and drug-related studies to be performed.

6. Expert opinion

6.1 Key findings and weaknesses in the research done so far

Monotherapy with antivirals holds a higher probability of resistance to treatment, in particular to those antivirals targeting external proteins with intrinsic relatively high mutation rate as NA, M2 and the variable region of HA.

To date, the antiviral candidates targeting NA are by far the most developed and for some of them extremely potent. The structure of the sialic acids derivatives in complex with the enzyme active site revealed an empty negatively charged pocket in the region of the C4 position on the sugar ring. The C4-OH was substituted by an amino or guanidinium group (oseltamivir, zanamivir, laninamivir) or by a bulky hydrophobic substituent (AV580 [19], A, B and C; Table 1). Substitution at other position(s) may help improve solubility and potency [23,24].

Novel antiviral candidates that target the RNP hold a great potential, both in terms of target conservation and of uniqueness/specificity compared to viral entry inhibitors. In terms of specificity, targeting the PB2 cap-binding domain of influenza virus may not differ so much from targeting cellular factors involved in cap-binding in the host cell. To date, inhibitors of the capping machinery are used to reverse chemoresistance since deregulation of cap-dependent translation is associated with cancer initiation [97]. Nevertheless, quite potent inhibitors of PB2 cap-binding were recently developed with few reported interactions with cellular partners [70,71]. The endonuclease domain represents a specific target with an excellent druggable site [73-78,83,84]. Moreover, the PA-PB1 interaction represents an attractive target for inhibitors since the association of PA and PB1 subunits is essential for replication [85-88], and the sequence of these domains is highly conserved [89,90]. The NP is also an excellent target for antivirals that was exploited through modulation of its self-association and its RNA-binding function [31,44,45,53,57,59,60,98].

6.2 Potential of this research and exciting area of research

The potential for disruption of specific protein-protein interactions within the RNP holds a lot of promises for the design of new antivirals. The field benefits from a breakthrough with the exciting crystal structures of influenza A and B polymerase with bound viral RNA promoter [68,69] that now opens brand new perspectives for the design of new antivirals perturbing

the polymerase assembly. Additionally, RNA-binding sites on the polymerase could represent newly defined targets probably through a better mechanistic understanding of the conserved RNA 3'-end recognition.

6.3 Antiviral combinations as a promising area of research

Multiple combinations as successfully exemplified by the tri-therapies developed against HIV may apply to severe influenza cases [1]. At present, combination therapies using dual NA antivirals did not reduce resistance. In contrast, pre-clinical studies showed additive effects of NA inhibitors with antiviral drugs with different mechanisms of action [1]. Future drug combinations targeting the viral RNP may involve compounds that inhibit more than one interaction type (protein-protein, protein-RNA). Alternatively, antiviral combination that target both a viral protein and a cellular factor involved in the host response to Influenza virus, holds promises for improved results for the host, human or animal [1,2]. The targeted host protein could be a factor required for transcription/replication or factors involved in chaperoning or transport of the RNP or its constituents. Moreover, combination therapies may reduce doses, thereby reducing the side effects and importantly, resistance against cellular factor are not observed.

Exploiting drug promiscuity (drug binding to more than one target) may be an advantage in severe cases of influenza [1]. A 'two-in-one' combination was recently suggested for naproxen, an approved anti-inflammatory drug and also inhibits influenza A NP [59].

6.4 How can it facilitate the discovery of new compounds or reduce attrition rates?

Improving our understanding of the structural biology of the virus and the underlying mechanisms of the targeted proteins in the viral lifecycle would be extremely important in the near

future. As for any drug development, overcoming toxicity is an issue requiring careful and relevant assays. Improving solubility and cellular targeting to the proper cellular compartment may decrease side effects.

6.5 Alternative technologies/approaches for understanding structural and mechanistic issues on influenza

Mechanistic assembly and function of multiprotein complexes may now benefit from high-resolution cryo-EM detectors. It is clear that function of the polymerase involves coordinated conformational changes and appropriate tools monitoring these movements in real time as time-resolved fluorescence, force measurements using single molecule techniques will flourish and greatly add to our understanding on influenza virus.

Declaration of interest

The authors are funded by the Agence Nationale de la Recherche, FLUNUCLEOVIR ANR 2010 Blan-130701 and RNAP-IAV ANR Blanc -14-CE09-0017. This work was granted access to the HPC resources of IDRIS made by Grand Equipment National de Calcul Intensif under the grant 2012-076378 attributed. This work used the platforms of the Grenoble Instruct Center (Integrated Structural Biology Grenoble; UMS3518 CNRS-CEA-UJF-EMBL) with support from The French Infrastructure for Integrated Structural Biology (ANR-10-INSB-05-02) and The Alliance Grenobloise pour la Biologie Structurale et Cellulaire Intégrées (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

Bibliography

Papers of special note have been highlighted as either of interest (●) or of considerable interest (●●) to readers.

- Dunning J, Baillie JK, Cao B, et al. Antiviral combinations for severe influenza. *Lancet Infect Dis* 2014;14(12):1259-70
- Du J, Cross TA, Zhou HX. Recent progress in structure-based anti-influenza drug design. *Drug Discov Today* 2012;17(19-20):1111-20
- Wang J, Wu Y, Ma C, et al. Structure and inhibition of the drug-resistant S31N mutant of the M2 ion channel of influenza A virus. *Proc Natl Acad Sci USA* 2013;110(4):1315-20
- Das K. Antivirals targeting influenza A virus. *J Med Chem* 2012;55:6263-77
- Das K, Aramini JM, Ma LC, et al. Structures of influenza A proteins and insights into antiviral drug targets. *Nat Struct Mol Biol* 2010;17(5):530-8
- Muller KH, Kakkola L, Nagaraj AS, et al. Emerging cellular targets for influenza antiviral agents. *Trends Pharmacol Sci* 2012;33(2):89-99
- Loregian A, Mercorelli B, Nannetti G, et al. Antiviral strategies against influenza virus: towards new therapeutic approaches. *Cell Mol Life Sci* 2014;71(19):3659-83
- Das K. **Structure of a potent inhibitor of matrix 2 (M2) of amantadine-resistant influenza.**
- Du J, Cross TA, Zhou HX. **Review of structural domains of the polymerase and non-structural protein 1 (NS1) as antiviral targets and their inhibitors.**
- Muller KH, Kakkola L, Nagaraj AS, et al. **Review of structural domains of the polymerase and NS1 as antiviral targets and their inhibitors.**
- Muller KH, Kakkola L, Nagaraj AS, et al. **Interesting overview of novel strategies targeting the virus and/or the host cell for counteracting influenza virus infection.**
- Dunning J, Baillie JK, Cao B, et al. **Nice review of structure-based antiviral candidates targeting influenza proteins.**
- Dunning J, Baillie JK, Cao B, et al. **Review of preclinical and clinical studies involving combination antiviral therapy for severe influenza.**

A. Monod et al.

8. Smee DF, Bailey KW, Wong MH, et al. Treatment of influenza A (H1N1) virus infections in mice and ferrets with cyanovirin-N. *Antiviral Res* 2008;80(3):266-71
9. Brancato V, Peduto A, Wharton S, et al. Design of inhibitors of influenza virus membrane fusion: synthesis, structure-activity relationship and in vitro antiviral activity of a novel indole series. *Antiviral Res* 2013;99(2):125-35
10. Leneva IA, Russell RJ, Boriskin YS, et al. Characteristics of arbidol-resistant mutants of influenza virus: implications for the mechanism of anti-influenza action of arbidol. *Antiviral Res* 2009;81(2):132-40
11. Russell RJ, Kerry PS, Stevens DJ, et al. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci USA* 2008;105(46):17736-41
12. Tang G, Lin X, Qiu Z, et al. Design and synthesis of benzenesulfonamide derivatives as potent anti-influenza hemagglutinin inhibitors. *ACS Med Chem Lett* 2011;2(8):603-7
13. Fleishman SJ, Whitehead TA, Ekiert DC, et al. Computational design of proteins targeting the conserved stem region of influenza hemagglutinin. *Science* 2011;332(6031):816-21
14. Basu A, Antanasijevic A, Wang M, et al. New small molecule entry inhibitors targeting hemagglutinin-mediated influenza A virus fusion. *J Virol* 2014;88(3):1447-60
15. Kim CU, Lew W, Williams MA, et al. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J Am Chem Soc* 1997;119(4):681-90
16. Russell RJ, Haire LF, Stevens DJ, et al. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 2006;443(7107):45-9
17. Spanakis N, Pitiriga V, Gennimata V, Taskris A. A review of neuraminidase inhibitor susceptibility in influenza strains. *Expert Rev Anti Infect Ther* 2014;12:1325-36
18. Vavricka CJ, Li Q, Wu Y, et al. Structural and functional analysis of laninamivir and its octanoate prodrug reveals group specific mechanisms for influenza NA inhibition. *PLoS Pathog* 2011;7(10):e1002249
19. Ivachtchenko AV, Ivanenkov YA, Mitkin OD, et al. Novel oral anti-influenza drug candidate AV5080. *J Antimicrob Chemother* 2014;69(7):1892-902
20. Vavricka CJ, Liu Y, Kiyota H, et al. Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. *Nat Commun* 2013;4:1491
21. Kim JH, Resende R, Wennekes T, et al. Mechanism-based covalent neuraminidase inhibitors with broad-spectrum influenza antiviral activity. *Science* 2013;340(6128):71-5
- **Novel sialic acid derivatives active against pandemic H1N1 and avian H5N1 influenza that target the active site.**
22. Rudrawar S, Dyason JC, Rameix-Welti MA, et al. Novel sialic acid derivatives lock open the 150-loop of an influenza A virus group-1 sialidase. *Nat Commun* 2010;1:113
- **Novel sialic acid derivatives active against pandemic H1N1 and avian H5N1 influenza that target the active site.**
23. Mohan S, Kerry PS, Bance N, et al. Serendipitous discovery of a potent influenza virus a neuraminidase inhibitor. *Angew Chem Int Ed Engl* 2014;53(4):1076-80
24. Kerry PS, Mohan S, Russell RJ, et al. Structural basis for a class of nanomolar influenza A neuraminidase inhibitors. *Sci Rep* 2013;3:2871
25. Lee CM, Weight AK, Haldar J, et al. Polymer-attached zanamivir inhibits synergistically both early and late stages of influenza virus infection. *Proc Natl Acad Sci USA* 2012;109(50):20385-90
26. Schnell JR, Chou JJ. Structure and mechanism of the M2 proton channel of influenza A virus. *Nature* 2008;451(7178):591-5
27. Cady SD, Schmidt-Rohr K, Wang J, et al. Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. *Nature* 2010;463(7281):689-92
28. Gu RX, Liu LA, Wei DQ. Structural and energetic analysis of drug inhibition of the influenza A M2 proton channel. *Trends Pharmacol Sci* 2013;34(10):571-80
29. Duque MD, Ma C, Torres E, et al. Exploring the size limit of templates for inhibitors of the M2 ion channel of influenza A virus. *J Med Chem* 2011;54(8):2646-57
30. Chenavas S, Estrozi LF, Delmas B, et al. Monomeric nucleoprotein of Influenza A virus. *Plos Pathogens* 2013;9(3):e1003275
- **Structure of monomeric nucleoprotein mutant R416A showing that the oligomerization loop folds in the RNA binding groove.**
31. Cianci C, Gerritz SW, Deminie C, et al. Influenza nucleoprotein: promising target for antiviral chemotherapy. *Antivir Chem Chemother* 2012;23(3):77-91
32. Davis AM, Chabolla BJ, Newcomb LL. Emerging antiviral resistant strains of influenza A and the potential therapeutic targets within the viral ribonucleoprotein (vRNP) complex. *Virology* 2014;11:167
33. Arranz R, Coloma R, Chichon FJ, et al. The structure of native influenza virion ribonucleoproteins. *Science* 2012;338(6114):1634-7
- **Cryo-microscopic studies of the ribonucleoprotein (RNP) and discovery of the nucleoprotein (NP)-RNA double-stranded structure.**
34. Moeller A, Kirchdoerfer RN, Potter CS, et al. Organization of the influenza virus replication machinery. *Science* 2012;338(6114):1631-4
- **Cryo-microscopic studies of the RNP and an electron microscopic (EM) model of the native influenza RNP.**
35. Coloma R, Valpuesta JM, Arranz R, et al. The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog* 2009;5(6):e1000491
- **High-resolution cryo-EM structure of a recombinant RNP containing an RNA molecule of 248 nucleotides, 9 NP subunits and a polymerase complex. The atomic structures of NP are also given.**
36. Ng AK, Chan WH, Choi ST, et al. Influenza Polymerase Activity Correlates with the Strength of Interaction between Nucleoprotein and PB2 through the Host-Specific Residue K/E627. *PLoS ONE* 2012;7(5):e36415

37. Marklund JK, Ye Q, Dong J, et al. Sequence in the influenza A virus nucleoprotein required for viral polymerase binding and RNA synthesis. *J Virol* 2012;86(13):7292-7
38. Gui X, Li R, Zhang X, et al. An important amino acid in nucleoprotein contributes to influenza A virus replication by interacting with polymerase PB2. *Virology* 2014;464-465:11-20
39. Gavazzi C, Yver M, Isel C, et al. A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc Natl Acad Sci USA* 2013;110(41):16604-9
40. Turrell L, Lyall JW, Tiley LS, et al. The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nat Commun* 2013;4:1591
41. Portela A, Digard P. The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *J Gen Virol* 2002;83(Pt 4):723-34
42. Newcomb LL, Kuo RL, Ye Q, et al. Interaction of the influenza A virus nucleocapsid protein with the viral RNA polymerase potentiates unprimed viral RNA replication. *J Virol* 2009;83(1):29-36
43. Ruigrok RW, Baudin F. Structure of influenza virus ribonucleoprotein particles. II. Purified RNA-free influenza virus ribonucleoprotein forms structures that are indistinguishable from the intact influenza virus ribonucleoprotein particles. *J Gen Virol* 1995;76(Pt 4):1009-14
44. Ng AK, Zhang H, Tan K, et al. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J* 2008;22(10):3638-47
- **Atomic structure of H5N1 nucleoprotein and mutational analysis of the RNA-binding site.**
45. Ye Q, Krug RM, Tao YJ. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* 2006;444(7122):1078-82
- **Atomic structure of H1N1 influenza virus NP showing a domain exchange polymerization domain and a putative RNA-binding surface.**
46. Ye Q, Guu TS, Mata DA, et al. Biochemical and structural evidence in support of a coherent model for the formation of the double-helical influenza A virus ribonucleoprotein. *MBio* 2012;4(1):e00467-12
47. Ng AK, Lam MK, Zhang H, et al. Structural basis for RNA binding and homo-oligomer formation by influenza B virus nucleoprotein. *J Virol* 2012;86(12):6758-67
48. Ruigrok RW, Crepin T, Kolakofsky D. Nucleoproteins and nucleocapsids of negative-strand RNA viruses. *Curr Opin Microbiol* 2011;14(4):504-10
49. Reguera J, Cusack S, Kolakofsky D. Segmented negative strand RNA virus nucleoprotein structure. *Curr Opin Virol* 2014;5:7-15
50. Reguera J, Malet H, Weber F, et al. Structural basis for encapsidation of genomic RNA by La Crosse Orthobunyavirus nucleoprotein. *Proc Natl Acad Sci USA* 2013;110(18):7246-51
51. Tarus B, Bakowicz O, Chenavas S, et al. Oligomerization paths of the nucleoprotein of influenza A virus. *Biochimie* 2012;94:776-85
52. Tarus B, Chevalier C, Richard CA, et al. Molecular dynamics studies of the nucleoprotein of influenza A virus: role of the protein flexibility in RNA binding. *PLoS ONE* 2012;7(1):e30038
53. Boulo SA, Lotteu H, Mueller V, et al. Human importin alpha and RNA do not compete for binding to Influenza A nucleoprotein. *Virology* 2011;409(1):84-90
54. Amorim MJ, Kao RY, Digard P. Nucleozin targets cytoplasmic trafficking of viral ribonucleoprotein-Rab11 complexes in influenza A virus infection. *J Virol* 2013;87(8):4694-703
55. Kao RY, Yang D, Lau LS, et al. Identification of influenza A nucleoprotein as an antiviral target. *Nat Biotechnol* 2010;28(6):600-5
- **Identification of nucleozin by high-throughput screening, first inhibitor of NP, enhancing non-functional oligomerization of NP and impeding its transport to the nucleus.**
56. Cheng H, Wan J, Lin MI, et al. Design, synthesis, and in vitro biological evaluation of 1H-1,2,3-triazole-4-carboxamide derivatives as new anti-influenza A agents targeting virus nucleoprotein. *J Med Chem* 2012;55(5):2144-53
57. Gerritz SW, Cianci C, Kim S, et al. Inhibition of influenza virus replication via small molecules that induce the formation of higher-order nucleoprotein oligomers. *Proc Natl Acad Sci USA* 2011;108(37):15366-71
- **Structure of a nucleozin derivative stabilizing the interface between two NP subunits.**
58. Shen YF, Chen YH, Chu SY, et al. E339 R416 salt bridge of nucleoprotein as a feasible target for influenza virus inhibitors. *Proc Natl Acad Sci USA* 2011;108(40):16515-20
- **Targeting the oligomerization loop of NP by new compounds to decrease its oligomerization.**
59. Lejal N, Tarus B, Bouguyon E, et al. Structure-Based Discovery of the Novel Antiviral Properties of Naproxen against the Nucleoprotein of Influenza A Virus. *Antimicrob Agents Chemother* 2013;57(5):2231-42
- **Competing with RNA binding to the NP by naproxen, a known inhibitor of COX2.**
60. Tarus B, Bertrand H, Zedda G, et al. Structure-based design of novel naproxen derivatives targeting monomeric nucleoprotein of Influenza A virus. *J Biomol Struct Dyn* 2014; [Epub ahead of print]
61. Dias A, Bouvier D, Crepin T, et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 2009;458(7240):914-18
- **Atomic structure of the endonuclease domain of the PA subunit containing two bound manganese ions and biochemical demonstration that the isolated domain has intrinsic divalent cation-dependent nuclease activity.**
62. Yuan P, Bartlam M, Lou Z, et al. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* 2009;458(7240):909-13
- **Atomic structure of the endonuclease domain in the PA subunit containing a single-bound magnesium ion. The biological function of this domain was confirmed by a mutational analysis using recombinant polymerase showing that mutants of the metal-ligand**

- residues were active in replication but inactive in cap-dependent transcription.
63. He X, Zhou J, Bartlam M, et al. Crystal structure of the polymerase PA(C)-PB1 (N) complex from an avian influenza H5N1 virus. *Nature* 2008;454(7208):1123-6
- **Atomic structure of the C-terminal domain of the PA subunit in complex with a short N-terminal peptide from the PB1 subunit.**
64. Obayashi E, Yoshida H, Kawai F, et al. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* 2008;454(7208):1127-31
- **Atomic structure of the C-terminal domain of the PA subunit in complex with a short N-terminal peptide from the PB1 subunit.**
65. Sugiyama K, Obayashi E, Kawaguchi A, et al. Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. *EMBO J* 2009;28(12):1803-11
66. Guilligay D, Tarendeau F, Resa-Infante P, et al. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* 2008;15(5):500-6
- **Atomic structure of the PB2 cap-binding domain in complex with 7-methylguanosine triphosphate as well as binding studies and analysis using recombinant polymerase showing that mutations of residues involved in cap-binding leads to the inactivation of cap-dependent transcription but not cap-independent transcription or replication. The paper also describes the ESPRIT method for screening soluble internal fragments of proteins.**
67. Tarendeau F, Crepin T, Guilligay D, et al. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* 2008;4(8):e1000136
- **Atomic structure of the C-terminal double domain of the PB2 subunit rich in amino acids that are mutated when avian viruses adapt to mammals. The most prominent of these is residue 627 that is a lysine in viruses that infect the human respiratory tract. The human virus domain contains an extensive positively charged surface that is disrupted by a glutamate substitution.**
68. Pflug A, Guigay D, Reich S, et al. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* 2014;516(7531):355-60.
- **Atomic structure of the C-terminal double domain of the PB2 subunit rich in amino acids that are mutated when avian viruses adapt to mammals. The most prominent of these is residue 627 that is a lysine in viruses that infect the human respiratory tract. The human virus domain contains an extensive positively charged surface that is disrupted by a glutamate substitution.**
69. Reich S, Guilligay D, Malet H, et al. Structural insight into cap-snatching and RNA synthesis by influenza polymerase. *Nature* 2014;516(7531):361-6.
- **Structures of influenza B polymerase and evidence of movements of PB2 cap-binding domains. Conformational changes required to initiate polymerase function are suggested.**
70. Pautus S, Sehr P, Lewis J, et al. New 7-methylguanine derivatives targeting the influenza polymerase PB2 cap-binding domain. *J Med Chem* 2013;56(21):8915-30
71. Clark MP, Ledebor MW, Davies I, et al. Discovery of a novel, first-in-class, orally bioavailable azaindole inhibitor (VX-787) of influenza PB2. *J Med Chem* 2014;57(15):6668-78
72. Wathen MW, Barro M, Bright RA. Antiviral in seasonal and pandemic influenza-future perspectives. *Influenza Other Respir Viruses* 2012;7(Suppl 1):76-80
73. Bauman JD, Patel D, Baker SF, et al. Crystallographic fragment screening and structure-based optimization yields a new class of influenza endonuclease inhibitors. *ACS Chem Biol* 2013;8:2501-8
74. Kowalinski E, Zubieta C, Wolkerstorfer A, et al. Structural analysis of specific metal chelating inhibitor binding to the endonuclease domain of influenza pH1N1 (2009) polymerase. *PLoS Pathog* 2012;8(8):e1002831
- **Structure of PA endonuclease domain with a series of specific inhibitors, including four diketo compounds and a green tea catechin.**
75. DuBois RM, Slavish PJ, Baughman BM, et al. Structural and biochemical basis for development of influenza virus inhibitors targeting the PA endonuclease. *PLoS Pathog* 2012;8(8):e1002830
- **Inhibitors of PA endonuclease.**
76. Parhi AK, Xiang A, Bauman JD, et al. Phenyl substituted 3-hydroxypyridin-2 (1H)-ones: inhibitors of influenza A endonuclease. *Bioorg Med Chem* 2013;21(21):6435-46
77. Sagong HY, Bauman JD, Patel D, et al. Phenyl Substituted 4-Hydroxypyridazin-3 (2H)-ones and 5-Hydroxypyrimidin-4 (3H)-ones: inhibitors of Influenza A Endonuclease. *J Med Chem* 2014;57(19):8086-98
78. Sagong HY, Parhi A, Bauman JD, et al. 3-Hydroxyquinolin-2(1H)-ones As Inhibitors of Influenza A Endonuclease. *ACS Med Chem Lett* 2013;4(6):547-50
79. Tefsen B, Lu G, Zhu Y, et al. The N-terminal domain of PA from bat-derived influenza-like virus H17N10 has endonuclease activity. *J Virol* 2014;88(4):1935-41
80. Tomassini J, Selnick H, Davies ME, et al. Inhibition of cap (m7GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob Agents Chemother* 1994;38(12):2827-37
81. Muratore G, Mercorelli B, Goracci L, et al. Human cytomegalovirus inhibitor AL18 also possesses activity against influenza A and B viruses. *Antimicrob Agents Chemother* 2012;56(11):6009-13
82. Muratore G, Goracci L, Mercorelli B, et al. Small molecule inhibitors of influenza A and B viruses that act by disrupting subunit interactions of the viral polymerase. *Proc Natl Acad Sci USA* 2012;109(16):6247-52
83. Pagano M, Castagnolo D, Bernardini M, et al. The fight against the influenza A virus H1N1: synthesis, molecular modeling, and biological evaluation of benzofurazan derivatives as viral RNA polymerase inhibitors. *ChemMedChem* 2014;9(1):129-50
84. Tintori C, Laurenzana I, Fallacara AL, et al. High-throughput docking for the identification of new influenza A virus polymerase inhibitors targeting the PA-PB1 protein-protein interaction.

- Bioorg Med Chem Lett 2014;24(1):280-2
- **Novel inhibitors targeting the PA–PB1 protein–protein interaction.**
85. Brownlee GG, Sharps JL. The RNA polymerase of influenza A virus is stabilized by interaction with its viral RNA promoter. *J Virol* 2002;76(14):7103-13
 86. Fodor E, Crow M, Mingay LJ, et al. A single amino acid mutation in the PA subunit of the influenza virus RNA polymerase inhibits endonucleolytic cleavage of capped RNAs. *J Virol* 2002;76(18):8989-9001
 87. Perales B, Ortin J. The influenza A virus PB2 polymerase subunit is required for the replication of viral RNA. *J Virol* 1997;71(2):1381-5
 88. Perez DR, Donis RO. Functional analysis of PA binding by influenza A virus PB1: effects on polymerase activity and viral infectivity. *J Virol* 2001;75(17):8127-36
 89. Ghanem A, Mayer D, Chase G, et al. Peptide-mediated interference with influenza A virus polymerase. *J Virol* 2007;81(14):7801-4
 90. Wunderlich K, Mayer D, Ranadheera C, et al. Identification of a PA-binding peptide with inhibitory activity against influenza A and B virus replication. *PLoS ONE* 2009;4(10):e7517
 91. Li C, Ba Q, Wu A, et al. A peptide derived from the C-terminus of PB1 inhibits influenza virus replication by interfering with viral polymerase assembly. *FEBS J* 2013;280(4):1139-49
 92. Engel DA. The influenza virus NS1 protein as a therapeutic target. *Antiviral Res* 2013;99(3):409-16
 93. Yin C, Khan JA, Swapna GV, et al. Conserved surface features form the double-stranded RNA binding site of non-structural protein 1 (NS1) from influenza A and B viruses. *J Biol Chem* 2007;282(28):20584-92
 94. Cho EJ, Xia S, Ma LC, et al. Identification of influenza virus inhibitors targeting NS1A utilizing fluorescence polarization-based high-throughput assay. *J Biomol Screen* 2012;17(4):448-59
 95. You L, Cho EJ, Leavitt J, et al. Synthesis and evaluation of quinoxaline derivatives as potential influenza NS1A protein inhibitors. *Bioorg Med Chem Lett* 2011;21(10):3007-11
 96. Qin S, Liu Y, Tempel W, et al. Structural basis for histone mimicry and hijacking of host proteins by influenza virus protein NS1. *Nat Commun* 2014;5:3952
 97. Cencic R, Hall DR, Robert F, et al. Reversing chemoresistance by small molecule inhibition of the translation initiation complex eIF4F. *Proc Natl Acad Sci USA* 2011;108(3):1046-51
 98. Chenavas S, Crepin T, Delmas B, et al. Influenza virus nucleoprotein: structure, RNA binding, oligomerization and antiviral drug target. *Future Microbiol* 2013;8:1537-45
 - **Structure of monomeric nucleoprotein (NP)-mutant R416A showing that the oligomerization loop folds in the RNA-binding groove.**

Affiliation

Alexandre Monod¹, Christopher Swale¹, Bogdan Tarus², Alice Tissot¹, Bernard Delmas², Rob WH Ruigrok¹, Thibaut Crépin¹ & Anny Slama-Schwok^{†2}

[†]Author for correspondence

¹University of Grenoble Alpes-EMBL-CNRS, Unit for Virus Host-Cell Interactions, 71 avenue des Martyrs, 38042 Grenoble, France

²Virologie et Immunologie Moléculaires, INRA UR 892, 78350, Jouy en Josas, France

E-mail: Anny.Schwok@gmail.com

Publication n°3:

Thibaut Crépin, Alexandre Dias, Andrés Palencia, Christopher Swale, Stephen Cusack, And Rob W. H. Ruigrok

Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit

Journal of Virology, Sept. 2010, p. 9096–9104

Mutational and Metal Binding Analysis of the Endonuclease Domain of the Influenza Virus Polymerase PA Subunit[∇]

Thibaut Crépin,¹ Alexandre Dias,¹ Andrés Palencia,^{1,2} Christopher Swale,¹
Stephen Cusack,^{1,2} and Rob W. H. Ruigrok^{1*}

UJF-EMBL-CNRS, UMI 3265, Unit of Virus Host-Cell Interactions, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France,¹
and European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France²

Received 7 May 2010/Accepted 23 June 2010

Influenza virus polymerase initiates the biosynthesis of its own mRNAs with capped 10- to 13-nucleotide fragments cleaved from cellular (pre-)mRNAs. Two activities are required for this cap-snatching activity: specific binding of the cap structure and an endonuclease activity. Recent work has shown that the cap-binding site is situated in the central part of the PB2 subunit and that the endonuclease activity is situated in the N-terminal domain of the PA subunit (PA-Nter). The influenza endonuclease is a member of the PD-(D/E)XK family of nucleases that use divalent metal ions for nucleic acid cleavage. Here we analyze the metal binding and endonuclease activities of eight PA-Nter single-point mutants. We show by calorimetry that the wild-type active site binds two Mn²⁺ ions and has a 500-fold higher affinity for manganese than for magnesium ions. The endonuclease activity of the isolated mutant domains are compared with the cap-dependent transcription activities of identical mutations in trimeric recombinant polymerases previously described by other groups. Mutations that inactivate the endonuclease activity in the isolated PA-Nter knock out the transcription but not replication activity in the recombinant polymerase. We confirm the importance of a number of active-site residues and identify some residues that may be involved in the positioning of the RNA substrate in the active site. Our results validate the use of the isolated endonuclease domain in a drug-design process for new anti-influenza virus compounds.

Influenza virus is a segmented negative-strand RNA virus that replicates in the nucleus of infected cells. Its eight viral RNA (vRNA) segments are covered by the viral nucleoprotein (NP) with a stoichiometry of 24 nucleotides per nucleoprotein protomer (25). The 3' and 5' ends of the vRNA are bound to the RNA-dependent RNA polymerase, a heterotrimeric complex composed of PB1, PB2, and PA. The complex between vRNA, NP, and the polymerase is called a ribonucleoprotein particle (RNP). Both ends of the viral RNA are necessary for polymerase activity, and together they form the promoter for transcription (12, 18, 33). Influenza virus RNA polymerase differs from the polymerases of the nonsegmented negative-strand RNA viruses, in that it does not carry the enzymatic functions required for the 5' capping of its own mRNAs (guanylyl- and methyltransferase activities). Instead, influenza virus mRNAs are capped by a unique cap-snatching mechanism (27). The polymerase binds to cellular pre-mRNAs via their cap structure and then cleaves them to generate capped 10- to 13-residue oligonucleotides, which are used as primers to initiate the transcription of viral mRNAs. The viral mRNAs are terminated by a 3'-end poly(A) sequence generated by the stuttering of the RNA polymerase at an oligo(U) motif located at the 5' end of the template (28, 29). The endonuclease activity for cleavage of the host mRNAs is not active in recombinant trimeric polymerase in the absence of vRNA. For viral transcription, the binding of the 5' end of the genomic vRNA

is necessary to stimulate the cap binding (4). The binding of both ends of the vRNA is also required to stimulate the endonuclease activity (12), and direct binding of annealed 3' and 5' vRNA ends stimulates cap binding and endonuclease activity even more strongly (22). These results suggest that significant conformational changes may take place in the RNA polymerase complex upon binding of the vRNA. When the three-dimensional (3D) electron microscopy model of the polymerase on a recombinant RNP was compared with that of free recombinant polymerase, structural differences were indeed observed (1, 5, 35).

The cap-binding site was known to be located within the PB2 subunit (23) but was only recently shown by structural analysis to reside in an independently folding domain of the PB2 subunit, between amino acids 320 and 483 (10). This isolated domain binds to cap analogues in the absence of other parts of the RNA polymerase and of vRNA. The position of the endonuclease site was also controversial but has recently been proven to reside in an also independently folding amino-terminal domain of the PA subunit (6, 37). The crystal structure of the first 196 residues of PA (PA-Nter) (Fig. 1A) shows structural homology with nucleases of the PD-(D/E)XK superfamily, which contains bacterial restriction enzymes such as *Escherichia coli* EcoRV and *Pyrococcus furiosus* Holiday junction resolvase. This family of enzymes binds to one or two divalent metal ions, in particular, Mg²⁺ or Mn²⁺. One of the ions is clearly involved in catalysis, whereas the role of the second one is not clearly established and may have a modulatory role (inhibition or stimulation, depending on the concentration and the nature of the ion) (17, 26). PA-Nter shows endonuclease activity in the absence of the rest of the poly-

* Corresponding author. Mailing address: UJF-EMBL-CNRS, UMI 3265, Unit of Virus Host-Cell Interactions, 6 rue Jules Horowitz, BP 181, 38042 Grenoble Cedex 9, France. Phone: 33 4 76 20 72 73. Fax: 33 4 76 29 71 99. E-mail: ruigrok@embl.fr.

[∇] Published ahead of print on 30 June 2010.

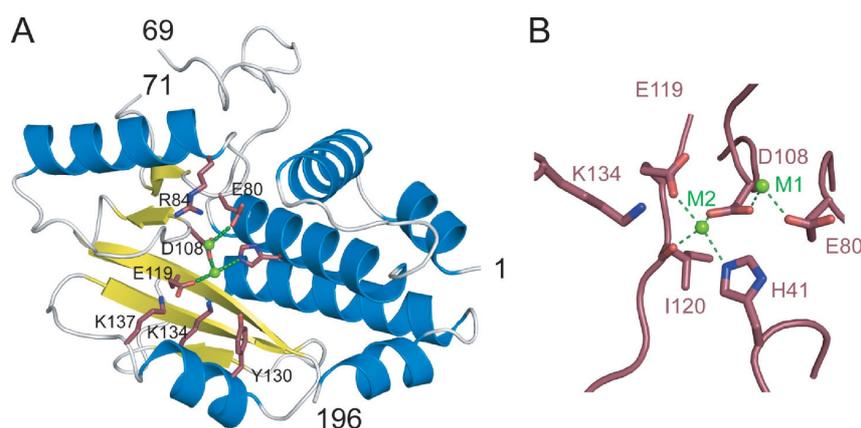


FIG. 1. Active site of PA-Nter. (A) Ribbon diagram of the structure of influenza A/Victoria/3/1975 PA-Nter (Protein Data Bank accession number 2W69) with α helices in blue and β strands in yellow. The key active-site residues mutated in this study are indicated in pink, and the manganese ions are in green. (B) Blowup of the active site of PA-Nter indicating the two metal binding sites and the metal binding ligands.

merase and has the same metal ion dependence as the polymerase in intact RNPs. The domain cleaves most efficiently with manganese ions, followed by cobalt ions (6, 7). In the crystal structure of PA-Nter determined by Dias et al., two divalent cation sites were identified, and both are occupied by manganese (6); metal site 1 (M1) is liganded by Glu80, Asp108, and Glu119 (through a water molecule); and metal site 2 (M2) is liganded by His41, Asp108, Glu119, and the carbonyl oxygen of Ile120 (Fig. 1A and B). On the other hand, Yuan and coworkers crystallized PA-Nter in the presence of only $MgCl_2$ and observed a single Mg^{2+} ion in site M1 (37). The active site also contains Lys134, which could correspond to the catalytic lysine of the PD-(D/E)XK motif (6, 37, 38).

Both the isolated cap-binding domain of PB2 and the endonuclease domain of PA have significant advantages compared to intact viral RNPs for inhibitor screening and structure-based antiviral drug optimization. However, it needs to be ascertained that the isolated domains have the same function and activity as the domains inside the intact trimeric RNA polymerase. A mutational analysis of the cap-binding domain confirmed that mutations in the isolated domain have the same effects as those in the context of the intact polymerase (10). Here we present the results of a mutational analysis of the active site of the isolated endonuclease domain. The mutations include the active-site residues involved in metal binding (His41, Glu80, Asp108, Glu119), the putative catalytic Lys134 residue, and three strictly conserved residues located on the rim surrounding the active site: Arg84, Tyr130, and Lys137 (Fig. 1A). The effects of these mutations introduced in the isolated domain on endonuclease activity are compared to those already tested in the intact recombinant polymerase (13, 37). We also tested the affinities of Mn^{2+} and Mg^{2+} ions for the wild-type (wt) and mutant endonuclease domains and correlate metal binding with enzymatic activity.

MATERIALS AND METHODS

Mutants. The clones containing the mutated DNA coding for PA-Nter (residues 1 to 209), cloned in pETM11, were obtained from Geneart (Germany). The clones code for a poly(His) sequence separated from the protein-coding sequence by a TEV cleavage site. The vectors were used to transform the *E. coli* BL21(DE3) RIL CodonPlus strain (Stratagene). The mutant proteins were ex-

pressed in LB medium overnight at 15°C after induction with 0.1 mM isopropyl- β -D-thiogalactopyranoside (IPTG). All PA-Nter mutants were expressed in the soluble fraction of the bacterial cells and were purified by an immobilized metal affinity column (IMAC). A second IMAC step was performed after cleavage by the His-tagged TEV protease, followed by gel filtration on a Superdex 200 column (GE Healthcare). Finally, the protein was concentrated to 5 to 10 $mg \cdot ml^{-1}$. For H41A and H41E, the yields were only about 5% of those obtained for the other mutants. The elution profile during the final gel filtration step showed the presence of aggregated material for these mutants.

Biophysical characterization. (i) **Far-UV CD spectra.** Far-UV circular dichroism (CD) spectra were recorded with a 1-mm path length at 20°C on a Jasco model J-810 CD spectropolarimeter equipped with a Peltier thermostat, as described previously (6). The PA-Nter concentration was 10 μM in 10 mM Tris-HCl, pH 7.0–10 mM NaCl.

(ii) **Thermal shift assays.** Thermal shift assays were performed with 10 μM PA-Nter in 20 mM Tris-HCl, pH 7.0–100 mM NaCl and a 5 \times dilution of SYPRO orange dye (Invitrogen), as described. The dye was excited at 490 nm, and the emission light was recorded at 575 nm while the temperature was increased by increments of 1°C per minute from 25 to 75°C. Control assays were carried out in the absence of protein or dye to check that no fluorescence signal was recorded. The experiments gave virtually identical results when they were performed under the same conditions. The variation between experiments came only from the estimation of the flexion point of the curve and was less than 0.5°C in triplicate experiments. For this reason, the results shown in Fig. 2C do not have error bars.

(iii) **Isothermal titration calorimetry.** Isothermal titration calorimetry (ITC) experiments were performed using a high-precision VP-ITC system (Microcal Inc., Northampton, MA). Proteins were first extensively dialyzed against the titration buffer (20 mM Tris-HCl, pH 8.0, 100 mM NaCl). All solutions were filtered, degassed to avoid bubble formation, and equilibrated to the corresponding temperature before each experiment. Protein solutions at about 60 μM in the calorimetric cell were titrated with the appropriate metal (0.8 to 9 mM) dissolved in dialysis buffer. Depending on the binding affinities, titrations were carried out either by constant-volume injections (30 injections of 7.5 μl) or by increasing-volume injections (27 injections from 4 to 20 μl) in order to better define the titration curves. The heat evolved after each metal injection was obtained from the integral of the calorimetric signal. The resulting binding isotherms were analyzed by nonlinear least-squares fitting of the experimental data to models corresponding to a single set of identical sites or corresponding to two sets of independent sites. Analysis of the data was performed using the Microcal Origin program (OriginLab Corporation, Northampton, MA).

Endonuclease assays. Endonuclease assays were carried out using an unstructured U-rich RNA probe of 51 nucleotides (6, 31) or a short panhandle RNA of 36 nucleotides comprising just the conserved 3' and 5' ends of the viral RNA with a short linker (6). RNA cleavage was performed by incubating 13 μM PA-Nter with various RNA substrates (all at 10 μM) at 37°C in a final volume of 50 μl . The reaction buffer was 20 mM Tris-HCl, pH 8, 100 mM NaCl, and 10 mM β -mercaptoethanol plus 1 mM $MnCl_2$ or 1 mM $MgCl_2$. Incubations were stopped by addition of EGTA at a final concentration of 20 mM. The reaction products

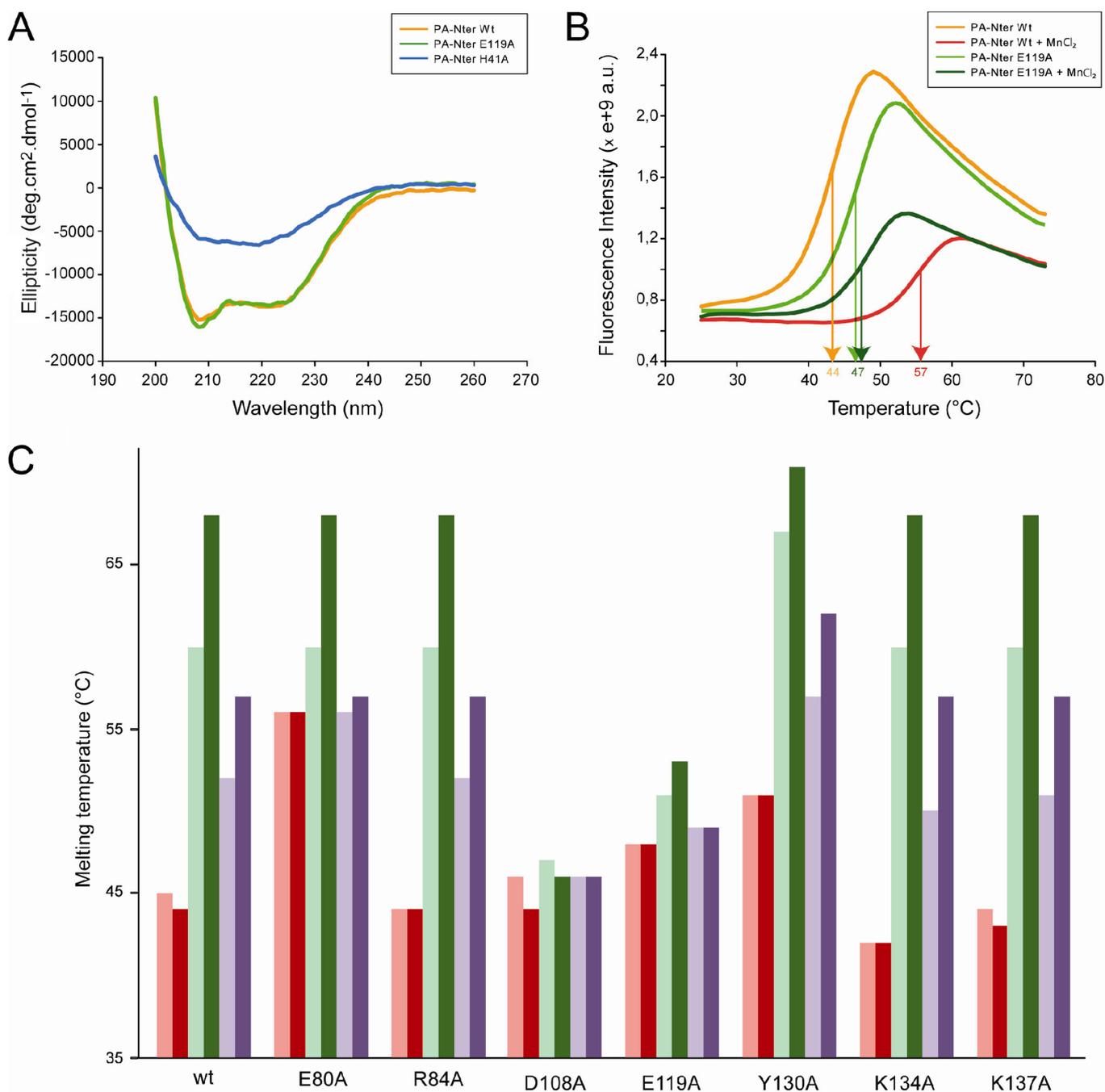


FIG. 2. Biophysical characterization of PA-Nter mutants. (A) Circular dichroism on wt PA-Nter and two of the eight mutants. CD spectra of the wt, H41A, and E119A PA-Nters are in orange, blue, and green, respectively. (B) Thermal stabilization by MnCl₂. Thermal shift assays on wt and E119A PA-Nters were done in the presence (red and dark green, respectively) or absence (orange and green, respectively) of 0.5 mM MnCl₂. (C) Effects of metals and DPBA binding on the thermal stability. Thermal shift assays to test the metal ion stabilization were performed on wt PA-Nter and the corresponding mutants in the presence (dark colors) and absence (light colors) of DPBA. The proteins were incubated with 0.1 mM MnCl₂ (green) or 5 mM MgCl₂ (violet) or without any metal (red). All experiments were repeated at least three times on different occasions and with different protein preparations. The results were identical when the same experimental conditions were used.

were loaded on 8 M urea–15% polyacrylamide gels and stained with methylene blue. The results of these assays were not quantified. The activities were estimated from the disappearance of the band of the intact panhandle RNA after incubation for 80 min and 6 h as ++ for wt enzyme, + for mutants like R84A and Y130A that digested less than 50% of the substrate RNA after 80 min but all of it after 6 h, ± for mutants like K137A that still had a significant amount of intact RNA after 6 h, and – for mutants like D108A, E119A, and K134A that were inactive.

RESULTS

All PA-Nter mutants were expressed and purified, and all mutants except the H41A mutant had yields comparable to the yield of the wt. This mutant yielded only about 5% of that for the wild type, and the elution profile of the final gel filtration step showed the presence of aggregates. The folding of the

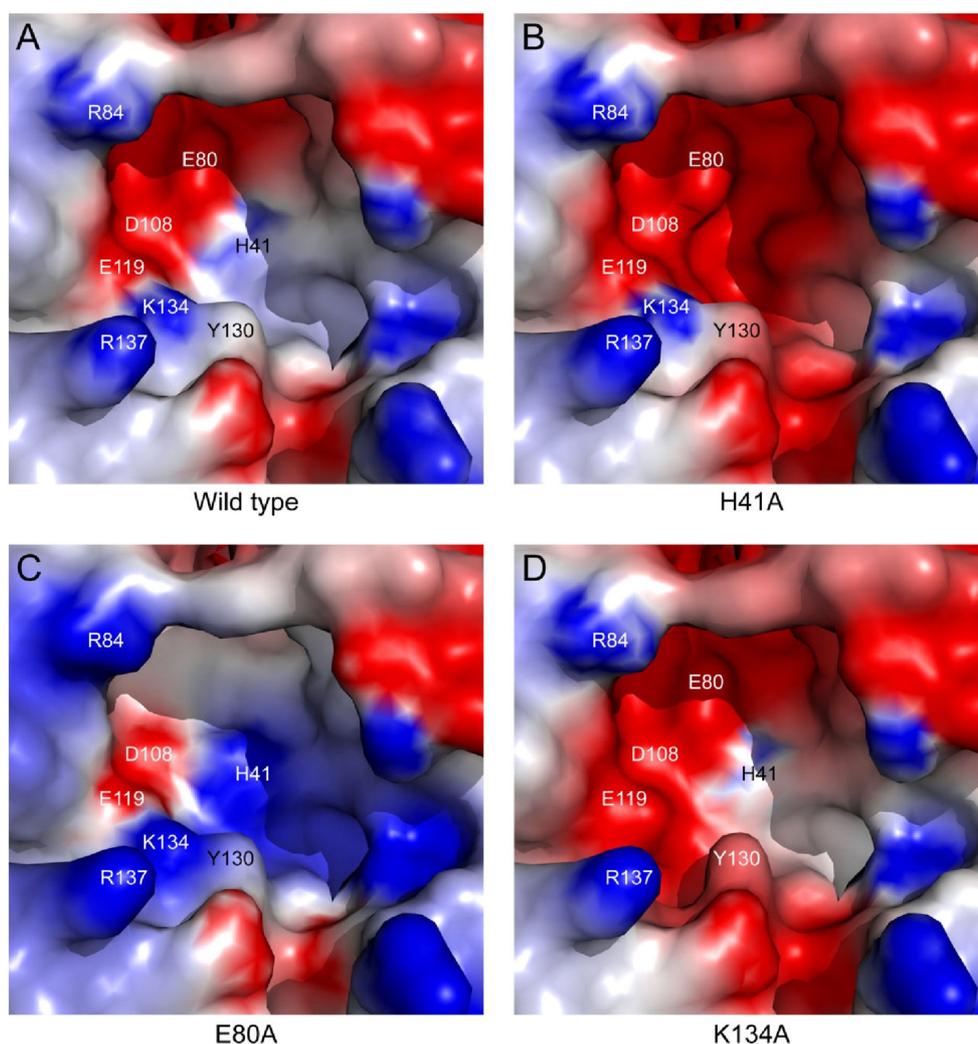


FIG. 3. Electrostatic surface potentials of the active sites of wt and three mutant PA-Nters. The surface of wt PA-Nter (A) was calculated from the crystal structure (Protein Data Bank accession number 2W69), whereas those of mutant PA-nters H41A (B), E80A (C), and K134A (D) correspond to models generated *in silico*. The electrostatic surfaces were calculated using the DelPhi program (30) with H41 fully protonated. The potential scales range from -5.0 kT/e (red) to 5.0 kT/e (blue).

mutants was checked by circular dichroism. All mutants showed a spectrum identical to that of the wt PA-Nter, as shown in Fig. 2A only for the E119A mutant, except that the H41A mutant showed little secondary structure (Fig. 2A).

Thermal stability. The thermal stability of the mutants was tested by Thermofluor assays in which a hydrophobic fluorophore has little affinity for native proteins but binds to denatured proteins, leading to an increase of the fluorescence (9). The apparent melting temperature (T_m) of denaturation can be obtained from the temperature dependence of the fluorescence (Fig. 2B). The thermal stability of the H41A mutant could not be derived since the fluorescent probe bound to the protein at room temperature, another indication that it is not properly folded. The R84A, D108A, and K137A PA-Nters have the same T_m as wt PA-Nter (Fig. 2C, pink bars) and the K134A PA-Nter is slightly less stable, whereas the E80A, E119A, and Y130A PA-Nters are more stable (Fig. 2C). The active site of wt PA-Nter is strongly negatively charged (Fig. 3A). In general, mutations that increase the negative charge in

the active-site pocket destabilize the domain (e.g., see Fig. 3B and D for the H41A and K134A PA-Nters, respectively), whereas mutations that reduce the negative charge increase the stability of the domain (e.g., Fig. 3C for the E80A PA-Nter). The fact that the H41A PA-Nter does not fold properly suggests that a basic residue at position 41 is important to provide the necessary electrostatic compensation for correct folding.

Nuclease activity. The RNase activity of the mutants was tested in the presence of 1 mM $MnCl_2$ or 1 mM $MgCl_2$ using short panhandle RNA (36 nucleotides) and unstructured U-rich RNA (51 nucleotides) as substrates (6). The substrate RNA was analyzed on gels after digestion for 80 min or 6 h, as shown in Fig. 4 and summarized in Table 1. Two of three mutants with mutations of acidic residues directly involved in metal binding (E80A, D108A, and E119A) are inactive both with $MnCl_2$ and with $MgCl_2$. The exception was the E80A mutant, which retained activity only in the presence of manganese ions. The K134A mutant is also completely inactive,

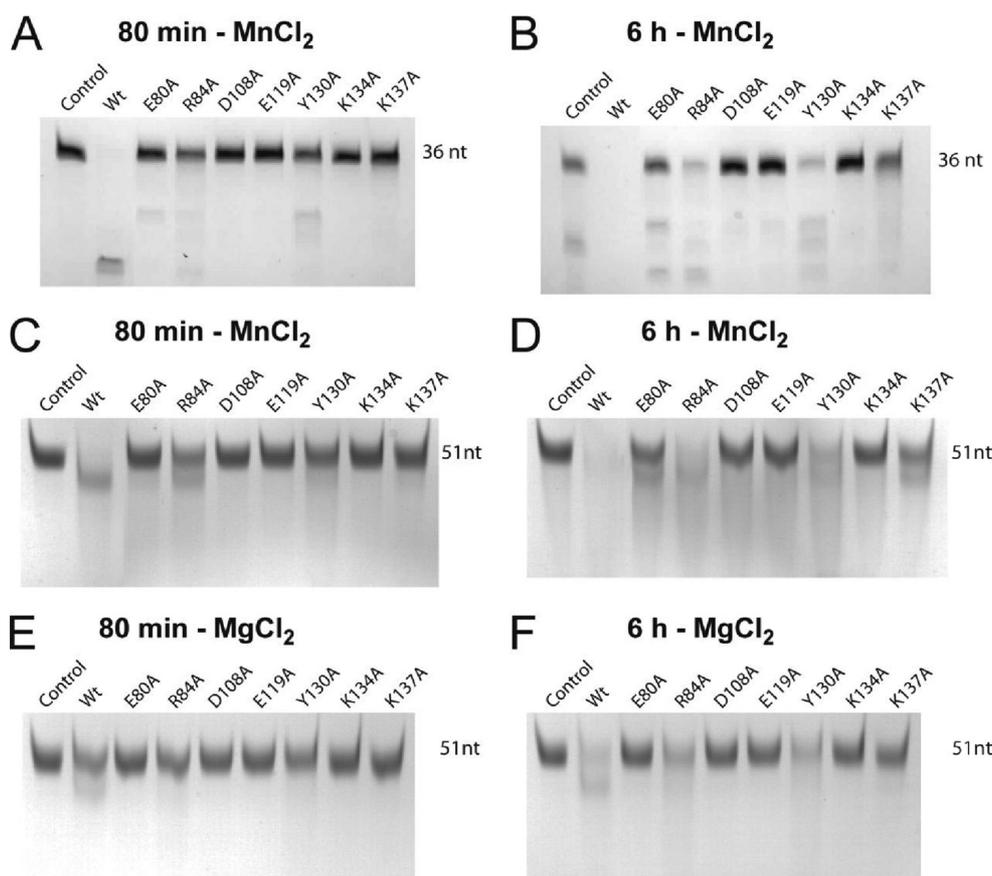


FIG. 4. Endonuclease activities of PA-Nter mutants. (A and B) RNase activities of wild-type and mutant PA-Nters in the presence of 1 mM $MnCl_2$ using a short panhandle RNA of 36 nucleotides (6) for 80 min (A) and 6 h (B). (C to F) The same experiment described for panels A and B but using U-rich RNA (51 nucleotides) plus 1 mM $MnCl_2$ (C and D) and using U-rich RNA plus 1 mM $MgCl_2$ (E and F). RNA cleavage was performed by incubating 13 μM PA-Nter for 80 min (A, C, and E) or 6 h (B, D, and F) with 10 μM RNA at 37°C in a final volume of 50 μl .

which is consistent with it being the catalytic lysine. The R84A, Y130A, and K137A mutants situated on the active-site rim are all less active than the wild type to various degrees but are not inactive. Also note that the activities of the wild type and all active mutants were higher in the presence of 1 mM $MnCl_2$ than in the presence of 1 mM $MgCl_2$ (Fig. 4C to F).

Metal binding monitored by thermal stabilization. We previously showed that the addition of 0.5 mM Mn^{2+} ions

significantly enhances the thermal stability of wt PA-Nter and that 0.5 mM dioxo-4-phenylbutanoic acid (DPBA), a known inhibitor of the influenza virus endonuclease (34), supershifts the T_m but only in the presence of metal ions (6). It is likely that the three oxygens on this inhibitor ligate the two resident metal ions in a similar manner, as has recently been observed in the integrase-inhibitor complex of retroviruses (14). To monitor metal binding to the mutant proteins, we therefore used the Thermofluor assay to measure the T_m in the presence of 0.1 mM $MnCl_2$ (Fig. 2C, light green), 5 mM $MgCl_2$ (Fig. 2C, light violet), or no metal (Fig. 2C, pink). We also measured the T_m in the presence of metal ions plus DPBA (Fig. 2C, red, dark green, and dark violet). Two of the three mutations that knock out the metal-ligating acidic residues (E80A, D108A, and E119A) also abolish metal ion binding. The exception was the E80A mutant, which was stabilized by Mn^{2+} but not by Mg^{2+} ions, consistent with the results of the endonuclease assay. All other mutations were stabilized by both types of cations and were further stabilized when DPBA was added in the presence of metal ions, similar to the result for the wt. In the absence of ions, the inhibitor did not stabilize the proteins.

Metal binding measured by ITC. In the PD-(D/E)XK nuclease superfamily, the positions equivalent to Asp108,

TABLE 1. Effects of mutations on RNase activity of PA-Nter and comparison with transcription and replication activities of the recombinant trimeric polymerase

Mutation	Role	RNase activity		Transcription ^a	Replication ^a
		$MnCl_2$	$MgCl_2$		
wt		++	++	+	+
H41A	Ligand Mn^{2+}	ND ^b	ND	-	-
E80A	Ligand Mn^{2+}	±	-	-	+
R84A	RNA positioning	+	+	+	+
D108A	Ligands Mn^{2+} and Mn^{2+}	-	-	-	+
E119A	Ligand Mn^{2+}	-	-	-	+
Y130A	RNA positioning	+	+	+	+
K134A	Catalytic	-	-	-	+
K137A	RNA positioning	±	±	+	+

^a The results obtained with influenza virus recombinant RNP (13, 37).

^b ND, not determined.

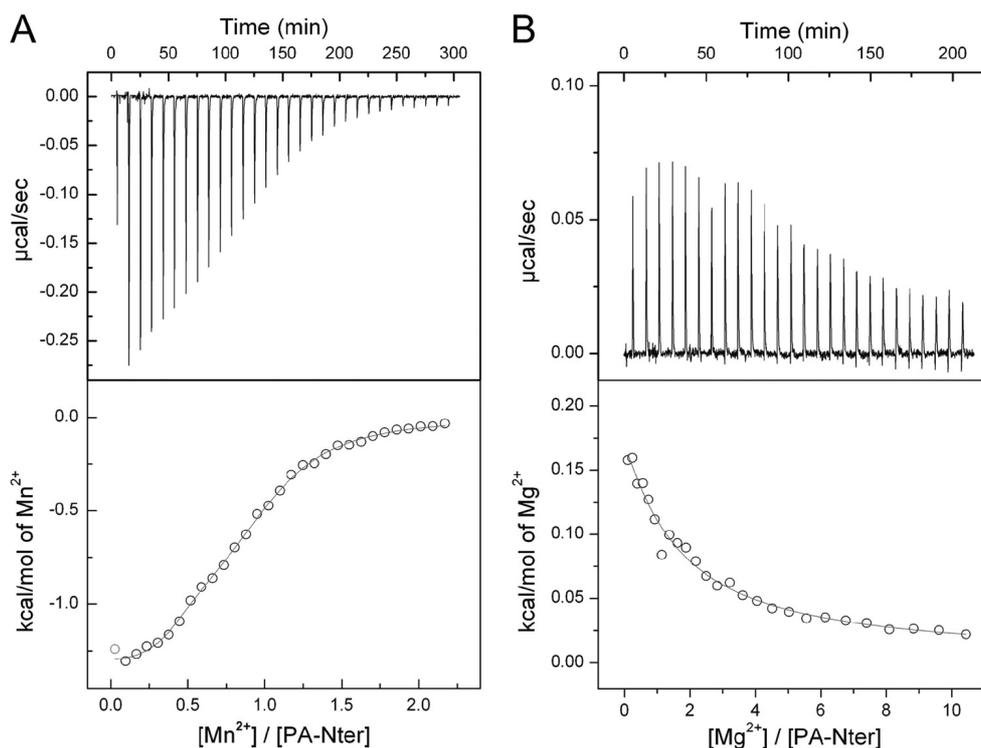


FIG. 5. Isothermal titration calorimetry of wild-type PA-Nter with metal ions. (A) MnCl_2 at 0.3 mM was added to 60 μM PA-Nter at 25°C in 20 mM Tris-HCl (pH 8.0) plus 100 mM NaCl; (B) MgCl_2 at 9 mM was added to 60 μM PA-Nter at 25°C in 20 mM Tris-HCl (pH 8.0) plus 100 mM NaCl. In the lower panels, the circles represent experimental data and the continuous lines correspond to the best fit to a model with two binding sites.

Glu119, and Lys134 are always present, giving the family its name. Glu80 and His41 are not always conserved, and indeed, this combination seems to be specific for the influenza virus endonuclease (6, 37). Glu80 is particularly interesting, as E80A shows binding and residual nuclease activity with Mn^{2+} ions but not with Mg^{2+} ions. In order to understand these observations, we directly measured Mn^{2+} and Mg^{2+} ion binding to wt PA-Nter and E80A by ITC (Fig. 5 and Table 2), which is the most appropriate technique for directly measuring the thermodynamics of protein-ligand binding (21). We find that binding of Mn^{2+} ions to wt PA-Nter is exothermic (Fig. 5A), whereas binding of Mg^{2+} ions is endothermic (Fig. 5B). The dissociation constants for the Mn^{2+} and Mg^{2+} ions are obtained by model fitting (Table 2), in which the number of binding sites and the values for the dissociation complex (K_d^{-1}) and enthalpy change (ΔH) are variables.

TABLE 2. Metal specificity of PA-Nter^a

Ion	Site	K_d (μM)	
		wt	E80A
Mn^{2+}	High affinity	0.3	5.2
	Low affinity	6.5	46
Mg^{2+}	High affinity	148	
	Low affinity	4,000	

^a Isothermal titration calorimetry experiments were performed on a VP-ITC calorimeter using protein solutions at a concentration of 60 μM and metal solution concentrations between 0.8 and 9 mM.

For the wt protein, the titration data for both Mn^{2+} and Mg^{2+} are best fitted with a two-site model since the quality of the fit (by the chi-square test) is better by a factor of 2 than that for a single-site model. Note that at this stage that we make no assumption about whether the two Mn^{2+} sites revealed by ITC are the same as the two Mg^{2+} sites or whether they correspond to sites M1 and M2 defined by the crystal structure (6). The affinity of wt PA-Nter for Mn^{2+} ions (K_d s for the two sites, 0.3 and 6.5 μM) is 500 to 600 times higher than that for Mg^{2+} ions (K_d s for the two sites, 148 and 4,000 μM). This is consistent with the higher endonuclease activity shown by wt PA-Nter in the presence of 1 mM Mn^{2+} compared to that in the presence of 1 mM Mg^{2+} ions (compare Fig. 4C and E). The measured affinities lie in the range of those obtained for other divalent cation binding proteins, such as RNases and proteases (3, 20, 36). The fact that the binding enthalpy is exothermic for Mn^{2+} ions and endothermic for Mg^{2+} ions suggests a more optimal coordination for bound Mn^{2+} ions, which is reflected in a more favorable enthalpy.

Addition of MgCl_2 to the E80A mutant produced no heat effect, in agreement with the absence of thermal stabilization and nuclease activity upon addition of Mg^{2+} ions. The binding data for Mn^{2+} could be modeled using a single site with a K_d of 77 μM and an occupancy of 2.1 or with two sites with K_d values of 5.2 and 46 μM . Because the chi-square test of the single-site model provides a value twice that of the two-site-model, we favor the two-site model. Although the Mn^{2+} binding affinities for the two-site model are only 10 times lower than those for wt PA-Nter, the endonuclease activity of this

mutant is impaired in the presence of 1 mM MnCl_2 , which may be due to a difference in the geometry of metal binding due to the absence of Glu80.

DISCUSSION

This work was undertaken with dual aims: first, to determine whether mutations in the isolated endonuclease domain had the same effects on activity as the equivalent mutations made in the trimeric polymerase and, second, to clarify whether the nuclease activity of the domain depends on binding to one or two Mg^{2+} or Mn^{2+} ions.

Concerning the first aim, all the alanine mutations that we made in PA-Nter have already been studied in the context of the trimeric recombinant polymerase (13, 37). As Table 1 shows, mutations that inactivate the endonuclease activity in the isolated PA-Nter domain knock out transcription but not replication in the intact polymerase. The experiments with the intact recombinant polymerase were performed in the presence of MgCl_2 without added MnCl_2 . Bearing in mind our results for this mutant, this could explain why the intact polymerase containing the E80A mutation was inactive in transcription in the assay of Hara and coworkers (13). The results for this mutation presented by Yuan and coworkers (37) are more ambiguous; although the mutant was inactive in the endonuclease assay, some globin mRNA-primed transcription activity was retained. Concerning the conserved Arg84, Tyr130, and Lys137 residues on the active-site rim, we hypothesized that these could be necessary for the binding and correct positioning of the substrate RNA in the active site. In particular, Arg84 is bound to a well-ordered sulfate ion in the crystal structure of PA-Nter (6). This sulfate is in the same position as one of the phosphates of the EcoRV restriction enzyme product complex (16). Tyr130 and Lys137 coordinate water molecules that bind to the monophosphate group in a complex of PA-Nter with nucleoside monophosphates (38). In the isolated endonuclease domain, the enzymatic activities of R84A, Y130A, and K137A were reduced, whereas in the context of the intact trimeric polymerase, the mutations retain full transcription activity (13). We may explain this difference by the fact that in the intact polymerase the prebinding of the cellular mRNA to the PB2 cap-binding domain probably considerably enhances substrate affinity by increasing the local concentration. Furthermore, the presence of other positively charged polymerase domains such as the highly basic surfaces of the PB2 627 domain (19, 32) and the C-terminal two-thirds of PA (15, 24) may also assist with the correct positioning of the substrate RNA over the endonuclease active site.

An additional observation is that the active site of the endonuclease is very acidic, like that of other endonucleases (Fig. 3A); and mutations that increase the negative charge destabilize the protein, whereas those that decrease the acidity stabilize PA-Nter. The single exception is the D108A mutant PA-Nter, which has the same stability as wt PA-Nter, for which we do not have an explanation. The H41A mutation in the context of the intact polymerase results in disruption not only of transcription but also of replication (13, 37). Therefore, it seems that the correct folding or the stability of PA-Nter is a prerequisite to the correct formation of the active site of the polymerase on the PB1 subunit. These observations also suggest a

rationale for why PA-Nter has a histidine at position 41 rather than a glutamate, as in EcoRV (5).

In conclusion, the accordance of the activities of PA-Nter with those of the intact recombinant trimer suggests that isolated PA-Nter has the same structure in the context of the intact trimer and validates the use of the isolated domain for drug screening and structure-based design. A similar conclusion was drawn for the PB2 cap-binding domain, i.e., that the isolated domain has the same structure as that in recombinant RNP (10). This would imply that activation of the cap-binding and endonuclease functions in the intact polymerase, by the binding of the 5' end of the vRNA or a 3'-5' duplex (12, 22), are likely not due to the induced formation of the two active sites but rather to the removal of inhibition (perhaps steric) of these functions. The structural differences that are observed between free polymerase and RNP-bound polymerase (1, 5, 35) suggest important domain rearrangements that may change the disposition of the cap-binding and endonuclease sites and the overall binding of host mRNA by the intact polymerase.

Concerning the second aim of our study, in a careful enzymatic analysis of the endonuclease activity of purified RNPs, Doan et al. showed that there are two interacting metal binding sites that need to be occupied with divalent metal ions for full nuclease activity (7). They found that manganese ions are two times stronger than magnesium ions at activating the endonuclease and showed that the affinity for Mn^{2+} ions is stronger than for Mg^{2+} ions. The metal dependence of the endonuclease activity shown here is in agreement with these results. The nuclease activity of PA-Nter and the mutants is higher in the presence of 1 mM MnCl_2 than in the presence of 1 mM MgCl_2 , and the shift in T_m is higher in the presence of 0.1 mM Mn^{2+} ions than in the presence of 5 mM Mg^{2+} ions.

Yuan and coworkers grew crystals of PA-Nter in the presence of 100 mM MgCl_2 and observed a single metal ion only in position M1, even in the presence of mononucleotide phosphates (38). Therefore, in the absence of substrate, Mg^{2+} ions seem to bind only to the M1 site even at concentrations that lay several orders of magnitude above the low affinity K_d (4 mM). Because a second magnesium ion was never seen in the active site, it is possible that the low-affinity binding site for Mg^{2+} lies outside the active site. Histidine is one of the ligands of the M2 site in the crystal structure (6). Whereas manganese ions can be favorably coordinated by both acidic residues and histidine (2, 11), ligation of Mg^{2+} ions by histidine is uncommon (8). This agrees with the biochemical and enzymatic data presented here for the E80A mutant (Glu80 is an M1 ligand), which showed neither RNase activity nor binding of Mg^{2+} ions, although the mutant could still bind two manganese ions. All crystallographic data (37, 38) and our enzymatic and ITC data presented here suggest that only a single magnesium ion can bind to the enzymatic cavity of PA-Nter in the absence of substrate. However, Doan et al. (7) found a Hill coefficient of 2 for the endonuclease activity of intact viral RNPs in the presence of Mg^{2+} , suggesting that two ions can bind in the presence of substrate. The finding that the stability of wt PA-Nter supershifts in the presence of Mg^{2+} plus the inhibitor also suggests that two ions can bind when PA-Nter is stabilized by the inhibitor.

Dias et al. grew crystals in a mixture of 2.5 mM MnCl_2 and

5 mM MgCl₂ and located two Mn²⁺ ions in positions M1 and M2 (6). Two manganese ions were also found by Zhao and co-workers upon adding MnCl₂ (38). Our ITC results explain these structural observations, since they show that Mn²⁺ can bind with a high affinity to two sites (K_d s, 0.3 and 6.5 μM) and that Mn²⁺ binding to its second site is 20-fold higher than that of Mg²⁺ ions to their high-affinity site (K_d , 148 μM). Although with these data we cannot prove that the two Mn²⁺ binding sites obtained through ITC correlate with the two binding sites observed in the crystal structure, it is likely that it is the case. This is supported by the results on activity and Mn²⁺ binding of the E80A mutant, which has an active site that resembles that of EcoRV, which also binds to two manganese ions.

In conclusion, our results are consistent with previous results on the metal dependence of the endonuclease in intact RNPs. We show quantitatively that the endonuclease active site binds to two Mn²⁺ ions and has a significantly higher affinity for Mn²⁺ ions than for Mg²⁺ ions. As was mentioned by Zhao et al., the cellular concentration of free magnesium ions is in the millimolar range, whereas that of manganese ions is in the micromolar range (38), making roles for both ions in the endonuclease activity during infection by influenza virus entirely possible. This suggestion is strengthened by the findings of Doan et al. that indicated a synergistic activation of cleavage activity with combinations of different metal ions (7).

ACKNOWLEDGMENTS

We acknowledge the Partnership for Structural Biology for an integrated structural biology environment.

The work was partially funded by the EU FLUPOL contract (SP5B-CT-2007-044263), the ANR FLU INTERPOL contract (ANR-06-MIME-014-02), and Lyon Biopôle. Alexandre Dias was the recipient of a Ph.D. fellowship from the French MENRT.

REFERENCES

1. Area, E., J. Martin-Benito, P. Gastaminza, E. Torreira, J. M. Valpuesta, J. L. Carrascosa, and J. Ortin. 2004. 3D structure of the influenza virus polymerase complex: localization of subunit domains. *Proc. Natl. Acad. Sci. U. S. A.* **101**:308–313.
2. Bewley, M. C., and J. M. Flanagan. 2001. Arginase, p. 952–962. In A. Messerschmidt, R. Huber, T. Poulos, and K. Wieghart (ed.), *Handbook of metalproteins*. John Wiley & Sons, Ltd., Chichester, United Kingdom.
3. Brazier, M. W., P. Davies, E. Player, F. Marken, J. H. Viles, and D. R. Brown. 2008. Manganese binding to the prion protein. *J. Biol. Chem.* **283**:12831–12839.
4. Cianci, C., L. Tiley, and M. Krystal. 1995. Differential activation of the influenza virus polymerase via template RNA binding. *J. Virol.* **69**:3995–3999.
5. Coloma, R., J. M. Valpuesta, R. Arranz, J. L. Carrascosa, J. Ortin, and J. Martin-Benito. 2009. The structure of a biologically active influenza virus ribonucleoprotein complex. *PLoS Pathog.* **5**:e1000491.
6. Dias, A., D. Bouvier, T. Crepin, A. A. McCarthy, D. J. Hart, F. Baudin, S. Cusack, and R. W. Ruigrok. 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458**:914–918.
7. Doan, L., B. Handa, N. A. Roberts, and K. Klumpp. 1999. Metal ion catalysis of RNA cleavage by the influenza virus endonuclease. *Biochemistry* **38**:5612–5619.
8. Dupureur, C. M. 2008. Roles of metal ions in nucleases. *Curr. Opin. Chem. Biol.* **12**:250–255.
9. Ericsson, U. B., B. M. Hallberg, G. T. Detitta, N. Dekker, and P. Nordlund. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**:289–298.
10. Guilligay, D., F. Tarendeau, P. Resa-Infante, R. Coloma, T. Crepin, P. Sehr, J. Lewis, R. W. Ruigrok, J. Ortin, D. J. Hart, and S. Cusack. 2008. The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15**:500–506.
11. Guss, J. M., and H. C. Freeman. 2001. Aminopeptidase, p. 973–980. In A. Messerschmidt, R. Huber, T. Poulos, and K. Wieghart (ed.), *Handbook of metalproteins*. John Wiley & Sons, Ltd., Chichester, United Kingdom.
12. Hagen, M., T. D. Chung, J. A. Butcher, and M. Krystal. 1994. Recombinant influenza virus polymerase: requirement of both 5' and 3' viral ends for endonuclease activity. *J. Virol.* **68**:1509–1515.
13. Hara, K., F. I. Schmidt, M. Crow, and G. G. Brownlee. 2006. Amino acid residues in the N-terminal region of the PA subunit of influenza A virus RNA polymerase play a critical role in protein stability, endonuclease activity, cap binding, and virion RNA promoter binding. *J. Virol.* **80**:7789–7798.
14. Hare, S., S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov. 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**:232–236.
15. He, X., J. Zhou, M. Bartlam, R. Zhang, J. Ma, Z. Lou, X. Li, J. Li, A. Joachimiak, Z. Zeng, R. Ge, Z. Rao, and Y. Liu. 2008. Crystal structure of the polymerase PA(C)-PB1(N) complex from an avian influenza H5N1 virus. *Nature* **454**:1123–1126.
16. Horton, N. C., and J. J. Perona. 2004. DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry* **43**:6841–6857.
17. Imhof, P., S. Fischer, and J. C. Smith. 2009. Catalytic mechanism of DNA backbone cleavage by the restriction enzyme EcoRV: a quantum mechanical/molecular mechanical analysis. *Biochemistry* **48**:9061–9075.
18. Klumpp, K., R. W. Ruigrok, and F. Baudin. 1997. Roles of the influenza virus polymerase and nucleoprotein in forming a functional RNP structure. *EMBO J.* **16**:1248–1257.
19. Kuzuhara, T., D. Kise, H. Yoshida, T. Horita, Y. Murazaki, A. Nishimura, N. Echigo, H. Utsunomiya, and H. Tsuge. 2009. Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J. Biol. Chem.* **284**:6855–6860.
20. Lai, B., Y. Li, A. Cao, and L. Lai. 2003. Metal ion binding and enzymatic mechanism of Methanococcus jannaschii RNase HII. *Biochemistry* **42**:785–791.
21. Leavitt, S., and E. Freire. 2001. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11**:560–566.
22. Lee, M. T., K. Klumpp, P. Digard, and L. Tiley. 2003. Activation of influenza virus RNA polymerase by the 5' and 3' terminal duplex of genomic RNA. *Nucleic Acids Res.* **31**:1624–1632.
23. Li, M. L., P. Rao, and R. M. Krug. 2001. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. *EMBO J.* **20**:2078–2086.
24. Obayashi, E., H. Yoshida, F. Kawai, N. Shibayama, A. Kawaguchi, K. Nagata, J. R. Tame, and S. Y. Park. 2008. The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454**:1127–1131.
25. Ortega, J., J. Martin-Benito, T. Zurcher, J. M. Valpuesta, J. L. Carrascosa, and J. Ortin. 2000. Ultrastructural and functional analyses of recombinant influenza virus ribonucleoproteins suggest dimerization of nucleoprotein during virus amplification. *J. Virol.* **74**:156–163.
26. Pingoud, V., W. Wende, P. Friedhoff, M. Reuter, J. Alves, A. Jeltsch, L. Mones, M. Fuxreiter, and A. Pingoud. 2009. On the divalent metal ion dependence of DNA cleavage by restriction endonucleases of the EcoRI family. *J. Mol. Biol.* **393**:140–160.
27. Plotch, S. J., M. Bouloy, I. Ulmanen, and R. M. Krug. 1981. A unique cap(m⁷GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23**:847–858.
28. Poon, L. L., D. C. Pritlove, E. Fodor, and G. G. Brownlee. 1999. Direct evidence that the poly(A) tail of influenza A virus mRNA is synthesized by reiterative copying of a U track in the virion RNA template. *J. Virol.* **73**:3473–3476.
29. Robertson, J. S., M. Schubert, and R. A. Lazzarini. 1981. Polyadenylation sites for influenza virus mRNA. *J. Virol.* **38**:157–163.
30. Rocchia, W., S. Sridharan, A. Nicholls, E. Alexov, A. Chiabrera, and B. Honig. 2002. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: applications to the molecular systems and geometric objects. *J. Comput. Chem.* **23**:128–137.
31. Saito, T., D. M. Owen, F. Jiang, J. Marcotrigiano, and M. Gale, Jr. 2008. Innate immunity induced by composition-dependent RIG-I recognition of hepatitis C virus RNA. *Nature* **454**:523–527.
32. Tarendeau, F., T. Crepin, D. Guilligay, R. W. Ruigrok, S. Cusack, and D. J. Hart. 2008. Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.* **4**:e1000136.
33. Tiley, L. S., M. Hagen, J. T. Matthews, and M. Krystal. 1994. Sequence-specific binding of the influenza virus RNA polymerase to sequences located at the 5' ends of the viral RNAs. *J. Virol.* **68**:5108–5116.
34. Tomassini, J., H. Selnick, M. E. Davies, M. E. Armstrong, J. Baldwin, M. Bourgeois, J. Hastings, D. Hazuda, J. Lewis, W. McClements, et al. 1994. Inhibition of cap (m⁷GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob. Agents Chemother.* **38**:2827–2837.
35. Torreira, E., G. Schoehn, Y. Fernandez, N. Jorba, R. W. Ruigrok, S. Cusack, J. Ortin, and O. Lorca. 2007. Three-dimensional model for the isolated

- recombinant influenza virus polymerase heterotrimer. *Nucleic Acids Res.* **35**:3774–3783.
36. **Wehenkel, A., M. Bellinzoni, F. Schaeffer, A. Villarino, and P. M. Alzari.** 2007. Structural and binding studies of the three-metal center in two mycobacterial PPM Ser/Thr protein phosphatases. *J. Mol. Biol.* **374**:890–898.
37. **Yuan, P., M. Bartlam, Z. Lou, S. Chen, J. Zhou, X. He, Z. Lv, R. Ge, X. Li, T. Deng, E. Fodor, Z. Rao, and Y. Liu.** 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. *Nature* **458**:909–913.
38. **Zhao, C., Z. Lou, Y. Guo, M. Ma, Y. Chen, S. Liang, L. Zhang, S. Chen, X. Li, Y. Liu, M. Bartlam, and Z. Rao.** 2009. Nucleoside monophosphate complex structures of the endonuclease domain from the influenza virus polymerase PA subunit reveal the substrate binding site inside the catalytic center. *J. Virol.* **83**:9024–9030.

Abstract

Influenza A virus is a negative-strand RNA virus belonging to the Orthomyxoviridae family whose replication occurs in the nucleus of infected cells. The genome organisation of influenza virus is segmented in eight vRNA segments of negative polarity coding for at least 16 different viral proteins. Each vRNA is bound to multiple copies of nucleoprotein (NP) and to the heterotrimeric RNA-dependent RNA-polymerase complex (PA, PB1 and PB2) through its 5' and 3' extremities. This macromolecular assembly (vRNA/polymerase/NP) forms the ribonucleoprotein (RNP) particle. The RNP complex is at the core of viral replication and in the context of RNPs, the polymerase performs both transcription and replication of the vRNA genome. As such, the polymerase constitutes a major antiviral drug target. The research work presented within this thesis focuses on the underlying determinants of the RNA polymerase assembly process and its interaction with its vRNA genome. To fulfill these goals, our lab, in collaboration with other groups, has set up a novel polyprotein expression system to express the polymerase but also to reconstitute polymerase and cellular partner complexes, notably RanBP5, which belongs to the importin- β family.

Résumé

Le virus de la grippe A est un virus à ARN négatif appartenant à la famille des Orthomyxoviridae dont la réplication se produit dans le noyau des cellules infectées. L'organisation du génome est segmentée en huit segments d'ARNv de polarité négative, codant pour un minimum de 16 protéines virales différentes. Ces ARN viraux (ARNv) sont en complexe avec de nombreuses copies de nucléoprotéines et liés par leurs extrémités 5' et 3' au complexe hétérotrimérique de l'ARN-polymérase ARN-dépendante composé des sous unités PA, PB1 et PB2. Cet assemblage macromoléculaire (ARNv / polymérase / NP) nommée Ribonucléoprotéine (RNP) constitue une entité génomique indépendante. Dans le contexte de la RNP, l'ARN-polymérase assure à la fois la transcription et la réplication du génome ARNv. En assurant ces deux fonctions, l'ARN-polymérase joue un rôle majeur dans la réplication virale et constitue une cible antivirale privilégiée. Les travaux de recherche présentés dans cette thèse se concentrent sur les éléments structuraux participants à l'assemblage de l'ARN polymérase et son interaction avec les ARNv. Pour atteindre ces objectifs, notre laboratoire, en collaboration avec d'autres groupes, a mis en place un système d'expression en polyprotéines permettant d'exprimer la polymérase. Plus encore, cette méthode a aussi permis de reconstituer des complexes entre l'ARN-polymérase et des partenaires cellulaires, notamment RanBP5 qui appartient à la famille des importines- β .