Svoluji k zapůjčení své diplomové práce ke studijním účelům a prosím, aby byla vedena přesná evidence vypůjčovatelů. Převzaté údaje je vypůjčovatel povinen řádně ocitovat.

Charles University

Faculty of Science

Study programme: Chemistry

Branch of study: Biophysical Chemistry

Bc. Jakub Hejdánek

Identification of small compounds disrupting protein-protein interaction in influenza A polymerase

Identifikace sloučenin rozrušujících protein-proteinovou interakci u polymerasy viru chřipky

Diploma thesis

Supervisor: Jan Konvalinka Ph.D.

Prague, 2018

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením školitele doc. RNDr. Jana Konvalinky, CSc. a všechny použité prameny jsem řádně citoval.

V Praze 4.5.2018   ………………………………

Jakub Hejdánek

Acknowledgments

Contents

List of abbreviations

AA	Amino acid
BTTP	(3-[4-({bis[(1-terc-butyl-1H-1,2,3-triazol-4-yl)methyl]amino}methyl)-1H-1,2,3-triazol-1-yl]propanol
CBB	Coomassie Brilliant Blue
CDC	Centres for Disease Control and Prevention
CPA	C-terminal domain of polymerase acidic subunit
CPM	Cytoplasmatic membrane
Da	Dalton
DANA	2-deoxy-2,3-dihydro-N-acetylneuraminic acid
DIANA	DNA-linked Inhibitor Antibody Assay
DLS	Dynamic light scattering
DMSO	Dimethylsulfoxid
EA	Ethanolamine
EDC	1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride
ELISA	Enzyme-Linked Immuno Sorbent Assay
ER	Endoplasmatic reticulum
Fmoc	Fluorenylmethyloxycarbonyl chloride
GSH	L-glutathione reduced
GST	Glutathione-S transferase
HA	Hemagglutinin
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LCMS	Liquid chromatography-mass spectrometry
NA	Neuraminidase
NANA	N-acetyl neuraminic acid (sialic acid)
NEP	Nuclear export protein
NHS	N-hydroxysuccinimide
NLS	Nuclear localization signal
NP	Nucleocapsid protein
RNA	Ribonucleic acid
PA	Polymerase acidic protein

PB1	Polymerase basic protein 1
PB2	Polymerase basic protein 2
PCR	Polymerase chain reaction
PPI	Protein-protein interaction
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate-polyacryleamide gel electrophoresis
SPR	Surface Plasmon Resonance
SUMO	Small ubiquitin-like modifier
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TMD	Trans membrane domain
vRNA	Viral RNA
vRNP	Viral ribonucleoprotein particle
WHO	World health organization
WT	Wild type

Annotations:

IC_{50}: The concentration of an inhibitor that is required for 50-percent inhibition of an enzyme *in vitro*

C_q: the value of the cycle at which the curvature of the amplification curve during real-time PCR is maximal

Abstract

Influenza virus causes severe respiratory infections in birds and mammals and it is responsible for up to half a million deaths of human beings worldwide each year. Two molecular targets in influenza viral life cycle, neuraminidase and M2 proton channel are exploited in treatment. However, the recent emergence of new pandemic type along with increasing resistance against approved drugs has urged the need for a new drug target discovery and potential search of its inhibitor. Recently, an interesting protein-protein interaction between two subunits PA and PB1 of influenza A viral polymerase has been identified by X-ray crystallography as a new promising drug target. The fact that relatively few residues drive the binding and that the binding interface is highly conserved presents an intriguing possibility to identify antiviral lead compounds effective against all subtypes of influenza A virus.

In our laboratory, we expressed and purified two fusion tag constructs of the recombinant C-terminal domain of polymerase acidic subunit (CPA) from the pandemic isolate A/California/07/2009 H1N1. First, GST-CPA fusion protein was used for kinetic evaluation of PA-PB1 interaction by surface plasmon resonance. Moreover, this construct was used in the development of high-throughput screening method for search of interaction disrupting molecules based on AlphaScreen technology. We utilized this assay to determine the effect of truncating minimal PB1 peptide responsible for mediating CPA-PB1 interaction. Furthermore, we designed additional construct of CPA fused with protein His_6-SUMO for efficient over-expression and subsequent crystallization experiments. We set up co-crystallization experiments by the vapor-diffusion method and were later able to obtain 3D structure disclosing information employable in rational drug design. Lastly, the secondary high-throughput screening assay is being developed in order to screen broad compound libraries for the search of novel inhibitors.

Keywords: influenza polymerase, protein-protein interaction, high-throughput screening assay, protein crystallography, AlphaScreen assay, DNA-linked inhibitor antibody assay

Abstrakt

Virus chřipky napadá ptactvo a savce a způsobuje u nich závažnou infekci, která je ročně příčinou téměř půl milionu lidských obětí. Dodnes jsou terapeutickými cíly pouze dvě molekuly v životním cyklu viru – M2 protonový kanál a neuraminidasa. Vznik nových pandemických subtypů je současně s vývojem resistentních variant proti současným lékům závažným podnětem pro nalezení nových cílů a potenciálně jejich inhibitorů. V nedávné době byla pomocí strukturní krystalografie identifikována protein-proteinová interakce mezi PA a PB1 podjednotkou virové polymerasy jakožto zajímavý therapeutický cíl. Zajímavé je, že tato interakce je zprostředkována jen několika interagujícími aminokyselinovými zbytky. Další zajímavý fakt je, že tyto zbytky vykazují vysokou konzervovanost mezi variantami viru chřipky typu A, což by v případě nalezení inhibitoru mohlo umožnit zacílit širokospektrální léčbu proti všem kmenům tohoto typu viru.

V této práci jsem se zabýval expresí a purifikací dvou fúzních rekombinantních proteinových konstruktů C-terminální domény PA podjednotky (CPA) z pandemického subtypu viru chřipky A/California/07/2009 (H1N1). První konstrukt byl navrhnut k evaulaci kinetických vlastností interakce mezi PB1 peptidem a CPA proteinem. Dále byl tento konstrukt použit k vývoji vysoko-kapacitní testovací analýzy látek na bázi technologie AlphaScreen. Pomocí této analýzy jsme testovali zkracované PB1 peptidy s úmyslem vyhledat minimální peptid, který by si zachoval vazebné vastnosti vůči proteinu CPA. Dále jsem navrhl druhý rekombinantní protein sestávající se z hexahistidinové kotvy spojené s proteinem SUMO zlepšujícím rozpustnost, vázaným na CPA protein. Tento protein byl po odštěpení fúzní kotvy následně využit pro krystalizační experimenty metodou difuze par, které vedly k získání 3D struktury poskytující informace využitelné v racionálním návrhu léčiv. Na závěr byla navržena další vysokokapacitní testovací metoda pro hledání sloučenin v knihovnách pro nalezení nových látek rozrušujících protein-proteinové interakce.

Klíčová slova: polymerasa viru chřipky, protein-proteinová interakce, vysokokapacitní testovací metoda pro hledání sloučenin, proteinová krystalografie, testovací metoda AlphaScreen, testovací metoda DIANA

1. Theoretical part

1.1. Influenza

1.1.1. Introduction

Influenza, more commonly referenced as the "flu" is one of the most common and worldwide known human diseases. It has a history dating back to ancient ages and it was one of the key factors during development of major human civilizations. Influenza virus circulates in the very broad spectrum of species and transmits between them with no difficulty with the potential to cause local epidemics as well as global pandemics. Permanently increasing world interconnection supported by continual population growth has a major impact on influenza viral evolution and on its ability to cause epidemics. On the other hand, this human resource cross-connection enhances world joined forces in the fight against influenza. World Health Organization (WHO) estimates 1 billion influenza cases, 4 million severe illness cases, and approximately 500 000 influenza-related deaths each year [1]. In summary, influenza is one of the diseases with the highest surveillance priority as well as farmaceutical interest.

1.1.2. Influenza epidemiology

Influenza disease can affect all people, however, certain groups can be more or less susceptible to this illness based on various factors. Influenza presents a greater risk for the elderly, children, women during pregnancy, individuals with additional medical conditions (pulmonary, cardiac etc.) or people with the suppressed immune system. Influenza is characteristic of its rapid transmission mainly in crowded areas such as schools, public transportation etc. Virus is transferable through infectious droplets dispersed in air as result of a cough or sneeze of an infected individual. Seasonal influenza epidemics outbreaks in temperate climates occur particularly during cold seasons [1].

1.2. Influenza virus

1.2.1. Classification

Influenza virus belongs to the family of viruses called *Orthomyxoviridae.* This virus has segmented RNA genome of negative polarity (reverse complement strand corresponds to functional mRNA molecule) and it carries envelope acquired from host

cells. Influenza virus can be divided into four types: A, B, C, D. These four aforementioned types differ in hosts: type A infects humans and various animals, B is transmitted only amongst humans, C has two hosts in humans and swine and type D primarily affects cattle and is not known to infect humans. Moreover, these types can be contradistinguished based on their immunological and biological differences. The A and B types have 8 segments of single-stranded vRNA whereas type C has only 7 segments [2–5]. From all four types, type A is the most relevant in relation to public health risks since it is responsible for 80% of seasonal epidemics. This type has potential to cause pandemics due to reassortment (see section 1.2.2). Its reservoir can be found amongst aquatic birds. The virus can be further transmitted and reassorted in swine which can potentially lead to the new highly pathogenic subtypes. Viruses can be classified depending on their host of origin, for example, "human influenza", "swine influenza", "avian influenza", etc. There are several antigenically distinct subtypes that are defined by two major antigens (viral surface proteins): hemagglutinin (HA) and neuraminidase (NA). Up to this date, there are 18 hemagglutinin subtypes and 11 neuraminidase subtypes known. By combining two specific antigens, original influenza type A virus can be recognized as, for example, A(H1N1)[1,4–6]. Newly isolated strains are named using the internationally accepted convention for influenza viruses published by WHO in 1979. This nomenclature is composed of following components [4]:

- Antigenic type (e.g. A, B, C)
- A host of origin (e.g. swine, avian etc. For human-origin no host is designated.)
- Geographical origin (e.g. California, Puerto Rico, etc.)
- Strain number (e.g. 1, 07, etc.)
- Year of isolation (e.g. 57, 2009, etc.)
- For type A of influenza viruses, HA and NA antigen is presented in parentheses (e.g. (H1N1))

for example A/California/07/2009 (H1N1) or A/Duck/Alberta/60/1976 (H12N5).

1.2.2. Influenza virus evolution

Due to the broad and robust sequence analyses over years, many studies describe and discuss the diversity of influenza A viruses. The indicated complexity is the result of the natural selection (phylogeny, epidemiology). Natural selection is mainly influenced by host immunity thus variants of HA (and less of NA) able to avoid immune response drive viral evolution. This can be observed through time and this phenomenon is called antigenic drift [5,7,8]. These small genetic changes are usually very closely phylogenetically related and possess similar antigenic properties. Nevertheless, they can accumulate over time and result in the genesis of new antigenic subtype no longer recognized by host immunity. Furthermore, analysis of these phylogenetical regularities and patterns play great importance in the prediction of newly forming strains and subsequently in preparation of vaccines [4,9].

Antigenic drift has considerable relevance in the emergence of new epidemics, however, the key factor at the beginning of new pandemics is an antigenic shift (reassortment) present only in influenza A types. This event resembles formation of a new combination of two aforesaid protein antigens resulting in subtype to which population is no longer susceptible. These new subtypes are generated when two distinct isolates infect the same host cell and viral segments encoding main antigens are combined resulting in the creation of new antigenically unique virus to which is the population immunologically naive [5,10].

1.2.3. Influenza pandemic

Influenza pandemics are large-scale epidemics affecting the worldwide population in 10 to 50-year intervals. They are the result of newly emerged viral strains mainly due to reassortment and can be only hardly predicted. Pandemics are usually spreading rapidly, especially in recent years due to progressive global migration. In consequence, it is hard to come with the early response and proper precautions. Evidence of 10-13 pandemics can be dated back to AD 1590 by way of historical records of high morbidity and mortality instances [1,2,11].

1.2.3.1. 1918 Spanish flu

Until this day, it is still hard to evaluate all circumstances that contributed to this most severe pandemic ever registered. The only evidence supporting the severity of 1918

Spanish flu is a characterization of isolates that provided information of high virulence. It is estimated that 50 % of world's population was infected and 25 % of the population experienced serious clinical symptoms. Global death toll estimate is approximately 50 millions. The noticeable distinction of this pandemics is that it came in several waves during the years 1918-1919. Another irregularity in comparison to other pandemics is that it caused death amongst young people (20-39 years old), children under 1-year-old and pregnant woman with case fatality reaching 2 %. Such high mortality is abnormal since typical pandemic reaches approximately 0.1 % and elderly are the ones that suffer most [2,12,13]. Reason for elderly resilience is probably circulation of the antigenically similar virus before the year 1889. The high mortality, on the other hand, is caused by exceptional virulence of circulating subtype during this pandemic. A major factor was also (precedent) nonexistence of bacterial antibiotics since pandemic was accompanied by secondary bacterial infections. These superinfections were primarily caused by *Streptococcus pneumoniae*, *Streptococcus pyogenes* and *Haemophilus influenzae*. They caused severe bacterial pneumonia symptomatic with pulmonary hemorrhage and lung flooding [14,15].

1.2.3.2. 2nd Millennium initiative

Prior to identification of avian influenza A(H5N1) virus, there was the clear perception of avian influenza untransmisibillity to humans. In 1997, this belief was disproven when the highly pathogenic avian virus was found to successfully transmit and infect humans in Hong Kong [16,17]. This threat was quickly eliminated by a reduction of live bird market. During 2003, the origin of the new respiratory virus was observed: severe acute respiratory syndrome (SARS) virus. Fortunately, this threat was rapidly eliminated but later highly pathogenic (H5N1) returned in Vietnam and Thailand with an extremely high fatality rate of 80% [18,19]. More cases were later observed and direct transmission from poultry to humans was documented. However, there were clear cases where transmission from human to human was evident. That raised concerns about effective viral adaptation to humans [20]. Considering all that happened during 2003, the formation of a new initiative that would access national pandemic planning was highly advised. This concept was mainly taken up by World Health Organization which constituted International Health Regulations that came into effect in 2007 and later served as a framework for pandemic response. This regulation focused on the

classification of pandemic phases, overall characteristics, and basis of transmission. Thenceforth, revision of vaccine and antivirals manufacturing, as well as nonpharmaceutical interventions, commenced. Many countries began to stockpile antivirals in response to long-lasting (H5N1) pandemic imminence. Nonpharmaceutical interventions mainly focused on measures that were effective in historical context. Secondly, experiments testing the role of factors such as hygiene on the acquisition of flu disease were done. Besides that, broad computational analysis of border closing and closing the institutions such as schools was executed [2,21].

1.2.3.3. 2009 pandemic

The aforementioned initiative focused highly on (H5N1) virus. Consequently, the emergence of new pandemic (H1N1) virus that first occurred in 2009 came as a complete surprise. This virus was specific in a way that it did not originate from reassorting. H1N1 the virus was previously globally circulating as seasonal influenza virus and there was no thought of the possibility of this virus to cause a pandemic. This pandemic (H1N1) 2009 virus first emerged in Mexico and was detected in California in April. The virus was confirmed as antigenically related to 1918 Spanish influenza virus. Unfortunately, the pandemic was spreading rapidly as a result of frequent plane travel. Most countries followed International Health Regulations proposed by WHO. However, they differed in execution. Japan had large stockpiles of antivirals and they achieved lower mortality compared to other countries. The United Kingdom, on the other hand, tried to use antivirals as prophylaxis which was not efficient so they relinquished this approach and later focused on the treatment of infected patients. Some countries even processed nonpharmaceutical interventions depending on severity in a specific region. Overall characteristics were different compared to any previous pandemic. Morbidity percentage was very high however mortality rates did not reach high numbers. Secondary bacterial superinfections were not common. People older than 64 years were not usually infected due to acquired immunity from previous encounters with the antigenically similar virus. Centers for Disease Control and Prevention (CDC) estimated over 60 million cases, over 250 000 hospitalizations and around 13 000 death in the USA itself. Global people death estimate was between 150 000 to 600 000 [2,21,22].

1.2.4. Influenza virus morphology

As it was mentioned previously, we recognize four types of influenza viruses based on their immunological and biological properties. Type A and B have 8 segments of RNA in contrast to type C which has only 7 segments. Virus particles are pleomorphic. They can resemble rough spheroidal particle (Figure 1) approximately 100 nm length in diameter or filamentous shape that can be longer than 1 μm. Morphology is dependent on how the virus is grown (e.g. polarity of epithelial cells, lipid composition of the cytoplasmatic membrane) [2,23]. The viral shape is also highly dependent on local pH value. Three viral components are a viral envelope, matrix, and core.

1.2.4.1. Viral Envelope

The lipid bilayer is the main component of the viral envelope. There are three transmembrane proteins anchored inside this bilayer: hemagglutinin, neuraminidase, and M2 ion channel. Bilayer itself is "stolen" from host cells upon viral budding, thus the composition is similar to the cytoplasmatic membrane of host cells enriched in sphingolipids and cholesterol rafts. M2 channel is cholesterol-binding protein and is therefore buried inside these rafts. HA forms homotrimers and is most abundant envelope protein with a ratio of 80% to 17% of neuraminidase forming homotetramer. M2 channel is also homotetramer with approximately 16-20 molecules per virion [3,24].
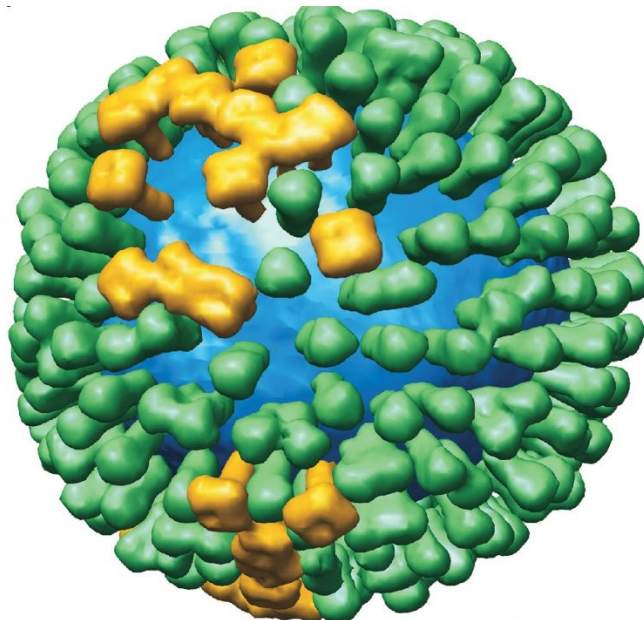
Figure 1: Spheroidical model of influenza virion coated with surface glycoproteins neuraminidase (Gold) and Hemagglutinin (Green). The lipid bilayer is blue. Reproduced from [25].

1.2.4.2. Viral matrix

Viral matrix is viral scaffold formed by M1 protein – the most abundant protein in influenza virus. This protein matrix is directly underneath lipid bilayer and is bound to it, hereafter forming an important bridge between envelope and core. M1 protein also interacts with HA, NA, and M2 on the outer side of the matrix and with nucleocapsid protein (NP) on the inner side. An additional role of M1 protein is during viral uncoating and disassembly, nuclear import and export of viral ribonucleoprotein particles (vRNPs), transcription, replication and finally budding from the host cells [3,24,26].

1.2.4.3. Viral core

Viral core is formed by helical ribonucleoprotein segments. RNA has a negative polarity which means it is first transcribed in reverse order. vRNA is coated with nucleocapsid proteins and with the low frequency with nuclear export protein (NEP). Every segment also bears polymerase complex at the end of vRNA. This complex is heterotrimeric RNA dependent RNA polymerase consisting of 3 subunits: polymerase basic protein 1 (PB1), polymerase basic protein 2 (PB2) and polymerase acidic protein (PA). vRNPs resemble rod-like structure (Figure 2) with a width about 13 nm in diameter and length up to 120 nm. vRNA is folded in half, coiled on itself and wrapped around NPs [24,27].
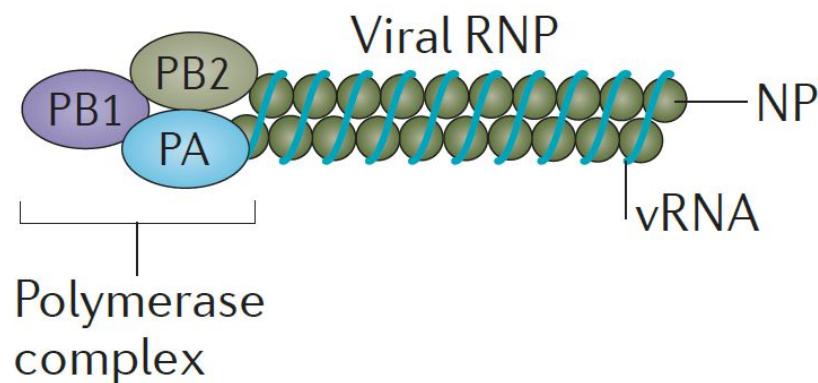
Figure 2: Graphical representation of ribonucleoprotein particle. Reproduced from [28].

1.2.5. Viral genome

The viral genome is segmented. These segments differ in size depending on which protein they encode, and are covered with nucleocapsid protein that binds to negatively charged vRNA through arginine-rich peptide sequences with a positive charge. All of

these NP are arranged alongside vRNA in a periodical fashion. Three largest segments (1, 2 and 3) encode PB2, PB1, and PA proteins, respectively, in decreasing manner. Segment 2 also contains genes of proteins PB1-F2 (proapoptotic virulence factor) and a truncated form of PB1, PB1-N40 [2,29,30]. Segment 3 also has a gene for protein PA-X that can originate from a frameshift [31]. This protein is responsible for virulence. The important viral protein hemagglutinin is situated on segment 4. Segments 5 a 6 encode nucleocapsid protein and neuraminidase, respectively. Segment 7 carries proteins M1 and M2 in particular: unspliced mRNA version is translated into M2 protein whereas spliced version encodes M1 protein. The last segment encodes nonstructural proteins 1 (NS1) and 2 (NS2) as well as nuclear export protein (NEP). The difference in influenza B virus is that segment 6 also encodes NB protein important for viral growth in addition to NA. Type B influenza virus has protein BM2 with similar function to M2 channel of influenza A virus. One of notable difference is protein NS1 which has similar but not the same functions in the prevention of host antiviral response as corresponding protein in influenza A viruses. Protein PB1-F2 has not been observed in types B or C. Virus C has been organized analogously to influenza viruses A and B with major divergence in HEF protein possessing the function of both NA and HA proteins.

All coding regions are flanked by regions that do not encode any proteins and these regions are highly conserved amongst all influenza A and B viruses. At each segment, termini are the same 13 nucleotides at the 5' end and 12 nucleotides at the 3' end and they are partially complementary. These regions play a crucial role in the specific packing of abovementioned eight segments into newly forming virions [2,32–34].

1.2.6. Viral proteins

Three proteins that so far received the most attention are hemagglutinin, neuraminidase, and M2 proton channel. These proteins play a crucial role in the viral life cycle and are the once most exploited as drug targets in influenza to date.

1.2.6.1. Hemagglutinin

Hemagglutinin is a surface glycoprotein that coats viral envelope with high abundance. Hemagglutinin forms homotrimer from identical subunits, each 540 – 550 amino acids long. Its polypeptide chain contains signal sequence guiding the protein to ER of host cells for posttranslational glycosylation. Each subunit is the product of

extracellular proteolytic cleavage of polypeptide precursor [35–37]. Its overall structure looks that N- as well as C- termini are stationed near the viral membrane. The outer part of HA extends 130 nm forming the structure of three subdomains: fusion subdomain, that is linked to central part from two antiparallel α-helices followed by vestigial esterase, and receptor binding site subdomains together forming globular structure at the far distant end from the viral membrane. Receptor binding site is formed by three conserved regions: active site consisting of amino acids Y98, W153, H183, and Y195 and two edge region elements with secondary structures 190 – α-helix, 130 – and 220 – loops [38,39].

The main function of HA is recognition of terminal N-acetyl neuraminic acid on cell-surface glycoproteins. Afterward, the virus can enter the cell via receptor-mediated endocytosis. This receptor binding is species specific. Avian and equine viruses preferably recognize sialic acid in α2,3-linkage to galactose while human prefer α2,6- linkage. Swine virus can recognize both and this fact is crucial for viral reassorting. Binding site can adopt various mutations potentially changing binding specificity and impacting virulence [2,40]. A secondary function of HA is membrane fusion during viral entry. During translation, HA is first translated into HA0 protein precursor. This precursor is further cleaved into two polypeptides HA1 and HA2 later forming a final trimer. The fusion peptide lies on HA2 upon acidification in endosome it repositions itself due to substantial structural changes in the whole HA and forms a bridge between viral and cellular membranes. This process leads to opening of endosome and release of vRNPs to the cytoplasm. HA protein sequence identity differs in subtypes from 40-70% and for the same subtypes 80-100%. HA can be phylogenetically and structurally divided into two separate groups. Up to date, only H1, H2, and H3 have been part of influenza viruses that caused human epidemics [2].

1.2.6.2. Neuraminidase

Neuraminidase is an enzyme with exosialidase activity (EC 3.2.1.18) cleaving α-ketosidic linkage between N-acetylneuraminic acid and adjacent galactose on cell surface glycoproteins. Neuraminidase subtypes are divided into two groups: first one consisting of N1, N4, N5 and N8 and the second one of N2, N3, N6, N7, and N9 [41,42]. One major difference in the first phylogenic group is that there is so-called "150 cavity" near the active site that possesses interesting properties for drug design. Homology of NA between strains of one subtype is 90% whereas homology of different subtypes is

approximately 50% and of influenza B is only around 30%. The polypeptide chain contains 470 amino acids. NA consists of 4 domains. The first is the cytoplasmatic domain followed by transmembrane domain anchored in the viral membrane. The transmembrane domain is then connected to globular domain bearing active site through "stem" domain. Whole NA is then formed as homotetramer resembling mushroom shape. The molecular weight of the whole tetramer is approximately 240 kDa. These tetramers can form clusters occupying viral envelope [41,43–45]. Globular domain carrying active site also called "head" is situated at far-end of protein. It binds calcium ion under the catalytic site that stabilizes the structure of this enzyme at low endosomal pH. The active site is formed through amino acids with Arg118, Asp151, Arg152, Arg224, Gln276, Arg292, Arg371, and Tyr406 and they are strictly conserved in all influenza A and B viruses. Secondary amino acids important for their structure role (Q119, Arg156, Trp178, Ser179, Asp198, Ile222, Gln227, Gln277, Asp293, and Gln425) may be subject to mutations [2,46]. Other conserved residues are few asparagines undergoing glycosylation modifications as well as several cysteines responsible for correct folding and maintenance of three-dimensional structure [43].

There have been numbers of proposed functions of NA, one of which is degradation of host secretary mucins during the approach of the virus to targeted cells. Second is probable function during fusion of viral and host membranes [47,48]. However, the most important role is during budding of newly formed viral particles when NA cleaves sialic acids surrounding new virus promoting viral release from the host cell membrane.

1.2.6.3. M2 proton channel

M2 proton channel is homotetramer of 97 amino acid-long protein of which 23 AAs form ectodomain, 19 AAs are part of the transmembrane region and last 54 AAs form the cytoplasmatic tail. Homotetramer is either linked through disulfide bridges or a pair of disulfide-linked dimers. Residues of the cytoplasmatic part near membrane have cholesterol binding function [49,50]. During the endosomal stage of the viral cycle, M2 channel conducts protons from cytoplasm inwards to virion core acidifying endosome and facilitating viral uncoating. This channel is highly selective for protons mainly due to histidine residues that facilitate hydrogen bonding. This net of bonds collectively forms water wiring that serves as a scaffold for hydrogen passage through channel [51,52].

1.2.7. Viral life cycle

Influenza viral life cycle can be divided into 5 stages: viral entry into the host cell, transport of vRNPs into the nucleus, transcription, and replication, vRNPs export from the nucleus and finally viral assembly and budding.
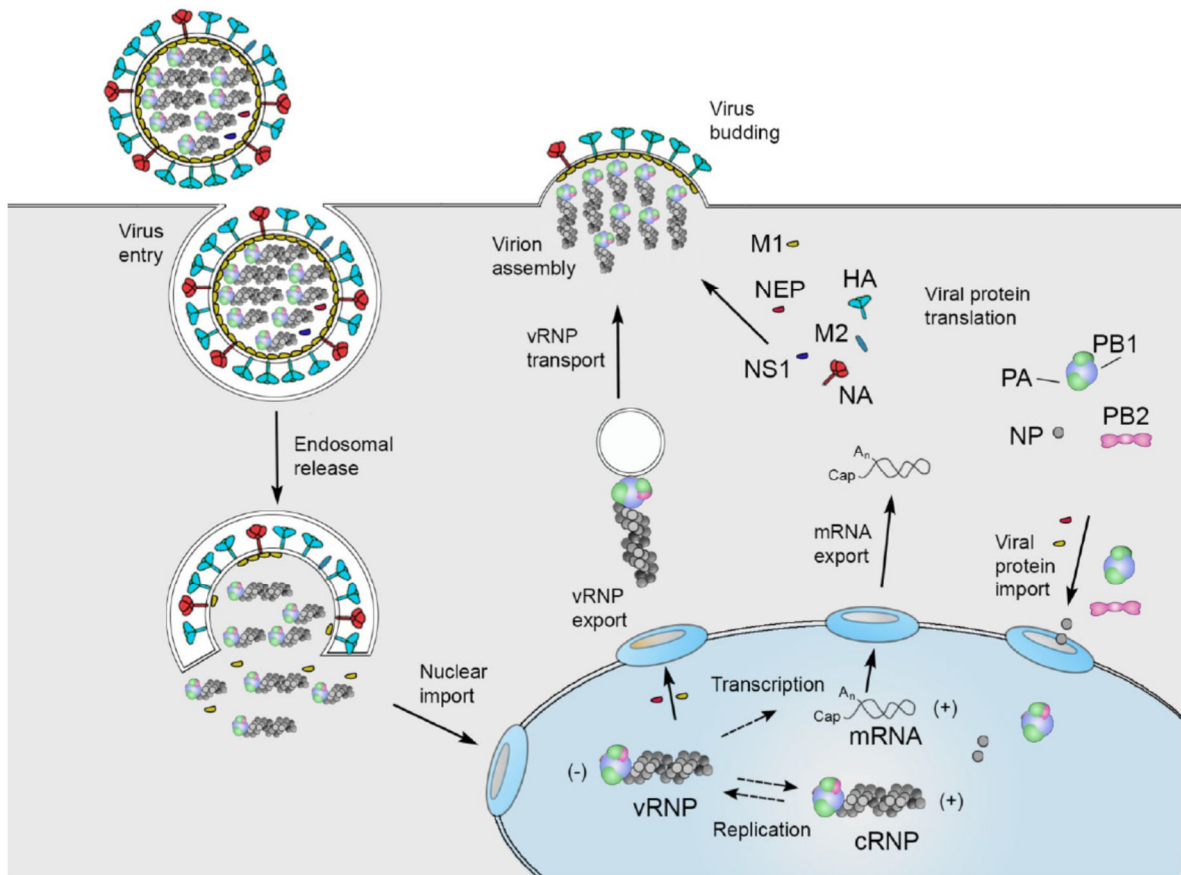
1.2.7.1. Host cell entry

As mentioned earlier, viral life cycle starts with recognition of terminal α-sialic groups of the host cells glycoproteins by viral protein hemagglutinin. This protein specifically recognizes terminal α-sialic acid on sphingolipids of host cells. Binding is multivalent as multiple copies of an HA1 subdomain of HA bind different sialic acids presented on surface glycoprotein and upon that virus enters the host cell via endosome. This endosome is guided towards nuclear membrane where it reaches the late endosomal stage. During this stage, endosome retains low pH 5. The low pH triggers substantial

conformational changes in HA and fusion peptide of HA2 subdomain repositions itself and inserts itself into the endosomal membrane which is then drawn near the viral membrane [54,55]. This starts hemifusion of these two membranes creating fusion pore. Furthermore, pore enlarges enabling vRNPs to exit viral core into the cytoplasm [56,57]. The release of vRNPs is also dependent on prior acidification of viral core. That is attained through M2 channel which pumps proton ions from intermediate endosomal layer to the viral core. Low pH allows vRNPs to release from M1 matrix which they are otherwise tightly bound to[58].

1.2.7.2. Transcription and replication

The distinctive feature of influenza viruses is that they replicate as well as transcribe their RNA in the cell nucleus. Transcription begins with 10 – 13 capped nucleotide primer which is "stolen" by polymerase from the host pre-mRNAs. mRNA cleavage starts with binding of host pre-mRNA to PB1 subunit which is then cut by N-terminal endonuclease domain of PA subunit. This 5' capped primer further binds to PB2 subunit. During the described process, the viral polymerase is associated with human RNA polymerase II responsible for the synthesis of these capped host pre-mRNAs. This interconnection mediates continuous access of primers during transcription [59–66].

Both vRNA ends are paired between themselves and bind to PB1 subunit where 3' end serves as a template for transcription. Primer then anneals to vRNA template initiating transcription elongation until polymerase reaches specific U rich region where it stutters due to steric hindrance and starts polyadenylation of 3' end resulting in generation nascent viral mRNA [53].

Replication starts with 3' vRNA end bound in the active site of PB1 subunit where dinucleotide of pppApG is formed and complement to starting dinucleotide of 3' vRNA. Elongation proceeds in a similar manner to transcription until termination where the template is released from binding pocket resulting in the formation of cRNA. cRNA is then assembled into premature RNP complex via associating with NPs recruited by the newly synthesized viral polymerase. During this step two, cRNA has to be used in the genesis of new vRNAs. At this stage, initiation starts with the production of same pppApG dinucleotide as in stage one, differentiating in initiation starting on dinucleotide in position 4-5 on cRNA. Elongation and termination are subject to current research.

Several models on how final vRNP complex is formed during creation of vRNA were proposed, none of them yet to be confirmed [2,53,67].

Replication, as well as transcription, are regulated by various factors. Proteins included in vRNPs are self-regulating factors as they are crucial for cRNA synthesis. It has been observed that vRNA of NP and NS1 proteins are preferably synthesized leading to the early synthesis of these two proteins. In contrast, M1 protein is synthesized as the last one. Regulatory mechanism for this time-dependent protein synthesis is still unclear. Viral mRNA level is concomitantly increased with vRNA synthesis for the first 4 hours of virus infection and then decreases whereas vRNA synthesis continues with high rate [2,53,67,68].

1.2.7.3. Transport, assembly, and budding

When all viral components are synthesized they have to be assembled into complete viral particles. Some components are transported towards plasmatic membrane separately and some pre-assemble into sub-complexes. Although assembly and budding are continuous processes, the release of an incompletely packaged viral particles can occur resulting in noninfectious virions. However, all requirements for successful budding are still not known. Nonetheless, some studies propose preferential budding of progeny infectious viral particles [26].

Viral components or pre-assembled complexes have to possess ability to reach viral membrane. The budding site is a location within apical plasmatic membrane in polarized epithelial cells. Proteins like HA, NA, and M2 contain apical determinants. During their synthesis, they undergo posttranslational modifications and are guided firstly through endoplasmatic reticulum and secondly through Golgi apparatus (see Figure 4). Then they associate with CPM via the apical determinants. These determinants in HA as well as in NA are tracked to glycan ectodomain and to the transmembrane domain (TMD). TMDs interact with lipid microdomains and can associate with lipid rafts. The main component of new viral particles are ribonucleoprotein particles [2,26,69,70].

As mentioned earlier, vRNAs are replicated in cell nucleus hence they have to be first exported into the cytoplasm. vRNPs are first formed inside nucleus from heterotrimeric polymerase, vRNA, and nucleocapsid protein. Then they form complex with M1 protein that is subsequently associating with NEP protein. Complete export

mechanism of vRNPs is still unclear. One hypothesis is that this big complex of NEP-M1-vRNP is further interacting with NS2 protein as well as with host factor RanGTP which enables vRNP nuclear export mediated by cellular export receptor Crm1 [71–75]. Then they have to be transported towards the apical side of the cell membrane. There are also two hypotheses on the mechanism by which is that accessed. First one is that M1-vRNPs can be transported via piggy-back interaction during exocytic pathway of HA and NA. The second hypothesis came from a publication that showed that NP, as well as vRNA, interacts with microfilaments of cellular cytoskeleton thereby possibly making it conveyor of transportation [76,77].

Considering all these pieces of information, M1 protein acts as a key player during viral assembly and budding. M1 interacts with each other forming multimers. This protein forms bridge between viral envelope through binding to cytoplasmatic parts of HA, NA, and M2 and viral core since it also binds to vRNPs. It is unclear whether M1 protein or envelope proteins select assembly and budding site although it is believed that HA plays a key role. When all components assemble at the apical side of the membrane, the initiation of budding is started by bending of the membrane from planar structure to a curved one. NA, HA, M1, and M2 are thought to be responsible for this initiation since they all interact with the inner leaflet of the lipid bilayer. Multimeric clustering of M1 protein then probably bends membrane leading to a final fusion and pinching-off and thus the creation of newly formed viral particles. M1 protein is also responsible for the exclusion of host factors during budding [26,78,79]. Final step during budding includes release from the outer layer of the cellular membrane. Progeny viral particles are after membrane fission still attached to sialic receptors that reside on the host cell membrane. NA than terminates viral life cycle by cleaving these sialic acids as well as it also cleaves all sialic acids present on viral particles, preventing viral aggregation [43].

All assembly and budding process are still not very well understood. Presumably, several unknown host factors may be of great importance. Studies have shown that viral particles can be formed in absence of either vRNPs, HA or NA confirming the essential role of M1 protein. One key question is whether all eight RNA segments are packed randomly or in an ordered fashion. Specific packing model is slightly favored since vRNA segments are in the equimolar ratios within viral particles [26,67,70].
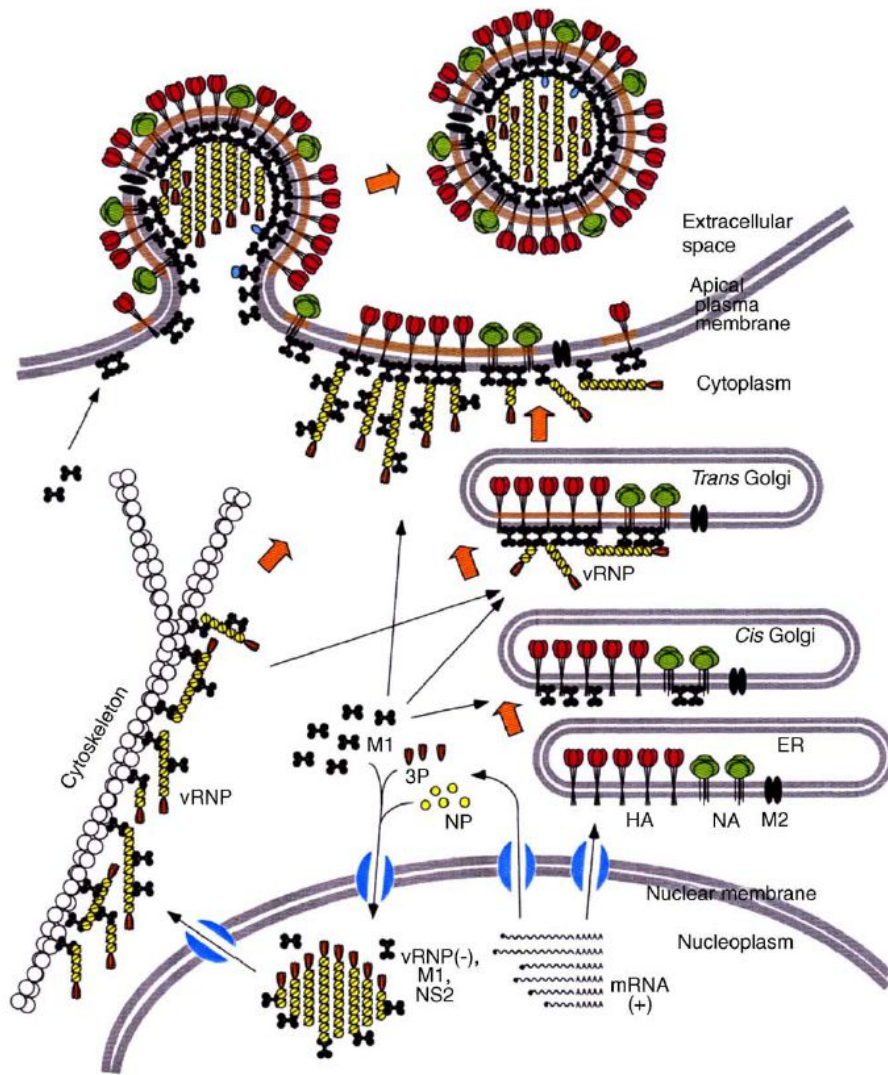
Figure 4: Scheme of viral assembly and budding. Reproduced from [26].

1.3. Influenza treatment

1.3.1. Antivirals

1.3.1.1. M2 channel inhibitors

Historically, the first influenza antiviral agent was developed serendipitously in the 1960s, twenty years before the discovery of M2 proton channel function. This drug, amantadine, was licensed in 1966 under trade name Symmetrel. Later in the former Soviet Union, the α-methylamine derivative of amantadine named rimantadine (Flumadine) was developed and afterward licensed in many countries [80]. Mechanism of inhibition was later explained. These antivirals were only successful targeting influenza A viruses. During viral endocytosis, M2 channel is responsible for acidification of virus core mediating dissociation of M1-vRNP from endosome. Hereafter, its inhibition results in the inability of vRNPs to reach cell nucleus for transcription/replication. The secondary M2 proton channel function is that during transport of HA, M2 through Golgi (see section. 1.2.7.3) M2 channel acidifies trans stage of Golgi apparatus enabling HA to travel to the cytoplasmatic membrane. When Golgi is not acidified due to inhibition with adamantanes HA cannot potently reach cytoplasmatic membrane [2,80,81].

Adamantanes bind inside the N-terminal part of channel interacting with residues 27-34. Today, these antivirals are notoriously famous for their rapid resistance emergence. The basis of resistance to adamantanes is nowadays very well established. It is caused by a single amino acid mutation in either of residues 26,27,30,31 or 34. Resistance first emerged during the 1980s and spread into various influenza subtypes. Today, most influenza A strains already poses this dominant phenotype making these drugs practically unavailable. Until this day, adamantane antivirals are still present on market but their usage is highly unadvised by WHO [80,82,83].

1.3.1.2. Neuraminidase inhibitors

Studies that revealed three-dimensional structure of neuraminidase in 1980´ shed light on a new approach for influenza treatment. Knowing that NA plays a crucial role during influenza viral cycle, it was considered as a valuable drug target. Antivirals were developed as derivatives of the transition state of N-acetylneuraminic acid (NANA) during cleavage with NA [84,85]. Inhibitor development first tried to capitalize on the closest analog to transition state of neuraminic acid bound to NA. This first analog was 2-deoxy-2,3-dihydro-N-acetylneuraminic acid (DANA). During development of these

derivatives, further X-ray structures with higher resolution were determined and provided knowledge for efficient drug design. The guanidinyl group was introduced into the core structure leading to the formation of additional interactions with enzymatically crucial residues (Glu119, Asp151, Glu227) and afterward improving binding affinity 1,000-fold compared to the NANA and 10,000-fold compared to DANA [46,86,87]. Another important factor was selectivity of the compound against NA compared to cellular sialic acid receptors. This derivative called zanamivir (Figure 5 on page 28) with the guanidinyl group showed high selectivity towards NA and low binding affinity to endogenous human sialidases. It was further developed as an inhaled drug by GlaxoSmithKline and later approved by FDA in 1999 under trade name Relenza as first drug targeting influenza neuraminidase [84,88].

After the discovery of zanamivir, broad structure-activity relationship studies on sialic acid derivatives were undertaken. Considerable effort was put towards the replacement of carbohydrate core with hexane as well as pentane inside the core of derivatives. Additional pentyl ether moiety that targets small hydrophobic pocket inside NA active site was perceived as an enhancement. Moreover improving lipophilicity while maintaining inhibitory activity was another key step in development for potential drug oral availability. Bioavailability was later supported by the development of ethyl ester prodrug later converted into an active form *in vivo*. Finally, all led to the development of a drug called oseltamivir by Gilead Sciences that was approved in 1999. Today it is marketed under tradename Tamiflu by Roche [84,89].
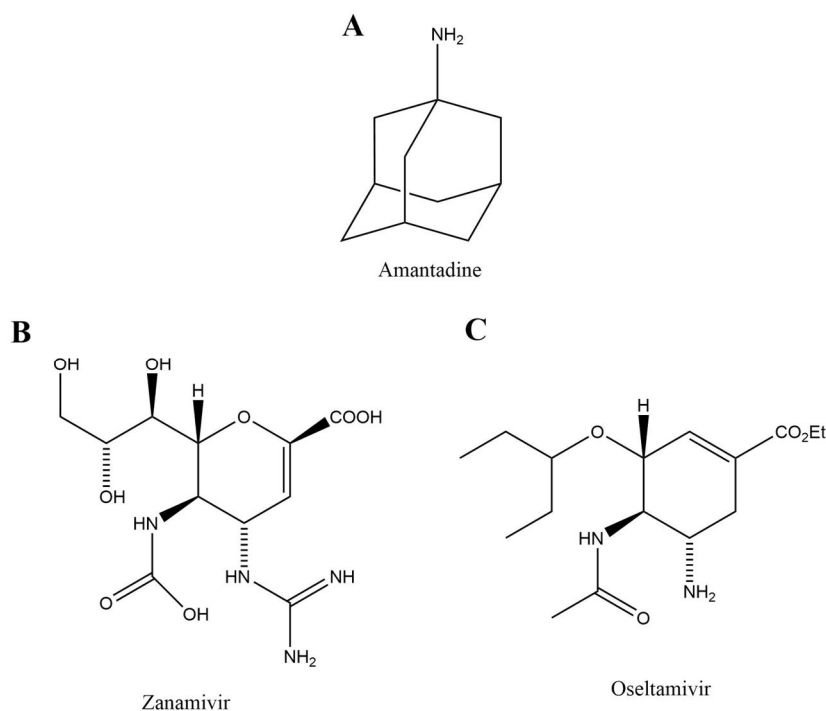
Figure 5: Structures of the inhibitors targeting influenza virus. A) Amantadine, an inhibitor of the M2 proton channel, formerly used drug in influenza treatment. B, C) Currently worldwide used inhibitors of influenza neuraminidase zanamivir and oseltamivir. Structures were created in software ChemDraw.

Zanamivir, as well as oseltamivir, were both found to be highly potent neuraminidase inhibitors. They also proved their prophylactic as well as therapeutic properties. When administered in early infection (within 48 hours) they significantly decreased illness duration [2,88,90]. However, zanamivir pharmacokinetic characteristics are poor due to its polarity. The drug is rapidly excreted and has very limited oral availability and thus it is prescripted mostly in severe cases as a secondary option when oseltamivir treatment fails.

As well as M2 channel inhibitors, zanamivir and oseltamivir select resistant variants. Since zanamivir resembles NANA transitional state, resistance pressure has never been seen as much as for influenza resistance to adamantanes. Resistance to zanamivir is mostly selected through mutations in HA reducing receptor binding affinity. On the other hand, oseltamivir is gradually more known for the resistance development. Some mutations such as Glu119Val by themselves confer oseltamivir resistance and it was also observed that this mutation does not impair viral transmissibility. The major

mutation causing oseltamivir resistance is substitution His274Tyr in NA. This substitution has only minor effect on zanamivir susceptibility in contrast to high resistance to oseltamivir. Later during 2009 pandemic, this key mutation was also accompanied by secondary mutations with a synergistic effect. Secondary mutations mainly consist of substitutions at positions 222, 246, 198 etc. Resistance primarily develops at a higher rate in immunocompromised patients. Mutations usually impair NA activity as well as viral fitness or transmissibility, however, they usually dominantly select through the development of compensatory mutations in either NA or HA enhancing their functions. After 2009, pandemic emerging mutations very closely monitored. Emergence and transmission of resistant viruses are today under close surveillance and it illustrates the possible ineffectiveness of oseltamivir treatment [2,91–95].

1.3.1.3. Other drug targets

Occurring resistance against traditional influenza treatment raised great interest in search of new drug targets for influenza treatment. Some studies tried to exploit influenza hemagglutinin as a possible target. Although there are 18 subtypes of influenza A virus with general sequence diversity, some conserved regions can be found. Researchers tried to isolate neutralizing antibodies from patients plasma samples. This approach could potentially allow influenza antibody-based treatment. Few antibodies recognizing HA epitope conserved in influenza A and B types were found and are currently subjects for clinical testing [96–99]. Other approaches exploit HA-mediated entry, with one compound in clinical trials, or transport of HA towards host cell membrane [100,101].

Another interest was directed towards Ribavirin (see Figure 7 on page 31), a broad-spectrum antiviral agent. Its activity against orthomyxoviruses (influenza) and paramyxoviruses (respiratory syncytial virus, measles) were previously shown. This compound was also FDA approved for treatment of hepatitis C virus as combination therapy with PEGylated interferon-α. Its significance is mainly due to very low resistance development since it acts against an IMP dehydrogenase, key enzyme in the biosynthesis of GTP: [88,102].

Another notable concept employs reducing viral production by silencing viral genes by small interfering RNA (siRNA). These small RNA sequences target conserved regions in nucleocapsid protein or in the segment coding polymerase upon inward cell

delivery via polycation carriers. They have been tested on animal models with promising results so far [103,104].

Last but not least approach focuses on neuraminidase protein fusion to the cell-surface-anchoring sequence. This fusion construct utilizes derivative of *Actinomyces viscosus* neuraminidase which when applied topically to airway epithelium cleaves sialic acids from the cell receptors preventing virus cell entry. A similar effect of blocking viral entry has been showed for sialic-acid polyacrylamide conjugates. Both approaches enregistered positive result especially in disease prophylaxis on animal models [105,106].

All above-mentioned drug targets or treatment options are currently subjects of a basic research study or clinical testing. Nevetherless all have disadvantages in either drug delivery, target specificity etc. Hereafter, the most promising target as of late is viral polymerase. Its essentiality during viral life cycle in replication/transcription stage directed researchers to search for possible drug targets within the enzyme. Recently solved crystal structures of whole heterotrimeric polymerase bound to viral promotor (Figure 6) revealed inter-subunit interactions as well as RNA binding site [99,107].
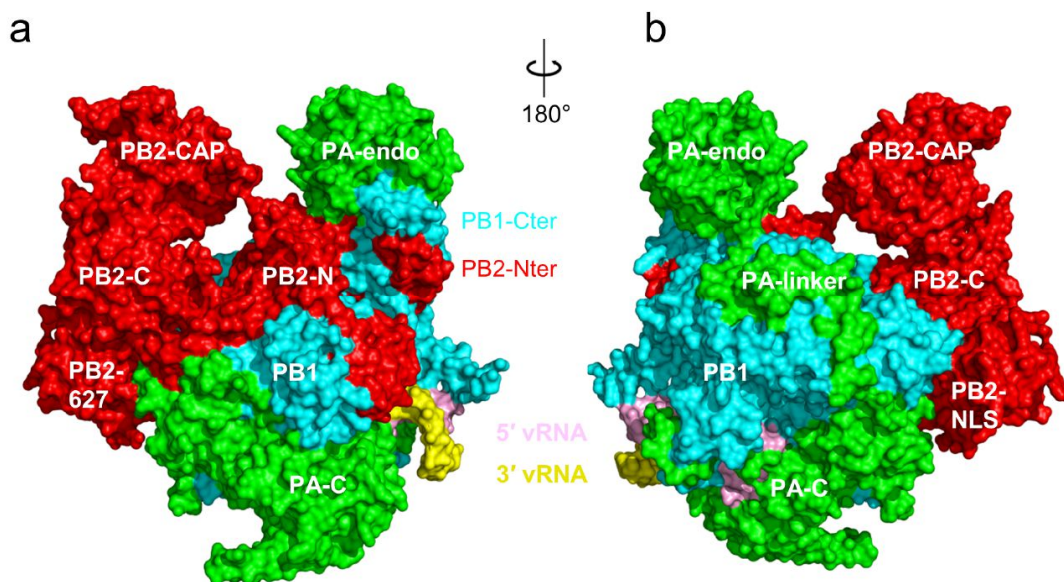
Figure 6: Space-filling representation of bat influenza A polymerase bound to viral promotor (180 degrees rotation across vertical axis). Each subunit is colored (PA green, PB1 teal, PB2 red). All protein domains and parts are correspondingly depicted (C/N-terminal parts, NLS (stands for nuclear localization signal), endonuclease and cap binding domains). Reproduced from [107].

One of the first recognized inhibitors targeting polymerase was compound T-705 (Favipiravir, see Figure 7), firstly reported in 2002. Since then, this compound was extensively tested and it seems to have significant potential. This compound acts as nucleoside mimetic (ATP/GDP) after its endogenous conversion to favipiravir-ribosyl-5′-triphosphate and it directly inhibits viral replication/transcription. However, this compound activation is a slow process and has relatively low antiviral potency. On the other hand, its advantage is that it is not a subject of high resistance development. Later in 2014, favipiravir was approved for influenza treatment in Japan and is in clinical testing in both USA and Europe [108,109].

Favipiravir
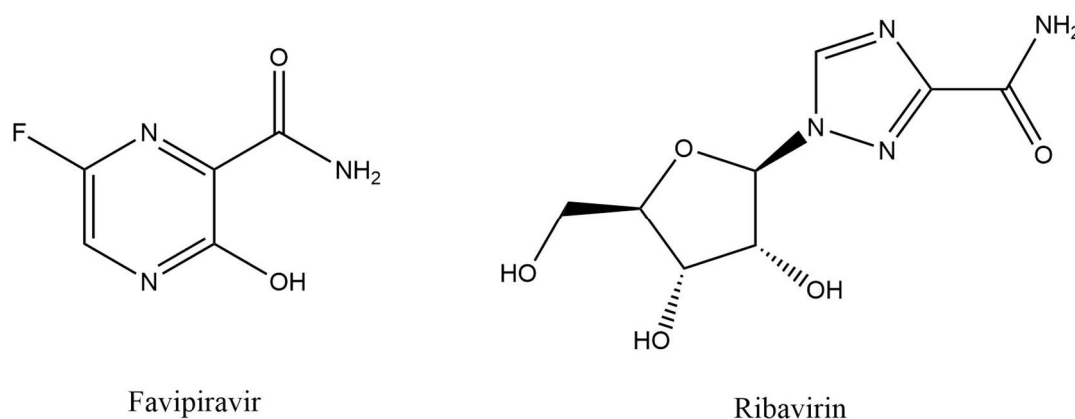
Ribavirin

Figure 7: Structural formulas of favipiravir and ribavirin, two compounds with the possibility of future utilization in influenza treatment. Structures created in ChemDraw.

Other potential targets within influenza polymerase were later discovered. One is a cap-binding domain in PB2 subunit. At first it was assumed that this target would not be therapeutically relevant considering it binds eukaryotic capped RNA proposing poor inhibitor specificity. Nevertheless, few compounds were found to inhibit the isolated PB2 cap-binding domain. However, they proved not to be active against virus probably due to poor cellular uptake considering their high negative charge [110]. Later the inhibitor VX-787 was developed through iterative synthesis and so far it is the best anti-PB2 compound. It is as azaindole derivative occupying cap-binding pocket and is nanomolar inhibitor with high antiviral activity and is currently in phase three of clinical trials [111,112].

Another interesting potential target is the N-terminal endonuclease domain in PA subunit. The main advantage of this target is the conservation of catalytic amino acids. The main strategy of targeting this endonuclease activity utilizes metal chelating agents

binding to two divalent ions located in the active center. Many inhibitors failed to prove their effectiveness of cellular uptake owing to their negative charge nature. In 1994 one compound, L-420,001, was found to possess anti-influenza activity[113]. It was later confirmed to be specific influenza endonuclease inhibitor[114]. As of late February 2018, the new drug, baloxavir marboxil (Xofluza) from Japanese pharmaceutical company Shionogi & Company was approved for sale in Japan and is still currently in late stage of clinical trials in the United States [65,99].

Another interest was focused on DNA aptamers that would bind to PA endonuclease domain. Few were found to be potent inhibitors nevertheless their pharmacokinetic/dynamics still needs to be significantly improved [115]. Last but not least target was found in the protein-protein interaction between two polymerase subunits PA and PB1.

1.4. Protein-protein interaction between PA and PB1 subunits of influenza polymerase

1.4.1. PPIs (Protein-protein interactions) classification

Protein-protein interactions form a very complex interplay network of various proteins. They play an extremely significant role in cellular signaling and multiple other biological pathways. However, they are yet to be broadly exploited in drug discovery in comparison to more traditional targets such as enzymes. Generally, targeting PPIs is considered as very challenging.

PPIs differ in their structural elements and can be classified accordingly (graphical representation on the next page) [116,117]:

- Two globular proteins interacting through discontinuous epitope and not undergoing conformational changes upon formation (**a**)
- Two globular proteins with a structural interface formed upon binding (**b**)
- Globular protein interacting with short single peptide chain (**c** and **d**)

- Two peptide chains interacting with each other (**e**)
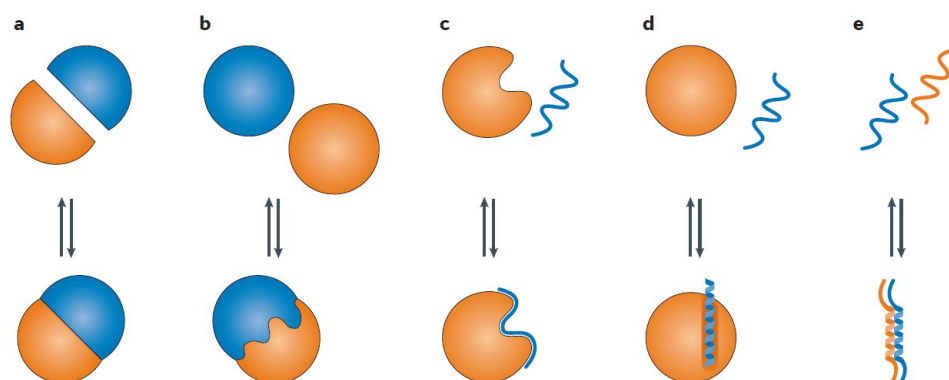
Figure 8: Classes of PPIs. Reproduced from [116].

All above mentioned are the interactions with an interface that usually does not have evenly distributed affinity driving forces across its surface. Some residues forming this interface are called "hot spots" and have a major contribution to binding affinity. Identifying these hot spots can be achieved through alanine scanning where interacting residues are mutated to alanines and change in binding affinity is observed and thus the severity of the change highlights the residue importance [116,118,119]. Another important fact is that there is no simple correlation between the size of the interacting surface area and binding affinity. Interactions can also be either obligate (long lasting with high affinity) or non-obligate (transient with low affinity) and affinity can span across six orders of magnitude from picomolar to high micromolar. Interface usually consists of hydrophobic patches surrounded by polar residues. However, a mixture of polar and hydrophobic residues was observed and water molecules can be part of them too [117,120,121].

Four PPI classes differ in their "druggability". First one with two globular proteins interacting through the large surface is hardly characterizable and usually undruggable. The second one poses difficulties since the interface is formed upon binding (Figure 8 b). That means that there is not stably formed binding pocket for the inhibitor, making this class also extremely challenging mainly during rational drug design. However, the interaction between globular protein and single chain peptide (Figure 8 c and d) can be an interesting drug target. These peptides can be relatively short making interface surface volume small and thus be possibly targeted with small compounds. Several studies tried

to target such interactions implementing various approaches [116,122]. Generally, high-throughput screening for PPIs results in low hit rates as well as greater number of false positives [117]. One very interesting approach is the development of peptidomimetics, organic compounds that structurally mimic peptide secondary structures. They can be either peptide or non-peptide-based. Few studies designed scaffold mimicking three recognized motifs in PPIs: α-helices, β-strands and reverse turns. In non-amino acid approach scaffolds such as terphenyl or polyamide were synthesized so they mimic α-helix representing consecutive amino acids (i, i+3 or i+4 and i+7) residues. This scaffold can be derived to occupy three-dimensional space similarly to corresponding amino acids side chains. Nonetheless, these molecules feature fairly simple structural elements and thus can result in bad selectivity [116,123–125].

Considering that in recent years we are almost out off enzymatic targets, protein-protein interactions have become of high interest in drug discovery. Modern methods in drug discovery such as computational modeling can assist in drug target evaluation. It should be mentioned here that relatively few targets have reached clinic testing. One notable fact is that the compounds targeting PPIs are usually incompatible with Lipinski's rules. However, recent advancements in drug delivery methods make essentiality of these rules questionable [116].

1.4.2. PA-PB1 interaction

The basis of this interaction was first discovered in 1995. Group of scientists from the University of Nebraska carried out a study about PB1 subunit deletion and identified that first 48 amino acids are sufficient to bind C-terminal part of PA subunit (CPA) [126]. Six years later, the same group published a study that focused on the mutational analysis of the first 16 residues of N-PB1. They mutated each residue to aspartic acid and looked for influenza virus recovery. They identified that first 12 AA constitute the interaction core and that 5 AA (Pro5, Leu7, Leu8, Phe9, and Leu10) are crucial as their mutation results in complete abolishment of polymerase activity [127]. Later in 2007, peptides derived from the first 25 amino acids of N-PB1 were shown as the inhibitors of the polymerase activity and viral spread [128]. The major impact for further research was as a result of two independent studies in 2008 that successfully obtained crystal structures of the complex of the N-terminal peptide and C-terminal domain of PA of influenza A polymerase. They revealed the 3D structure of the binding interface (Figure 9) and

identified interacting residues in CPA. C-terminal domain of PA resembles dragon´s head with PB1 peptide bound inside its jaws with N-terminus pointing out in back of mouth and C-terminus facing outwards.

Figure 9: Ribbon representation of the interaction between the C-terminal domain of PA subunit (Green) and N-terminal PB1 peptide (Red). Peptide binds inside hydrophobic cavity of CPA. Three pictures show a different point of view across the vertical axis. Picture was created in software PyMol using structure 3CM8 from PDB [129].

Seven AAs starting with Pro5 and ending with Lys11 form a short helix. In PA interacting amino acids are located in four helices forming hydrophobic cavity-like structure. Front-end residues Asp2, Asn4 interact with Ile621 and Glu623 within CPA as shown in Figure 10 on next page. Oxygen atoms in AAs Asp2, Val3, Phe9, Leu10 and Val12 form hydrogen bonds to CPA residues Gln408, Glu623, Gln670, Arg673, and Trp706. Furthermore, hydrogen bonds formed by backbone nitrogen atoms of Asp2, Val3, Asn4, Leu8, and Ala14 to Asn412, Pro620, Ile621, Glu623, and Gln670 also highly contribute to binding. Pro5 is one of key interacting AAs since it is part of a hydrophobic patch of residues Phe411, Trp706. In C-PA, Met595, Trp619, V636, and Leu640 form contacts with Leu8 [130,131].
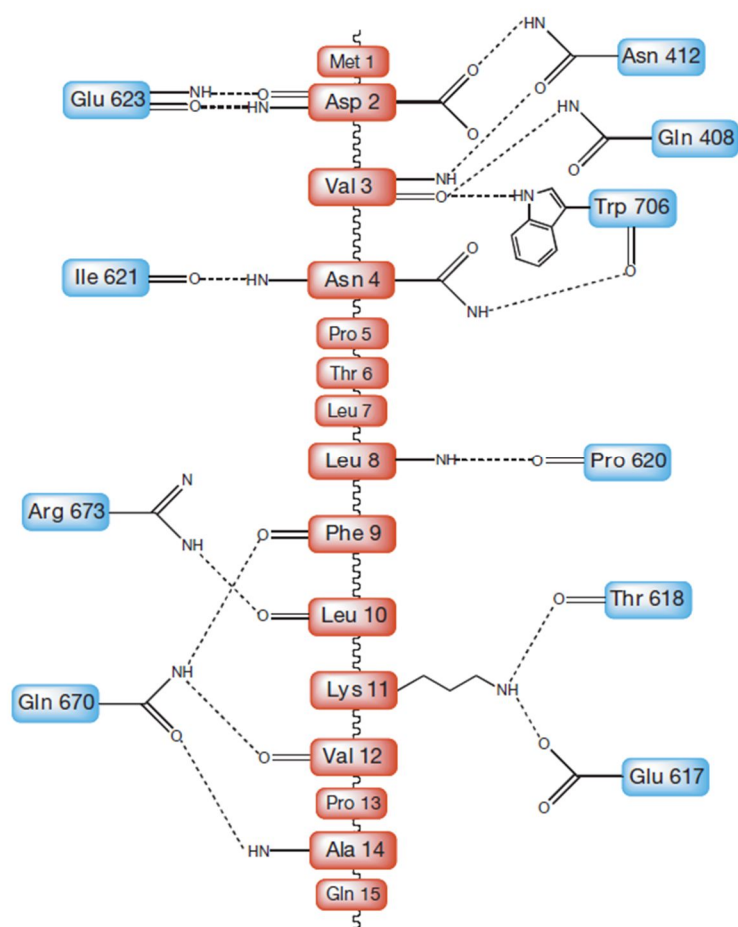
Figure 10: Graphical representation of interacting amino acids of N-PB1 and CPA. Red squares correspond to PB1 AAs and Blue to CPA. Reproduced from [130].

Crystal structure of apo CPA revealed only small structural difference compared to both structures of N-PB1 and CPA complexes as well as decreased stability determined by thermal shift assay. Stability of apo-CPA was decreased approximately by 18 °C, from 56°C for N-PB1/C-PA complex to 38°C for apo-CPA [132].

Based on all previously mentioned knowledge, this interaction has been evaluated as a possible drug target. The whole binding interface is relatively small compared to other PPIs [133]. Another study focused on mutational analysis when first 15 AAs were mutated to every other AA, resulted in a broad panel of possible AA substitutions further elaborating each residue importance [134,135].

The first search for small inhibitors was presented in 2012 as *in silico* screen. Compound showing hits from this screen were later tested in cell cultures and ELISA assays where one compound showed as promising as it inhibited the interaction with an IC_{50} value around 30µM, supported by data displaying no significant cell cytotoxicity [136]. Afterwards, few studies introduced similar results in compounds with comparable affinities [137,138]. Primary results from these studies were used later in the rational development of compounds with hopes of characteristics improvement. Several studies optimized previously reported compounds but without significant enhancement in binding affinities [139–142]. One study introduced the best compound so far with IC_{50} value in ELISA assay on the boundary of the micromolar/nanomolar range (Figure 11). However, this compound had a maximal solubility in the same range as IC_{50} value and thus supplementary characteristics such as cytotoxicity (100-fold higher) do not look convincing. So far no study showed any structural confirmation of identified inhibitors binding into CPA [143].
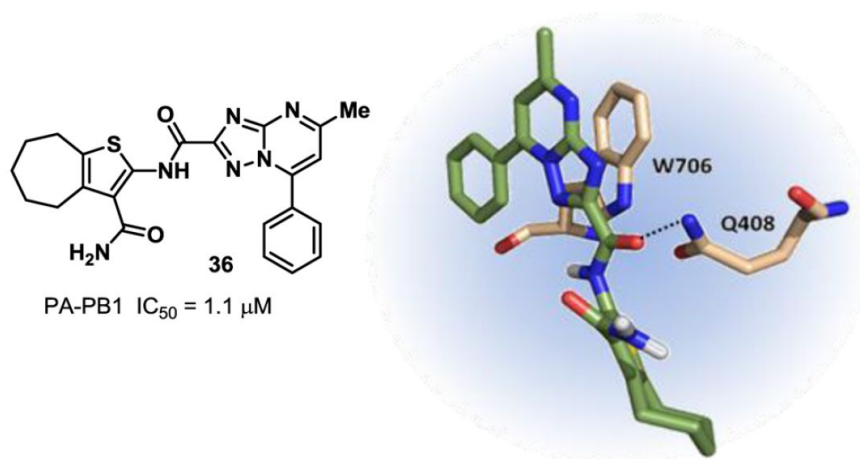
Figure 11: Promising compound targeting PA-PB1 interaction. On the left structural formula. On the right possible interaction scheme generated from docking studies of compound inside binding cavity surrounded by two interacting amino acids within PA subunit. Reproduced from [143].

To sum up, all research done on PA-PB1 interaction up to this date still has not identified relevant and potent inhibitors. Further research, as well as a search for novel compounds targeting N-PB1/PA interaction, is still of high importance.

2. Aims of the thesis

- Design of recombinant protein constructs of influenza polymerase subunits for high-throughput screening as well as for potential crystallization experiments
- Purification of recombinant proteins to desirable purity
- Development of high-throughput screening method for discovery of new inhibitors of PPI between influenza subunits PA and PB1
- Identification of the minimal PB1 peptide and structural characterization of its binding to CPA
- Crystallization of CPA complexed with newly identified inhibitory compounds from HTS for evaluation of binding interface

3. Materials and methods

3.1. Material, chemicals, instruments

3.1.1. Material

- Filter unit Sterivex GP 0.22 μm, Millipore (USA)
- Dialysis membrane Spectrapore (6,000-8,000 MWCO), Thermo Fisher Scientific (USA)
- Nunc 96-well plate transparent, Thermo Fisher Scientific (USA)
- Amicon Centrifugal filter units (MWCO 10,000, 30,000) 0.5, 4, 15 ml, Millipore (USA)
- 3 Lens Low Profile Plate, Swissci (Switzerland)
- EasyXtall 15-well plate, QIAGEN (Germany)
- FrameStar® 96 Well Semi-Skirted PCR Plate, 4titude (UK)
- HiLoad Superdex 75 16/600 column, GE Healthcare (USA)

3.1.2. Biological/special material

- 3 Lens Low Profile Plate, Swissci (Switzerland)
- *Escherichia coli,* Top10 strain, Novagen (USA)
- *Escherichia coli,* BL21 (DE3) RIL strain, Novagen (USA)
- ULP1 Protease, prepared by Michal Svoboda from IOCB of the CAS
- NeutrAvidin, Thermo Fisher Scientific (USA)
- GST protein, Sigma-Aldrich, (USA)
- U-35 antiHis i-Body, prepared by Vladimír Šubr from IMC of the CAS
- pGEX1-λT vector, obtained from Hans-Georg Kräusslich from the laboratory at the University of Heidelberg, Germany (originally from GE Healthcare Life Sciences)
- pETM11-SUMO3 vector, obtained from Dmytro Yushchenko Ph.D. from IOCB of the CAS (originally created at Protein Expression and Purification Core Facility, EMBL)

3.1.3. Instruments

- Centrifuges
   - Sorvall Evolution RC, Thermo Fisher Scientific (USA)
   - Allegra X-15R, Beckman Coulter (USA)
   - Microcentrifuge 5415R, Eppendorf (Germany)
- Rotary shaker Innova 44, Eppendorf (Germany)

- Trio 48 Thermocycler, Biometra (Germany)

- ThermoCell Mixing block. BIOER (China)

- Vertical Apparatus - PAGE, Bio-Rad (USA)

- Horizontal electrophoresis apparatus, Gibco (USA)

- Spectrophotometer, Specord 210, ChromSpec (Czech Republic)

- Microscope Olympus SZX10, Olympus Corporation (Japan)

- pH meter model pH 50, XS Instruments (Italy)

- Camera Olympus E-620, Olympus Corporation (Japan)

- Voltage source EPS 301, GE Healthcare (USA)

- Digital tube Roller SRT6D, Stuart (UK)

- Incubator IPP 400, Memmert (USA)

- Infinite Reader M1000 PRO, Tecan (Switzerland)

- EmulsiFlex-C3, Avestin (Canada)

- BlueWasher, Blue Cat Bio (Germany)

- Nanodrop ND-1000 spectrophotometer, Thermo Fisher Scientific (USA)

- Crystal Gryphon, Art Robbins Instruments (USA)

- Oryx 8, Douglas Instruments (UK)

- Xtal UV Detector, Rigaku (Japan)

- Gallery DT plate hotel, Rigaku (Japan)

- Laser-Spectroscatter 201, RiNA Netzwark RNA Technologien GmbH, (Germany)

- LightCycler®480 Instrument II, Roche Life Science (Germany)

- EnSpire Multimode Plate Reader, PerkinElmer (USA)

- IPC High Prescission Multimode Dispenser, Ismatec (Germany)

- Äkta Purifier 10, GE Healthcare Life Sciences (USA)

- Fraction Collector Frac-950, GE Healthcare Life Sciences (USA)

- MP-500V power supply, Major Science (USA)

- UV lamp UVT - 20 S/M/L, Herolab (Germany)

- SPR sensor platform developed at Institute of Photonics and Electronics, Prague [144]

3.1.4. Chemicals

- Biacore (Sweden)

  1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride (EDC), N-Hydroxysuccinimide (NHS), ethanolamine

- Biotika (Slovak Republic)

  ampicillin

- Hampton research (USA)

  silicone oil

- Lach-Ner (Czech Republic)

  ethanol, isopropanol, sacharose

- Sigma-Aldrich (USA)

  SDS, TEMED, APS, 2-mercaptoethanol, bromphenol blue, HEPES, acrylamide, dimethyl sulfoxide, PEG 3350, glycine, EDTA, kanamycin sulphate, TCEP, d-desthiobiotin, L-glutathione reduced (GSH), sodium phosphate dibasic heptahydrate, potassium dihydrogen phosphate dihydrate, LB Broth, imidazole, CBB-R250, sodium ascorbate, N, N´ - methylene bisacrylamide

- Penta (Czech Republic)

  glycerol, acetic acid, sodium chloride, hydrochloric acid, sodium acetate

- Prochimia (Poland)

  $HS-(CH_2)11-EG_4-OH$, $HS-(CH_2)_{11}-EG_6-OCH_2-COOH$

- Promega (USA)

  Tris base

- SDT (Germany)

  casein buffer 20×-4× concentrate

- Serva (Germany)

  agarose

- Thermo Fisher Scientific (USA)

  IPTG

- USB (USA)

  Tween 20

3.1.5. Others

- PPP Master Mix, Top-Bio (Czech Republic)

- Phusion® High Fidelity DNA Polymerase, New England Biolabs (UK)

- 5x Phusion HF Buffer, New England Biolabs (UK)

- Restriction enzyme NdeI, New England Biolabs (UK)

- Restriction enzyme BamHI, New England Biolabs (UK)

- Restriction enzyme XhoI, New England Biolabs (UK)

- Restriction enzyme NDEI, New England Biolabs (UK)

- CutSmart® buffer, New England Biolabs (UK)

- T4 DNA Ligase, New England Biolabs (UK)

- 10x T4 DNA Ligase Reaction buffer, New England Biolabs (UK)

- PCR dNTP Mix, Top-Bio, (Czech Republic)

- Zyppy Plasmid Miniprep Kit, Zymo Research (USA)

- QIAquick Gel Extraction Kit, Qiagen (Germany)

- Ni-NTA agarose, Qiagen (Germany)

- Crystallization Basic Kit for Proteins, Sigma Aldrich (USA)

- Crystallization Extension Kit for Proteins, Sigma Aldrich (USA)

- Morpheus® crystallization screen, Molecular Dimensions (UK)

- JBScreen JCSG++, Jena Biosciences (Germany)

- Index Hampton crystallization screen, Hampton Research (UK)

- GelRed Nucleic Acid Gel Stain, Biotum (USA)

- Pierce™ Glutathione Agarose, Thermo Fisher Scientific (USA)

- Protein Assay Dye Reagent Concentrate, Bio-Rad (USA)

- Sypro® Orange Protein Gel Stain, Sigma Aldrich (USA)

- Plasmon IV immersion oil, Gargille (UK)

- Alkaline Phosphatase, New England Biolabs (UK)

3.2. Molecular cloning

3.2.1. GST-CPA fusion recombinant construct for bacterial expression

The gene for influenza C-terminal domain of polymerase acidic subunit from viral strain A/California/07/2009 (H1N1) (Genebank accession number CY121685.1) was obtained from Jan Weber Ph.D. from the IOCB of the CAS by amplification from the virus. Construct for GST fusion protein was prepared by PCR amplification of segment

coding amino acid region 239 – 716 of CPA with additional restriction endonuclease sites NdeI (purple), BamHI (red) at both ends as well as two stop codons (grey) at 3´ terminus. Thrombin cleavage site (teal) was added to N-terminus of CPA for potential separation from GST fusion protein (was not put to use) as shown in Figure 12. Protein will be synthesized as a fusion protein of GST-CPA where GST protein would serve for purification purposes as well as in HTS AlphaScreen assay.

PCR was done in a two-step process. First PCR served for insertion of thrombin cleavage site as well as BamHI site and stop codons. Second PCR further elongated 5´ end with a nonspecific overhang for restriction cleavage as well as with NdeI site. Primers for amplification were following:

Forward primer 1: 5´CTGGTTCCGCGTGGGTCCGCTGCTAGGCTGTACTGCCAA

Forward primer 2: 5´ATATTTTCATATGCTGGTTCCGCGTGGGTCC

Reverse primer: 5´CACGATGGATCCTATCACTTCAGTGCATGTGTGAGGAAGGAG

```
   1 CATATGCTGG TTCCGCGTGG GTCCAACGGC TGCATTGAGG GCAAGCTTTC CCAAATGTCA
  61 AAAGAAGTGA ACGCCAAAAT TGAACCATTC TTGAGGACGA CACCACGCCC CCTCAGATTG
 121 CCTGATGGGC CTCTTTGCCA TCAGCGGTCA AAGTTCCTGC TGATGGATGC TCTGAAATTA
 181 AGTATTGAAG ACCCGAGTCA CGAGGGGGAG GGAATACCAC TATATGATGC AATCAAATGC
 241 ATGAAGACAT CTTTTGGCTG GAAAGAGCCT AACATAGTCA AACCACATGA GAAAGGCATA
 301 AATCCCAATT ACCTCATGGC TTGGAAGCAG GTGCTAGCAG AGCTACAGGA CATTGAAAAT
 361 GAAGAGAAGA TCCCAAGGAC AAAGAACATG AAGAGAACAA GCCAATTGAA GTGGGCACTC
 421 GGTGAAAATA TGGCACCAGA AAAAGTAGAC TTTGATGACT GCAAAGATGT TGGAGACCTT
 481 AAACAGTATG ACAGTGATGA GCCAGAGCCC AGATCTCTAG CAAGCTGGGT CCAAAATGAA
 541 TTCAATAAGG CATGTGAATT GACTGATTCA AGCTGGATAG AACTTGATGA AATAGGAGAA
 601 GATGTTGCCC CGATTGAACA TATCGCAAGC ATGAGGAGGA ACTATTTTAC AGCAGAAGTG
 661 TCCCACTGCA GGGCTACTGA ATACATAATG AAGGGAGTGT ACATAAATAC GGCCTTGCTC
 721 AATGCATCCT GTGCAGCCAT GGATGACTTT CAGCTGATCC CAATGATAAG CAAATGTAGG
 781 ACCAAAGAAG GAAGACGGAA AACAAACCTG TATGGGTTCA TTATAAAAGG AAGGTCTCAT
 841 TTGAGAAATG ATACTGATGT GGTGAACTTT GTAAGTATGG AGTTCTCACT CACTGACCCG
 901 AGACTGGAGC CACACAAATG GGAAAAATAC TGTGTTCTTG AAATAGGAGA CATGCTCTTG
 961 AGGACTGCGA TAGGCCAAGT GTCGAGGCCC ATGTTCCTAT ATGTGAGAAC CAATGGAACC
1021 TCCAAGATCA AGATGAAATG GGGCATGGAA ATGAGGCGCT GCCTTCTTCA GTCTCTTCAG
1081 CAGATTGAGA GCATGATTGA GGCCGAGTCT TCTGTCAAAG AGAAAGACAT GACCAAGGAA
1141 TTCTTTGAAA ACAAATCGGA AACATGGCCA ATCGGAGAGT CACCCAGGGG AGTGGAGGAA
1201 GGCTCTATTG GGAAAGTGTG CAGGACCTTA CTGGCAAAAT CTGTATTCAA CAGTCTATAT
1261 GCGTCTCCAC AACTTGAGGG GTTTTCGGCT GAATCTAGAA AATTGCTTCT CATTGTTCAG
1321 GCACTTAGGG ACAACCTGGA ACCTGGAACC TTCGATCTTG GGGGGCTATA TGAAGCAATC
1381 GAGGAGTGCC TGATTAATGA TCCCTGGGTT TTGCTTAATG CATCTTGGTT CAACTCCTTC
1441 CTCACACATG CACTGAAGTG ATAGGATCC
```

Figure 12: CPA$_{239-716}$ DNA segment for insertion into pGEX1-λT vector.

LVPRGSNGCIEGKLSQMSKEVNAKIEPFLRTTPRPLRLPDGPLCHQRSKFLLMDALKLSIEDPSHEG
EGIPLYDAIKCMKTFFGWKEPNIVKPHEKGINPNYLMAWKQVLAELQDIENEEKIPRTKNMKRTSQL
KWALGENMAPEKVDFDDCKDVGDLKQYDSDEPEPRSLASWVQNEFNKACELTDSSWIELDEIGEDVA
PIEHIASMRRNYFTAEVSHCRATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTKEGRRKTNL
YGFIIKGRSHLRNDTDVVNFVSMEFSLTDPRLEPHKWEKYCVLEIGDMLLRTAIGQVSRPMFLYVRT
NGTSKIKMKWGMEMRRCLLQSLQQIESMIEAESSVKEKDMTKEFFENKSETWPIGESPRGVEEGSIG
KVCRTLLAKSVFNSLYASPQLEGFSAESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVL
LNASWFNSFLTHALK

Figure 13: Protein sequence of CPA$_{239-716}$, GST protein will be synthesized at N-terminus of CPA$_{239-716}$ as it is built in used pGEX1-λT vector.
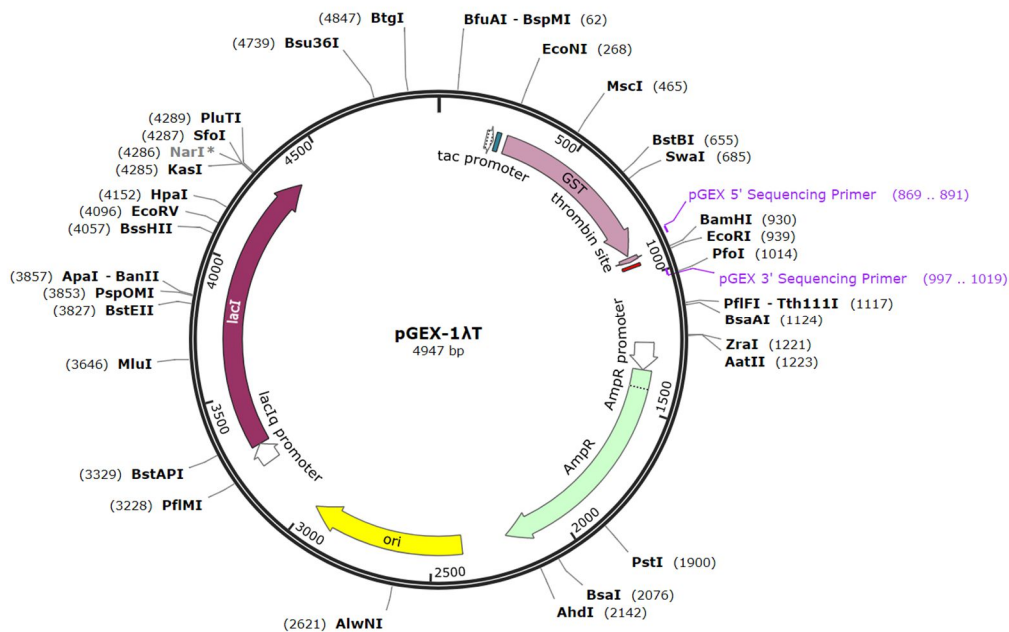
Figure 14: Scheme of vector pGEX-1λT used in the design of fusion protein construct of GST-CPA. Picture reproduced from [145].

The plasmid used for CPA gene insertion was commercially available plasmid pGEX1-λT coding Glutathione-S-Transferase (GST) previously modified (having extra NdeI restriction site preceding to BamHI site) and kindly provided to us from Hans-Georg Kräusslich of the laboratory at University of Heidelberg, Germany.

3.2.2. His$_6$-SUMO-CPA fusion recombinant construct for bacterial expression

The same template as in section 3.2.1. was used for the PCR amplification preparation of His$_6$-SUMO-CPA construct. PCR experiment was used to amplify a 257 – 716 amino acid segment of CPA. Amplified DNA contained restriction endonuclease sites BamHI (purple), XhoI (red) at 5′, 3′ ends respectively, as well as two stop codons (grey) at 3′ terminus (Figure 15). The construct was designed so that His$_6$ (hexahistidine) chelation tag for purification purposes accompanied with SUMO (small ubiquitin-like modifier) tag for improvement of protein solubility would be introduced at N-terminus of CPA protein as they are built in the used pETM11-SUMO3 vector. This SUMO protein can be also cleaved by ULP1 protease so the CPA protein could be later isolated from these tags and used for the crystallization experiments. Primers for amplification were following:

Forward primer:

5′ ATATTTTGGATCCATTGAACCATTCTTGAGGACGACACCAC

Reverse Primer:

5′ CACGATCTCGAGCTATCACTTCAGTGCATGTGTGAGGAAGGAGTT

```
   1 GGATCCATTG AACCATTCTT GAGGACGACA CCACGCCCCC TCAGATTGCC TGATGGGCCT
  61 CTTTGCCATC AGCGGTCAAA GTTCCTGCTG ATGGATGCTC TGAAATTAAG TATTGAAGAC
 121 CCGAGTCACG AGGGGGAGGG AATACCACTA TATGATGCAA TCAAATGCAT GAAGACATTC
 181 TTTGGCTGGA AAGAGCCTAA CATAGTCAAA CCACATGAGA AAGGCATAAA TCCCAATTAC
 241 CTCATGGCTT GGAAGCAGGT GCTAGCAGAG CTACAGGACA TTGAAAATGA AGAGAAGATC
 301 CCAAGGACAA AGAACATGAA GAGAACAAGC CAATTGAAGT GGGCACTCGG TGAAAATATG
 361 GCACCAGAAA AGTAGACTT TGATGACTGC AAAGATGTTG GAGACCTTAA ACAGTATGAC
 421 AGTGATGAGC CAGAGCCCAG ATCTCTAGCA AGCTGGGTCC AAAATGAATT CAATAAGGCA
 481 TGTGAATTGA CTGATTCAAG CTGGATAGAA CTTGATGAAA TAGGAGAAGA TGTTGCCCCG
 541 ATTGAACATA TCGCAAGCAT GAGGAGGAAC TATTTTACAG CAGAAGTGTC CCACTGCAGG
 601 GCTACTGAAT ACATAATGAA GGGAGTGTAC ATAAATACGG CCTTGCTCAA TGCATCCTGT
 661 GCAGCCATGG ATGACTTTCA GCTGATCCCA ATGATAAGCA AATGTAGGAC CAAAGAAGGA
 721 AGACGGAAAA CAAACCTGTA TGGGTTCATT ATAAAAGGAA GGTCTCATTT GAGAAATGAT
 781 ACTGATGTGG TGAACTTTGT AAGTATGGAG TTCTCACTCA CTGACCCGAG ACTGGAGCCA
 841 CACAAATGGG AAAAATACTG TGTTCTTGAA ATAGGAGACA TGCTCTTGAG GACTGCGATA
 901 GGCCAAGTGT CGAGGCCCAT GTTCCTATAT GTGAGAACCA ATGGAACCTC CAAGATCAAG
 961 ATGAAATGGG GCATGGAAAT GAGGCGCTGC CTTCTTCAGT CTCTTCAGCA GATTGAGAGC
1021 ATGATTGAGG CCGAGTCTTC TGTCAAAGAG AAAGACATGA CCAAGGAATT CTTTGAAAAC
1081 AAATCGGAAA CATGGCCAAT CGGAGAGTCA CCCAGGGGAG TGGAGGAAGG CTCTATTGGG
1141 AAAGTGTGCA GGACCTTACT GGCAAAATCT GTATTCAACA GTCTATATGC GTCTCCACAA
1201 CTTGAGGGGT TTTCGGCTGA ATCTAGAAAA TTGCTTCTCA TTGTTCAGGC ACTTAGGGAC
1261 AACCTGGAAC CTGGAACCTT CGATCTTGGG GGGCTATATG AAGCAATCGA GGAGTGCCTG
1321 ATTAATGATC CCTGGGTTTT GCTTAATGCA TCTTGGTTCA ACTCCTTCCT CACACATGCA
1381 CTGAAGTGAT GACTCGAG
```

Figure 15: CPA$_{257-716}$ DNA segment for insertion into pETM11-SUMO3 vector.

```
IEPFLRTTPRPLRLPDGPLCHQRSKFLLMDALKLSIEDPSHEGEGIPLYDAIKCMKTFFGWKEPNI
VKPHEKGINPNYLMAWKQVLAELQDIENEEKIPRTKNMKRTSQLKWALGENMAPEKVDFDDCKDVG
DLKQYDSDEPEPRSLASWVQNEFNKACELTDSSWIELDEIGEDVAPIEHIASMRRNYFTAEVSHCR
ATEYIMKGVYINTALLNASCAAMDDFQLIPMISKCRTKEGRRKTNLYGFIIKGRSHLRNDTDVVNF
VSMEFSLTDPRLEPHKWEKYCVLEIGDMLLRTAIGQVSRPMFLYVRTNGTSKIKMKWGMEMRRCLL
QSLQQIESMIEAESSVKEKDMTKEFFENKSETWPIGESPRGVEEGSIGKVCRTLLAKSVFNSLYAS
PQLEGFSAESRKLLLIVQALRDNLEPGTFDLGGLYEAIEECLINDPWVLLNASWFNSFLTHALK
```

Figure 16: His$_6$-SUMO protein will be synthesized at N-terminus of CPA$_{257-716}$ as it is built in the used pETM11-SUMO3 vector.

Plasmid for expression of His$_6$-SUMO-CPA protein was plasmid pETM11-SUMO3 GFP (Figure 17) obtained from Dmytro Yushchenko Ph.D. from IOCB of the CAS. GFP protein was not utilized in our case as it was cleaved out during plasmid endonuclease restriction cleavage (section 3.2.6.).
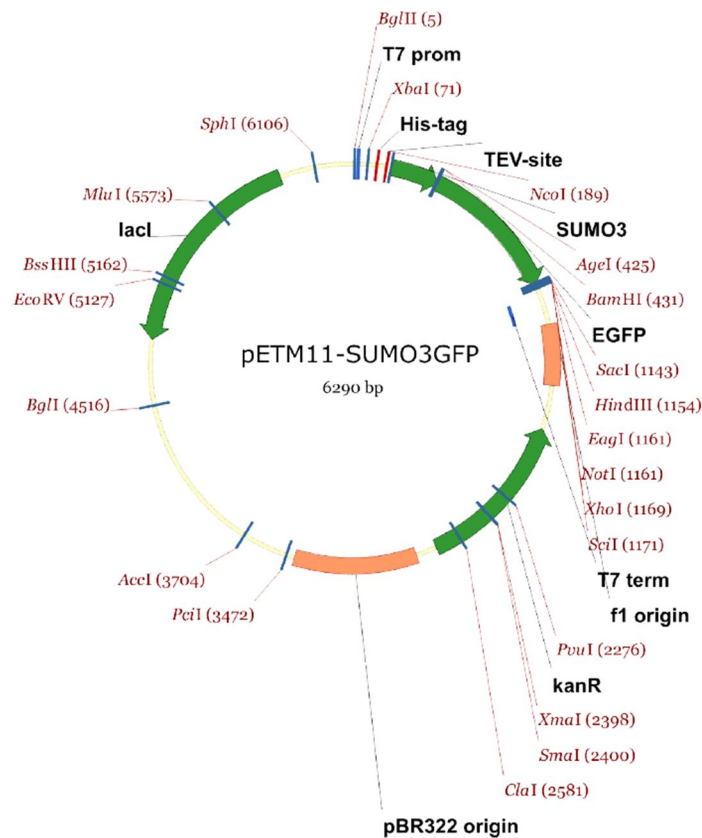
Figure 17: Scheme of the vector pETM11-SUMO3 GFP for design of His$_6$-SUMO-CPA fusion recombinant protein. Reproduced from [146].

46

3.2.3. Polymerase chain reaction

Two polymerase chain reactions setups were done utilizing two different polymerases. Phusion® High-Fidelity DNA polymerase (New England Biolabs® INC) was used for initial gene amplification. PPP Master Mix (Top-Bio) was used for testing of positively ligated bacterial clones. The general protocol for amplification of DNA by Phusion polymerase was done as follows:

Table 1: Preparation of PCR reaction solution (per one reaction)

Component	Reaction volume	Final concentration
Phusion HF 5x buffer	10 µl	
10 mM dNTPs	1 µl	200 µM
20 µM Forward Primer	1.25 µl	0.5 µM
20 µM Reverse Primer	1.25 µl	0.5 µM
Template DNA	0.5 – 3 µl	< 250 ng
DMSO	1.5	
Phusion DNA polymerase	0.5	1 U
miliQH$_2$O	to 50 µl	

Table 2: PCR amplifiation protocol

Step	Temperature	Time
DNA denaturation	98 °C	30 sec
30 cycles of amplification	98 °C	10 sec
	42 – 72 °C	30 sec
	72 °C	30 sec per kb
Final extension	72 °C	8 min

Testing of clones for determination of positive DNA gene integration into plasmid was done as described below:

Table 3: Preparation of PCR reaction sample per 1 colony

Component	Reaction volume
PPP master mix	5 μl
miliQH$_2$O	4.5 μl
20 μM Forward Primer	0.25 μl
20 μM Reverse Primer	0.25 μl

Table 4: PCR amplification protocol for testing of positively ligated clones

Step	Temperature	Time
DNA denaturation	94 °C	60 sec
	94 °C	15 sec
30 cycles of amplification	55 – 68 °C	15 sec
	72 °C	60 sec per kb
Final extension	72 °C	8 min

3.2.4. Agarose gel electrophoresis

- TAE Buffer: 40 mM Tris/Acetate, pH 8.3, 1 mM EDTA
- 6x Sample Buffer: 10% Sacharose (w/v), 0.1% bromphenol blue

40 ml of 1% agarose (w/v) dissolved in TAE buffer were mixed with 4 μl of the 10,000x GelRed solution (Biotum). The gel was left to solidify for 20 minutes after which it was transferred inside the electrophoretic cage and poured over with TAE buffer. DNA samples were premixed with 6x Sample buffer and loaded into gell wells. Voltage was set to 120 V and time was adjusted variably for optimal DNA fragment separation.

3.2.5. Agarose gel DNA extraction

After finishing electrophoresis, gels were transferred inside UV box for visualization of DNA. Appropriate DNA fragments were excised by scalpel and placed inside microtubes. Each gel fragment was then mixed with 700 μl of QG buffer (Qiagen Gel extraction kit) and incubated at 50°C for complete gell dissolution. The solution was then applied to DNA centrifugal filter units (Qiagen) and centrifuged (13,000 g, RT, 1min). The column was washed with 700 μl of PE buffer. Finaly, 40 μl of miliQ H$_2$O

was applied to the column and incubated at RT for 5 minutes to dissolve DNA. DNA sample was recovered after centrifugation (13,000 g, RT, 1min).

3.2.6. Restriction endonuclease treatment of DNA

PCR amplified products, isolated from agarose gel after electrophoresis, as well as commercial plasmids, were treated with restriction endonucleases. For GST-CPA construct both pGEX-1λT plasmid and PCR amplification product were double-digested with endonucleases as follows:

Table 5: Preparation of reagents for restriction reaction

Component	Reaction volume
Cutsmart Buffer	5 μl
DNA [ul]	1 – 40 μl
NdeI	1 μl
BamHI	1 μl
miliQH$_2$O	to 50 μl

For His$_6$-SUMO-CPA both pETM11-SUMO3 GFP plasmid and PCR amplification product were double-digested with endonucleases as follows:

Table 6: Preparation of reagents for restriction reaction

Component	Reaction volume
Cutsmart Buffer	5 μl
DNA [ul]	1 – 40 μl
BamHI	1 μl
XHOI	1 μl
miliQH$_2$O	to 50 μl

All reactions were incubated for 2 hours at 37 °C.

After successful endonuclease treatment plasmids (pGEX-1λT, pETM11-SUMO3 GFP) were dephosphorylated at their 5′, 3′ ends by alkaline phosphatase. At the end of incubation 1μl of alkaline phosphatase was added to reaction tubes containing cleaved plasmids. Reactions were further incubated for 20 minutes at RT for dephosphorylation. Agarose gel electrophoresis isolation was then performed to purify desired fragments.

3.2.7. Ligation of CPA DNA into plasmids

- 10x T4 DNA Ligase buffer (New England Biolabs)

After purification of DNA fragments of each plasmid and CPA were these fragments ligated together using T4 DNA Ligase. The proper amount of DNA was counted using NEBioCalculator tool [147] to provide 3-fold molar excess of CPA over plasmid. Reactions were then mixed accordingly:

Table 7 : Preparation of ligation reaction

Component	Reaction volume	Final mass
10 x T4 ligase buffer	4 µl	
CPA DNA	5 – 30 µl	75 – 85 ng
Plasmid DNA	0.5 – 2 µl	100 ng
T4 DNA Ligase	1 µl	
miliQH$_2$O	to 40 µl	

Reactions were then incubated at 16 °C for 4 hours followed by chemical transformation of bacterial Top10 *E.Coli* cells.

3.2.8. DNA plasmid chemical transformation to competent bacterial cells

Sixty µl of bacterial cell suspension (*Escherichia Coli*-Top10) was gently mixed with 8 ul of final ligation mixture and left on ice for 10 minutes. Microtube was transferred to heat block and incubated at 42 °C for 90 seconds followed by immediate transfer back to the ice for another 5 minutes. In a final step, 250 µl of LB media was added to the bacterial suspension and the sample was incubated at 37 °C for 60 minutes. After that, 250 µl of the suspension was spread onto agar plates containing antibiotics. After 16 – 20 hours, resistant bacterial colonies were tested by insert amplification (colony PCR) and positive ones were selected.

3.2.9. Minipreparation of plasmid DNA

10 ml of LB medium in 50 ml falcon tube containing antibiotic was inoculated with the preselected bacterial colony. The inoculum was incubated for 14 – 18 hours at 37 °C and afterward spun down (3,000 g, 4°C, 15 minutes). The supernatant was carefully decanted and pellets were resuspended in 600 µl of sterile water. Preparation

of DNA was conducted using Zyppy^{TM} plasmid miniprep kit from Zymo Research. Bacterial pellet resuspended in 600 µl of sterile water was transferred in microtube and 100 µl of 7X lysis buffer was added and mixed by repeated inverting. Within 2 minutes, 350 µl of prechilled neutralization buffer was added and rapidly inverted in hand for few times. The sample was then spun down (14,000 g, 20°C, 15 min) after which 900 µl of supernatant was carefully transferred into Zymo-Spin^{TM} IIN column and placed inside collection tube. Quick centrifugation (13,000 g, RT, 60sec) was executed and flow-through was discarded. Furthermore, the column was washed with 200 µl of Endo-Wash Buffer and spun down (13,000 g, RT, 60sec) followed by a wash with 400 µl of Zyppy^{TM} Wash Buffer and spun down at (13,000 g, RT, 60sec). The column was transferred to clean microtube and incubated with 40 µl of sterile water for 5 minutes. Final centrifugation was carried out and DNA sample was recovered. The concentration of DNA was measured using Nanodrop device (Thermo Fisher Scientific).

3.3. Bacterial expression of recombinant proteins

- Lysis buffer (GST-CPA): 25 mM Tris/HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA
- Lysis buffer (His$_6$-SUMO-CPA): 50 mM Tris/HCl, pH 8.0, 200 mM NaCl, 10 mM imidazole

DNA from positively selected clones was sent for Sanger sequencing (GATC- Biotech) to further evaluate correct DNA gene sequence within selected clones. 1 – 3 µl (200 – 600 ng) of selected miniprep DNA was chemically transformed (see section 3.2.8) into 60 µl of *Escherichia coli* BL21 (DE3) RIL strain suspension. Two agar plates with grown colonies were gently washed with 6 ml of LB media after 16 – 20 hours incubation to make inoculum for large-scale expression. 3 liters of LB media containing corresponding antibiotics were divided into 6 2-liter Erlenmeyer flasks covered with chemical pulp and each flask was inoculated with 1 ml of prepared inoculum. Flasks were transferred inside preheated (37 °C) rotary shaker (Eppendorf) and shook to the speed of 220 RPM. Expression process and growth curve were observed by measuring optical density (OD) at 595 nm at periodical intervals. The temperature was gradually decreased to a final temperature of 18 °C). When OD$_{595}$ reached approximately 0.8, 1 ml of 375 mM IPTG was added to each flask (final 0.75 mM) to induce protein expression. Flasks were then shaken at 18 °C for 16 – 18 hours.

Bacterial suspensions were spun down (4,000 g, 4°C, 20 minutes). Pellets were then resuspended in 150/130 ml (GST-CPA/His$_6$-SUMO-CPA) of corresponding lysis buffer and frozen at -20°C.

3.4. Protein purification

3.4.1. Bacterial cell lysis

Cell suspension was thawed and then homogenized on ice using hand glass homogenizer. Homogenate was strained through gaze. The suspension was repeatedly (3 times) run through the Emulsiflex device (Avestin) at pressure 1,200 bar for cell lysis. The lysate was spun down (35,000 g, 4°C, 40 min) and the supernatant was harvested by very careful decantation.

3.4.2. Affinity purification – Chelation chromatography

3.4.2.1. Purification of GST-CPA on the Glutathione-agarose resin

- Wash buffer: 50 mM Tris/HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA
- Elution buffer: 50 mM Tris/HCl, pH 7.5, 150 mM NaCl,10 mM reduced L-glutathione (GSH), 1 mM EDTA

Firstly, glutathione-agarose had to be equilibrated. Four ml of 50% ethanol suspension of glutathione-agarose was spun down (600 g, 4°C, 2min) and the supernatant was discarded. The resin was then washed twice with 10 ml of wash buffer to wash out all ethanol. The resin was then mixed with lysate supernatant after centrifugation, divided into 50 ml falcon tubes and incubated at 4°C on a rotator for 90 minutes. Resin suspension was spun down (600 g, 4°C, 5min) and the supernatant was recovered (flow-through fraction). The resin was then washed twice with 25 ml of wash buffer. Protein was lastly eluted with 25 ml of elution buffer in 3 steps (the first elution step took 10 minutes of incubation). First and second elution were then pulled together and concentrated to 10 ml using 15 ml Amicon centrifugal filter units (MWCO 30 kDa).

3.4.2.2. Purification of His$_6$-SUMO-CPA on Ni-NTA agarose resin

- Wash buffer: 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole
- Elution buffer: 50 mM Tris/HCl, pH 8.0, 200 mM NaCl, 250 mM imidazole
- Dialysis buffer: 50 mM Tris/HCl, pH 8.0, 200 mM NaCl, 1 mM TCEP, 1 mM EDTA

All steps were equivalent to protocol in section 3.4.2.1. except that the resin used in this case was Ni-NTA agarose. The protocol was divergent in that two resin wash steps were proceded each with 10 ml (instead of 25) and elutions were done with 6 ml (instead of 25) of elution buffer. As the last step the first and second elution were pulled together and dialyzed in dialysis tube overnight in dialysis buffer.

3.4.2.3. Cleavage of chelation/solubility tag from His-SUMO-CPA construct

- recombinant ULP1 protease (400 µg/ml)

After dialysis the recombinant protein was proteolytically cleaved by ULP1 protease (kindly provided by Michal Svoboda from the IOCB of the CAS) for removal of His_6-SUMO tag. Later 63 mg of protein was mixed with 1.6 ml of ULP1 protease for weight mass ratio 1:100 (protein : protease) and left incubating at 4°C for 8 hours. After cleavage, the chelation chromatography was performed in the same manner as in section 3.4.2.2 with one exception that the fraction containing CPA was flow-through (fraction without His_6-SUMO).

3.4.3. Gel permeation chromatography

- Buffer GST-CPA: 50 mM Tris/HCl, pH 7.4, 150 mM NaCl
- Buffer CPA: 50 mM Tris pH 8.0, 200 mM NaCl, 1 mM EDTA, 1 mM TCEP

GPC was used for further purification of both GST-CPA protein and for CPA after His_6-SUMO tag cleavage. In both cases, Superdex75 column (GE Healthcare) was connected to Äkta purifier (GE Healthcare), equilibrated with corresponding buffers and preincubated in the cooling box at 16 °C. Proteins were divided into fractions of 5 ml (diluted to 1 mg/ml) and filtered before load. Proteins were loaded into 5 ml loop and the flow rate was set to 0.3 ml/min. Chromatogram displaying UV (280 nm) absorption was recorded and 5/3 ml fractions (GST-CPA/CPA) were sampled by the automatic sampler. The process of GPC was analyzed by SDS-PAGE and required fractions were diluted to desired concentrations and frozen either in -80°C freezer (proteins for HTS) or in liquid nitrogen (aliquots for protein crystallography experiments).

3.5. SDS-PAGE

- Sample buffer: (6 × concentrate): 350 mM Tris/HCl, pH 6.8, 30% (v/v) glycerol, 350 mM SDS, 4% (v/v) 2 mM 2-mercaptoethanol, 180 μM bromphenol blue

- Electrode buffer: (5 × concentrate): 140 mM Tris/HCl, pH 8.8, 1.4 M glycine, 20 mM SDS

- 14% stacking gel: 375 mM Tris/HCl, pH 8.8, 14% acrylamide, 0.1% (w/v) SDS, 0.2% (v/v) TEMED, 0.1 % (w/v) APS

- 6% resolving gel: 250 mM Tris/HCl, pH 6.8, 6.6% acrylamide, 0.1% (w/v) SDS, 0.2% (v/v) TEMED, 0.1% (w/v) APS

SDS-PAGE was used to monitor whole purification process. For this purpose, 100 μl samples were pulled from every fraction during purification. Samples were mixed with sample buffer and heated to 96 °C for 10 minutes for protein denaturation. The pre-prepared gel was placed inside Bio-Rad apparature and poured over with electrode buffer. After cooling, samples and protein standard (All blue marker) were loaded into gel wells. The cooling block was placed inside apparatus as well to prevent gel overheating. The device was closed and electrodes were tuck inside voltage source. The program was set to 180V and gel was run for approximately 90 minutes. Sample dye was observed in time periods and the program was stopped 10 minutes after dye exited a lower end of the gel. The gel was removed from the glass frame and placed inside the plastic box for CBB staining:

- Staining buffer: 0.5% Coomassie Brilliant Blue (w/v), 50% methanol (v/v), 10% acetic acid (v/v)

- Washing buffer: 10% acetic acid

The gel was stained for 10 minutes and then washed repeatedly with washing buffer until protein bands were clearly visible. Afterwards, gels were placed inside plastic foil and scanned.

3.6. Protein concentration determination – Bradford assay

160 μl of protein samples, as well as standards for calibration, were pipetted to 96 well plate. Next 40 μl of Protein Assay Dye reagent (Bio-Rad) was added to each well

and mixed thoroughly. The plate was then incubated for 5 minutes at room temperature after which plate was placed inside reader M1000 Pro (Tecan). The program was set to measure the absorbance at 562 nm for each well. Concentration was counted from a calibration curve equation obtained by extrapolation of absorbance to the concentration of calibration samples.

3.7. Surface Plasmon Resonance (SPR)

- Protein buffer: 25 mM Tris/HCl, pH 7.4, 150 mM NaCl (0.05 – 0.1% Tween20)
- Alkanethiole solution: pure ethanol mixture of HS-$(CH_2)_{11}$-PEG_4-OH and HS-$(CH_2)_{11}$-PEG_6-O-CH_2-$COOH$ alkanethiols (molar ratio 7:3, Prochimia) with a final concentration of 0.2 mM
- SA10 buffer (10 mM sodium acetate, pH 5.0)
- Ethanolamine (EA) 1 M solution
- PBS buffer: 10 mM Na_2HPO_4, 1,8 mM KH_2PO_4, pH 7.4, 500 mM NaCl, 2.7 mM KCl
- NHS/EDC water solution: 100 mM N-hydroxysuccinimide, 400 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride
- Neutravidin 20 mg/l in SA10 buffer

The day before the experiment, the golden chip was washed with water then ethanol and thoroughly dried with gaseous nitrogen. Chip was placed inside glass tube containing an alkenethiole solution and incubated for 60 minutes at 37°C. Than it was removed from the incubator and placed into the dark compartment and left coating overnight. Protein samples were diluted to the desired concentration and dialyzed against protein buffer overnight.

Next day, SPR apparatus was turned on (peristaltic pump, temperating container, light source). SPR chip was washed with ethanol, then water and dried with gaseous nitrogen. A drop of immersion oil was applied to chip surface and the chip was then placed inside holding cage. The system was filled with water set to constant flow-rate of 30 µl/min and the reference and dark spectra were obtained. Record of measurement was started and the first solution applied to the system was NHS/EDC for activation of carboxylic terminal groups. After approximately 3 minutes, system was switched to

water to wash for another 3 minutes. Neutravidin solution was applied next for 7 minutes. Then 5 washing steps (each for approximately 3 minutes) followed in this order: SA10, PBS, SA10, EA, SA10. High ionic strength PBS buffer served to get rid of unspecifically bound neutravidin whereas EA deactivated residual carboxylic groups. The thickness of neutravidin layer was measured and the system was switched to protein buffer. From this step, channels differed in content as different controls were tested. For the test of GST-CPA binding, the biotinylated $PB1_{1-25}$ (MDVNPTLLFLKIPAQNAISTTFPYTGGSK-Biotin) peptide was bound to neutravidin layer. After peptide saturation, in the last step, GST-CPA protein (or individual GST as control) sample was introduced to system and growth of layer size was observed (k_{on}). For K_d determination was system switched to protein buffer to observe protein washing out (k_{off}). Curves were imported into TraceDrawer software (Ridgeview instruments AB) for fitting by 1:1 model (1 protein : 1 ligand) and obtaining of k_{on}, k_{off} parameters. K_d was calculated as follows:

$$K_d = \frac{k_{off}}{k_{on}}$$

- k_{on} is the association rate constant
- k_{off} is the dissociation rate constant
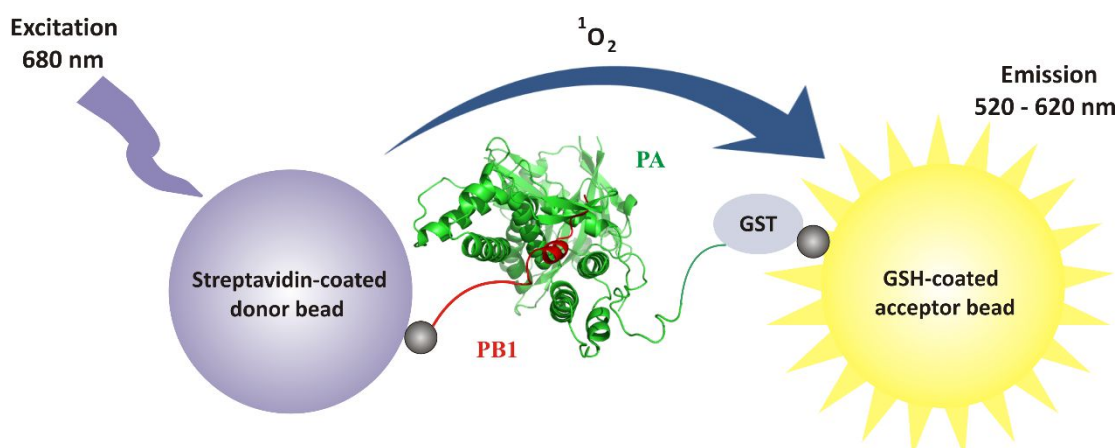
3.8. AlphaScreen technology

Figure 18: Schematic representation of the AlphaScreen experiment. The biotinylated PB1 peptide was coated onto the streptavidin containing donor latex bead. GST-CPA protein was coated onto GSH covered acceptor latex bead. After beads mixing, the solution is excited with light of a wavelength of 680 nm. Illumination of reactive groups within donor beads creates reactive oxygen species that can further diffuse in 200 nm distance before extinction. When the peptide is bound to its partner (CPA), reactive oxygen species can reach specific reactive groups within acceptor beads leading to the creation of emission in 520-620 nm spectra. If the interaction is disrupted, weaker or no emission signal is observed since the beads are no longer in the vicinity.

- AlphaScreen buffer: 25 mM Tris/HCl pH 7.4, 150 mM NaCl, 0.05% Tween20, 2% DMSO

The AlphaScreen [148] experiments were performed using Perkin Elmer Enspire plate reader in 96 well ProxiPlates. Biotinylated peptide PB1$_{1-25}$ (MDVNPTLLFLKIPAQNAISTTFPYTGGSK-Biotin) prepared by Fmoc-based peptide synthesis by Miroslava Blechová from the IOCB of the CAS was captured on Streptavidin-coated donor beads. Second reaction solution contained GST-CPA fusion protein that bound to GSH coated acceptor beads. Mixtures were incubated for 60 minutes (RT) at dark and subsequently mixed together and incubated for another 120 minutes while mixing. In experiments screening for potential inhibitors, compounds were mixed with both types of beads prior to 120 minutes incubation. First optimization experiment was proceeded to evaluate the optimal concentration of protein/peptide for further experiments. Finally 15 nM PB1$_{1-25}$-biotin and 60 nM GST-CPA were used in screening experiments. The concentration of beads was 2µg/ml in 50 µl reaction volume. All experiments were performed in AlphaScreen buffer. Assay evaluation was performed with PB1-0 peptide (MDVNPTLLFLKIPA) as well as with noninhibitory control

peptide (IYDPTLYGLEFD). Statistical evaluation of the assay was done as described previously [149]. Octaplicates of each positive signal (non-inhibited PB1-CPA interaction) and a negative background signal (uncoated beads) were measured and statistical evaluation was performed:

Signal to noise ratio determination:

$$S/N = \frac{\mu_{c+} - \mu_{c-}}{\sigma_{c-}}$$

Signal to background ratio determination:

$$S/B = \frac{\mu_{c+}}{\mu_{c-}}$$

Z´-factor – characteristic parameter for assay quality:

$$Z´ = 1 - \frac{(3\sigma_{c+} + 3\sigma_{c-})}{\mu_{c+} - \mu_{c-}}$$

Where:

- μ_{c+} is mean signal
- μ_{c-} is mean background
- σ_{c+} is standard deviation of the signal
- σ_{c-} is standard deviation of the background

For determination of a minimal binding motif of truncated PB1 peptides, 11 peptides were synthesized by Miroslava Blechová from the IOCB of the CAS and these peptides were tested for their inhibitory activity in titration experiments. IC_{50} values were obtained by fitting titration points in GraFit software (Erithacus Software).

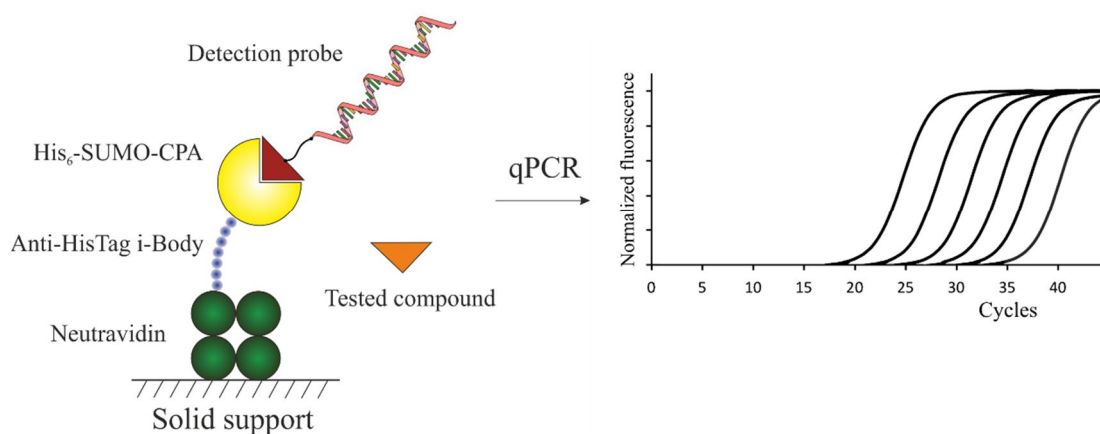## 3.9. DIANA (DNA-linked Inhibitor Antibody Assay)

Figure 19: Scheme of DIANA experiment [150] for high-throughput screening of potential inhibitors of the interaction between the C-terminal domain of influenza polymerase acidic subunit and the N-terminal peptide of polymerase basic subunit 1. In typical DIANA experiment, target protein (CPA) is bound to a specific antibody that is adsorbed to a solid support (in this case CPA fused to His_6-SUMO tag is bound to nickel chelating i-Body [151] that is linked by biotin substituents to neutravidin adsorbed to a solid support). Setout of this experiment is that neutravidin is adsorbed to a solid support (plate well) then the free excess surface is blocked with casein. Secondly, i-Body is incubated with neutravidin for one hour. Next complex of antigen (His_6-SUMO-CPA) and i-Body is formed within 2-hours incubation. Then each well is mixed with a solution of detection probe consisting of antigen-specific binding partner (PB1-0 peptide) covalently linked to reporter single-stranded DNA oligonucleotide. Sample well is thoroughly washed for displacement of the unbound probe. For inhibitor testing, the probe is mixed with the inhibitor for the establishment of equilibrium between them in context to the target protein. The final step is the addition of real-time PCR mix consisting of Taq polymerase, two probe-complementary primers, and a molecule that upon intercalation to DNA acts as a fluorophore. The plate is sealed and placed inside Lightcycler device (Roche) for real-time PCR experiment. As PCR reaches later cycles, the fluorescent signal is produced. Introduction of potent inhibitor results in a shift of the fluorescent signal production towards later PCR cycles in contrast to samples where the inhibitor is absent. Finally, C_q values are calculated for data analysis. The scheme was created in Inkscape software.

3.9.1. Probe preparation

For the probe preparation, the copper-catalyzed azide-alkyne click was preformed [152]. The PB1-0-TTDS-KN3 peptide (MDVNPTLLFLKIPA-TTDS-KN3) was synthesized and kindly provided by Miroslava Blechová from IOCB of the CAS. Alkyne-oligonucleotide with the same amplicon sequence as described previously[150] was purchased from Generi Biotech (Czech Republic). BTTP ligand was prepared by Milan Vrabel from the IOCB of the CAS. Click reaction solutions were mixed in microtube by mixing peptide and oligonucleotide both to a final concentration of 100 µM with the addition of catalysis mix solution consisting of BTTP ligand, $CuSO_4$, and sodium ascorbate all mixed in 100 mM HEPES buffer pH 7.0 (final molar ratio of peptide : oligonucleotide : Cu^{2+}: BTTP ligand : ascorbate beeing 1 : 1 : 7.7 : 15.4 : 15.4). All reagents were mixed to a final volume of 40 µl and then incubated in a thermocycler at 30 °C for 2 hours. In next step, reaction solution was repeatedly washed/reconcentrated with TBS (10,000 g, 23 °C) with Amicon centrifugal filter units (MWCO 10 kDa) to eliminate possible excess of unreacted peptide. Finally concentration was determined on Nanodrop spectrophotometer (Thermo Fisher Scientific). Quality of sample and determination of sample purity was performed using LCMS analysis that was carried out by Radko Souček from the IOCB of the CAS.

3.9.2. DIANA optimization experiment

- Casein solution: 5-fold diluted casein buffer (SDT)
- Buffer TBS: 20 mM Tris-HCl, pH 7.5, 150 mM NaCl
- Buffer TBST: 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% Tween20 (w/v)
- Buffer TBST´: buffer TBST: 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.1% Tween20 (w/v)
- Buffer TBST´c: buffer TBST´+ 500-fold diluted casein buffer (SDT)
- Buffer TBST´r: buffer TBST´ + 5 mM 2-mercaptoethanol
- Buffer TBST´rk: buffer TBST´r ´+ 500-fold diluted casein buffer (SDT)
- Buffer TBST´rkd1: buffer TBST´rk + 2% DMSO
- Buffer TBST´rkd2: buffer TBST´rk + 10% DMSO

On day 1 of DIANA experiment 10 ng/μl of neutravidin solution was added to 96-well plate (4titude) by multichannel pipette. The plate was incubated at room temperature for 60 minutes after which 100 μl of casein solution was added to each well for blockage of the free excess surface. The plate was sealed with foil and left incubating overnight at RT. Next day, the plate was placed in Blue Washer (Blue Cat Bio) to wash out excess casein. Next step included binding of anti-HisTag i-Body [151] to previously coated neutravidin. Each well was filled with 5 μl of 100 nM i-Body/50 μM $NiCl_2$ solution followed by 60 minutes incubation. Subsequently, wash step with Blue Washer was executed. In the next step binding of the target protein (His_6-SUMO-CPA protein) to i-Body was performed. Various concentrations of protein in TBST´r buffer were dispensed to each well followed by 120 minutes incubation at RT. Subsequently, wash step with Blue Washer was executed. Next step included incubation with the DIANA probe and in some wells with a PB1-0 inhibitory peptide of various concentrations in buffers TBST´rk, TBST´rkd1 or TBST´rkd2. Subsequently, thorough wash step with Blue Washer was executed. In the last step, 5 μl of the qPCR mix of the same composition as described in [150] was added to each well. The plate was tightly sealed with foil and placed inside Lightcycler device (Roche) to detect real-time PCR curves using the protocol as described previously [150].

Comment: The plate was spun down (2,000 g, RT, 5 min) after each step during plate preparation.

The final analysis was done directly in Light cycler 480 II software using the method of maxima of the second derivative from fluorescent curves for obtaining cycles C_q. K_d values of the probe were measured by titration of captured CPA by this probe. Finally, K_d values were obtained by fitting concentrations of the bound probe by function:

$$[EP] = \frac{E_{tot} * P_{tot}}{K_d + P_{tot}}$$

Where:

- [EP] is the amount of bound probe
- P_{tot} is the analytical concentration of the probe during the incubation
- E_{tot} is the amount of captured enzyme
- K_d is the dissociation constant

E_{tot}, as well as K_d, are parameters to be solved by fitting. Fitting was done using GraFit software.

K_i values for PB1-0 peptide were determined according to this function:

$$K_i = \frac{2^{-\Delta C_q}}{1 - 2^{-\Delta C_q}} * \frac{I_{tot}}{1 + \frac{P_{tot}}{K_d}}$$

Where:

- K_i is the inhibition constant
- ΔCq is the difference between the C_q value in well incubated with the particular compound and mean C_q value of wells incubated without any compound
- I_{tot} is the concentration of the tested compound
- Meaning of the other variables is the same as in previous equation

3.10. Protein crystallization

- crystallization buffer: 10 mM Tris/HCl, pH 8.0, 1 mM TCEP

Sufficient amount of protein (aliquots - 326 µg/ml) was concentrated to 5.5 mg/ml (10,000 g, 10 °C) with Amicon centrifugal filter units (MWCO 30 kDa). The concentrated protein was washed with 1 ml of crystallization buffer. Lastly, the protein was washed with 50 µM PB1-11 peptide (DYNPYLLFLK) in crystallization buffer + 2% DMSO) and re-concentrated to 5.5 mg/ml. Protein sample homogeneity (monodispersity) was tested on DLS instrument (Laser-Spectroscatter 201, RiNA Netzwark RNA Technologien GmbH).

3.10.1. Automated robotic screening

Sitting drop vapor-diffusion method was chosen for initial screening of the crystallization conditions. 96-well plate (MRC 3 well low, Swissci) was filled with 30 µl of crystallization screen conditions by the Crystal Gryphon robot (Art Robbins Instruments). Protein was spun down (15,000 g, 18°C, 30 min) prior to screening. Plate with screening conditions solutions was placed inside Oryx8 (Douglas Instruments) holding compartment. The fluidic tip was washed and channels were debubbled. Then 20 µl per 96 drops of protein sample was loaded inside micro microtube and placed inside designated holder (in case of seeding experiments 5 µl of seed stock per 96 drops was

placed inside the secondary tube). The program was set to pipet 400 nl drop per condition at ratio 1:1 protein: condition solution (in case of seeding 4 : 3 : 1 protein: condition solution: seed stock). The plate was sealed with foil and placed inside Gallery DT plate hotel for scheduled photo documentation (visible, UV inspections).

3.10.2. Hand optimization

In Hand optimization experiments vapor-diffusion hanging drop method was used. Optimized conditions of MORPHEUS® crystallization screen kit (Molecular Dimensions) were prepared from MORPHEUS® mixes to desired final conditions. 500 µl of prepared conditions were added to each well of EasyXtal 15-well plate (QIAGEN). Four 2 µl drops (5 : 4 : 1, Protein : condition solution : seed stock) were dispensed per well.

4. Results

4.1. Cloning experiments

All cloning experiments were successful as both genes coding C-terminal domain of influenza polymerase acidic subunit were cloned inside appropriate expression plasmid for GST and His_6-SUMO fusion which were later sequenced by Sanger sequencing with no mutations recorded within genes. These plasmids were furthermore used for protein expression experiments as they were successfully transformed into *E.Coli* BL21(DE3) RIL strain.
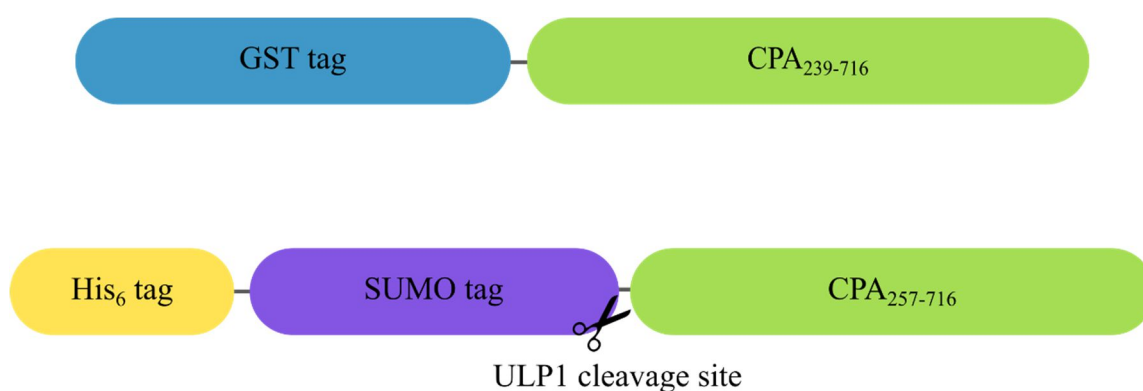

Figure 20: Scheme of recombinant protein constructs. GST-CPA was later used in development of AlphaScreen HTS assay. Second uncleaved construct of His_6-SUMO-CPA was also used in development of HTS assay based on DIANA technology and small fraction of protein was digested with ULP1 protease, further separated from tags and used in crystallization experiments.

4.2. Recombinant protein purification

4.2.1. GST-CPA protein

The recombinant C-terminal domain of polymerase acidic subunit of influenza strain A/California/07/2009 (H1N1) carrying GST affinity tag was purified by affinity chromatography on the GSH-agarose resin. Protein was over-expressed in *Escherichia coli* BL21(DE3) RIL strain for potential use in the development of high-throughput screening assays. Protein was successfully purified with an approximate yield of 10 mg from 3 liters of bacterial cells suspension with >90% purity. Protein was later re-purified using gel permeation chromatography for higher purity. The purification process was analyzed with the SDS-PAGE method and later visualized by CBB staining.

Lane	Load volume	Fraction
1	1 μl	All blue marker
2	5.4 μl	Before Induction
3	1 μl	After induction
4	1 μl	Soluble frac.
5	1 μl	Insoluble frac.
6	1 μl	Flow through
7	1.5 μl	1. elution
8	1.5 μl	2. elution
9	1.5 μl	3. elution

Figure 21: SDS-PAGE documentation of affinity chromatography purification process. On the left side, polyacrylamide gel is stained with Coomassie Brilliant Blue. All lanes correspond to table on the right side depicting individual sample loads. Molecular weight protein standard All blue marker (Bio-Rad) is in the first lane. Lanes 2-9 correspond to pooled fractions during purification steps.

As you can see in Figure 21, GST-CPA protein was expressed as a significant band around 79 kDa can be observed in lane 3. A substantial amount of protein was produced in an insoluble form, nevertheless substantial portion of protein was purified from the cell lysate and obtained after elution from affinity column.

Gel permeation chromatography was performed to further improve protein purity. The first and second elution from GSH-agarose resin were pooled together, concentrated and applied to Superdex 75 column in 3 separate identical experiments.

Figure 22: Chromatogram from the gel permeation chromatography experiment. GST-CPA protein was subjected to buffer exchange and subsequently purified from small molecular weight impurities. The fraction corresponding to V_e (53-57) was picked as a final fraction for further experiments.

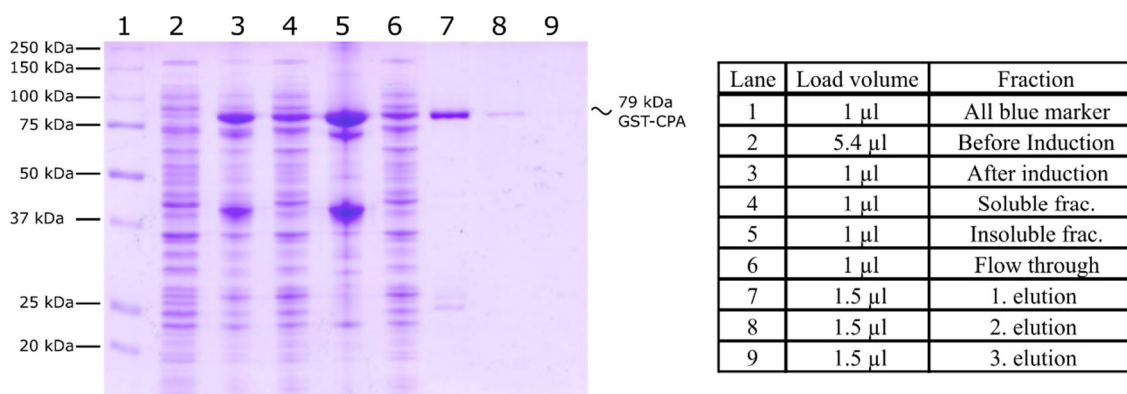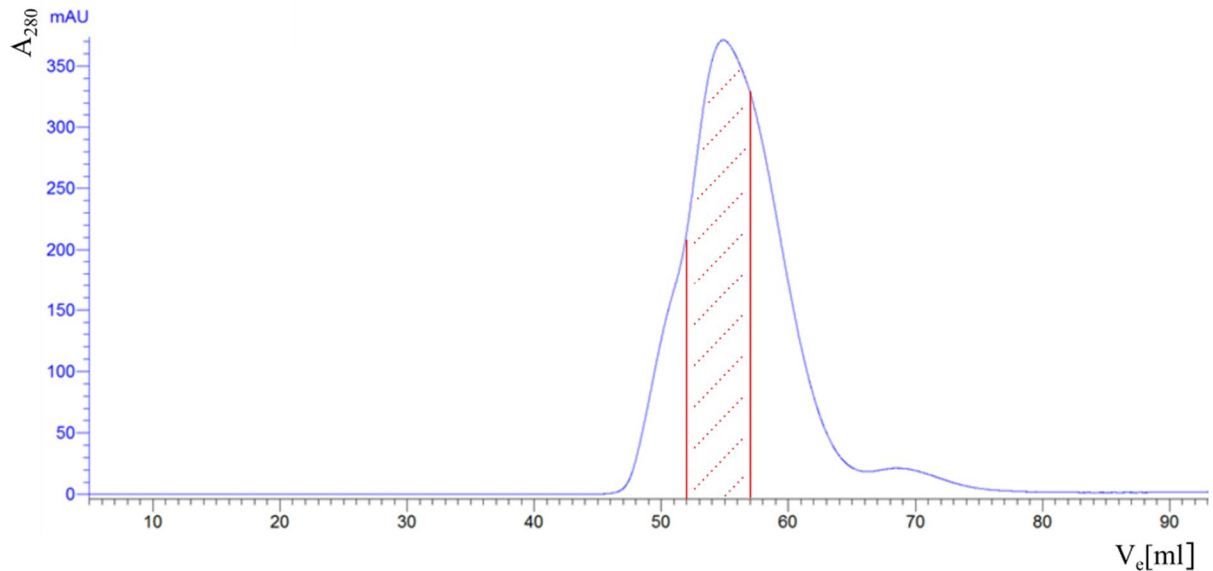| Lane	Load volume	Fraction
1	1.5 μl	All blue marker
2	6 μl	Before GPC
3	6 μl	Fraction 48-52 ml
4	6 μl	Fraction 53-57 ml
5	6 μl	Fraction 58-62 ml
6	6 μl	Fraction 63-67 ml
7	6 μl	Fraction 68-72 ml

Figure 23: SDS-PAGE documentation of the gel permeation chromatography purification process. On the left side, polyacrylamide gel is stained with Coomassie Brilliant Blue. All lanes correspond to table on the right side depicting individual sample loads. Molecular weight protein standard All blue marker (Bio-Rad) is in the first lane. Lanes 3-7 correspond to gathered fractions during elution from the column.

GST-CPA protein was successfully transferred into the final buffer as well as additionaly purified from small molecular weight impurities. As documented in

Figure 23, visible change in impurities comparing protein sample before and after GPC can be observed. Finally, 1.2 mg of sufficiently pure GST-CPA protein was obtained for later SPR/AlphaScreen experiments by obtaining 5 ml fraction (Lane 4 in Figure 23).

4.2.2. His$_6$-SUMO-CPA protein

Recombinant protein of C-terminal domain of polymerase acidic subunit of influenza strain A/California/07/2009 (H1N1) carrying His$_6$ affinity tag in addition to SUMO solubility tag was purified by chelation chromatography on the Ni-NTA-agarose resin. Protein was over-expressed in *Escherichia coli* BL21(DE3) RIL strain for potential use in the development of DIANA high-throughput screening assay as well as for crystallization experiments. Protein was successfully purified after affinity chromatography with an approximate yield of 66 mg from 3 liters of bacterial cells suspension with >90% purity. A small fraction was devided from elution fraction and proceeded directly to GPC to obtain higher purity and to be later used in DIANA assay (data not shown). Major fraction was further processed by cleavage of His$_6$-SUMO tag. Protein was later re-purified using gel permeation chromatography for higher purity.

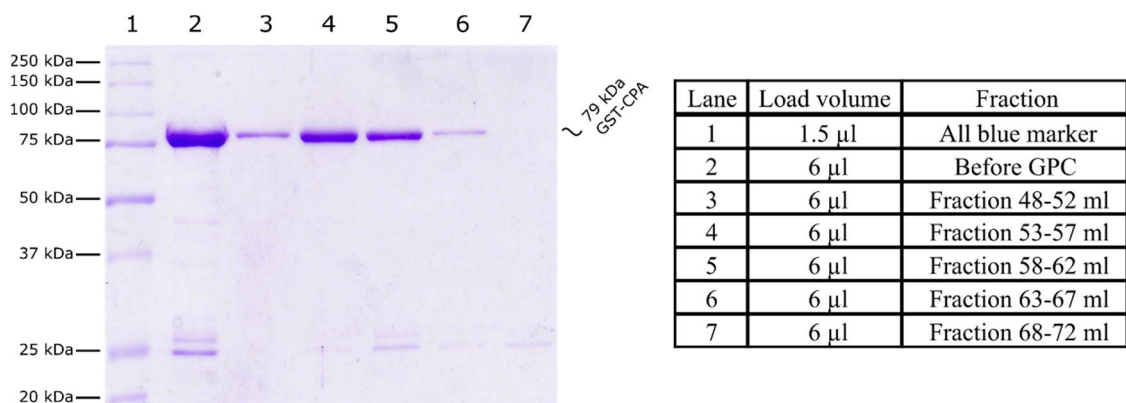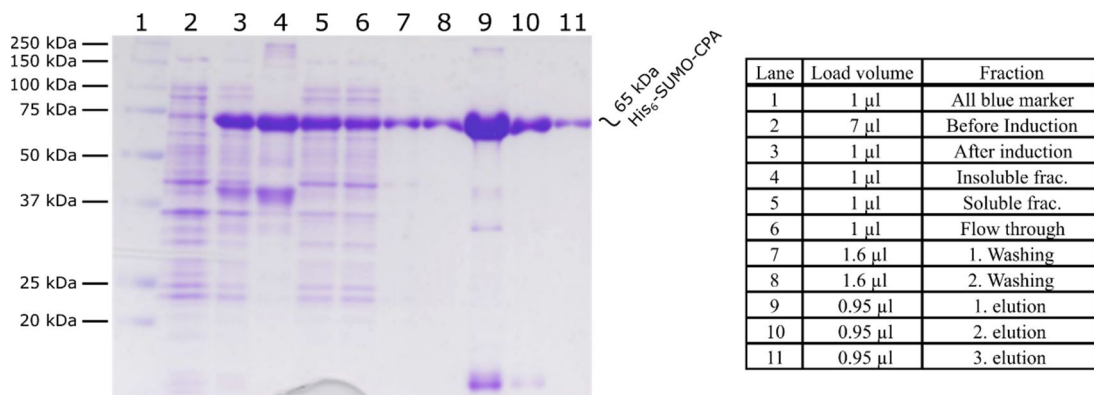| Lane	Load volume	Fraction
1	1 µl	All blue marker
2	7 µl	Before Induction
3	1 µl	After induction
4	1 µl	Insoluble frac.
5	1 µl	Soluble frac.
6	1 µl	Flow through
7	1.6 µl	1. Washing
8	1.6 µl	2. Washing
9	0.95 µl	1. elution
10	0.95 µl	2. elution
11	0.95 µl	3. elution

Figure 24: SDS-PAGE documentation of chelation chromatography purification process. On the left side, polyacrylamide gel is stained with Coomassie Brilliant Blue. All lanes correspond to table on the right side depicting individual sample loads. Molecular weight protein standard All blue marker (Bio-Rad) is in the first lane. Lanes 2-11 correspond to gathered fractions during purification steps. The purification process was analyzed with the SDS-PAGE method and later CBB staining.

Distinctive overexpression can be monitored in Figure 24 where substantial amount of protein in each insoluble and soluble protein fraction was obtained. High amount of protein was obtained after elution from resin.

After successful purification, His$_6$-SUMO tag was cleaved by a ULP1 protease from the major portion of pooled first and second elution fractions. Cleavage progress was also monitored by SDS-PAGE/CBB staining experiment.

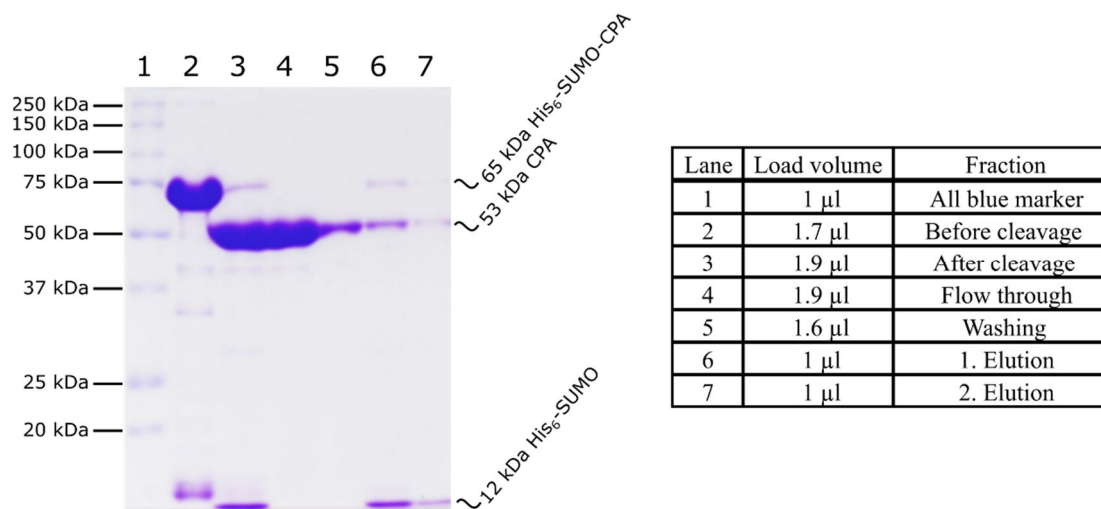| Lane	Load volume	Fraction
1	1 µl	All blue marker
2	1.7 µl	Before cleavage
3	1.9 µl	After cleavage
4	1.9 µl	Flow through
5	1.6 µl	Washing
6	1 µl	1. Elution
7	1 µl	2. Elution

Figure 25: SDS-PAGE documentation of chelation chromatography purification process after successful cleavage of fusion tags from CPA protein. On the left side, polyacrylamide gel is stained with Coomassie Brilliant Blue. All lanes correspond to table on the right side depicting individual sample loads. Molecular weight protein standard All blue marker (Bio-Rad) is in the first lane. Lanes 2-7 correspond to gathered fractions during purification steps

Cleavage of fusion tags was successful as documented in lanes 2 and 3 in Figure 25. Small fraction remained uncleaved but was removed by chelation chromatography later as you can see in the fourth lane. Also, the cleaved His$_6$-SUMO tag was removed by binding to the resin. Some protein was lost due to unspecific binding to Ni-NTA resin as observable in lane 5 where resin wash was performed. The main fraction, in this case, was flow-through fraction where individual CPA protein was obtained. This fraction was further processed by GPC to get protein of high purity.

Gel permeation chromatography was performed to further improve protein purity. The flow through fraction after His$_6$-SUMO cleavage was diluted and applied to Superdex 75 column in 9 separate identical experiments. In each experiment, 5 ml of protein sample was loaded inside injection loop and subsequently injected into the column. Three ml fractions were collected from fractions with registered significant absorption at 280 nm.
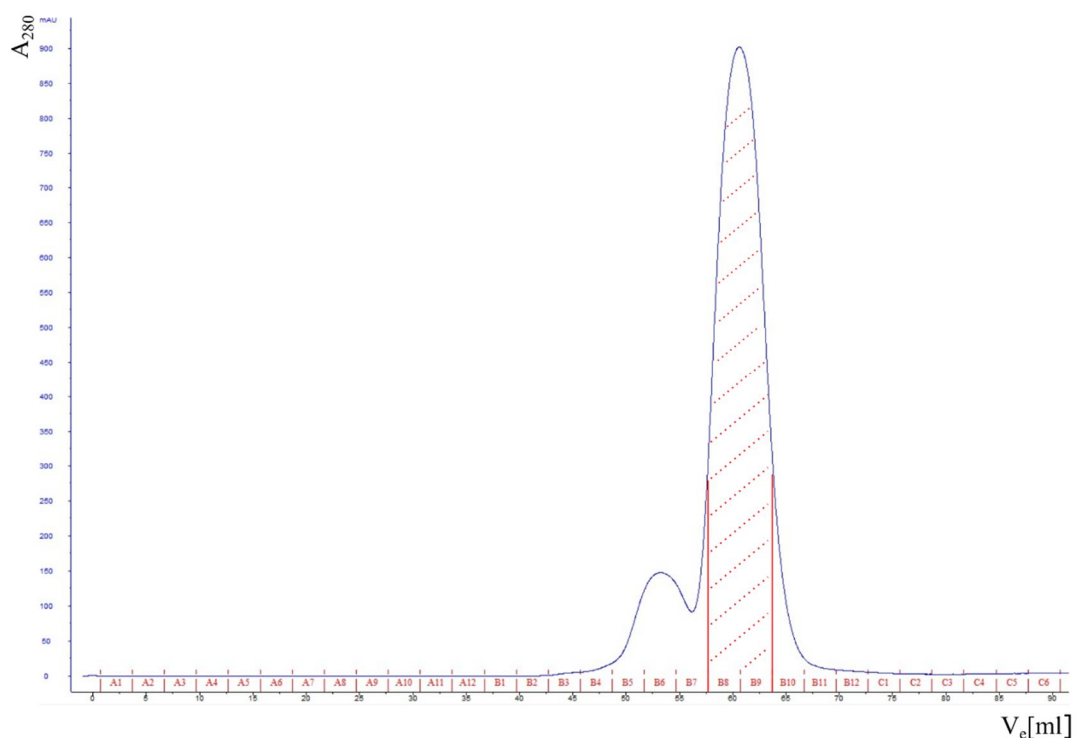
Figure 26: Chromatogram from the gel permeation chromatography experiment documenting CPA protein purification from impurities as well as from CPA in the multimeric state. Highest peak with a maximum at fraction B8-B9 was main exclusion volume for monomeric CPA.

Mainly multimeric states of CPA corresponding to first peek (B5-B7) were separated from monomer CPA. Fractions B5-B11 were tested for purity in SDS-PAGE/CBB staining experiment to further evaluate protein purity.

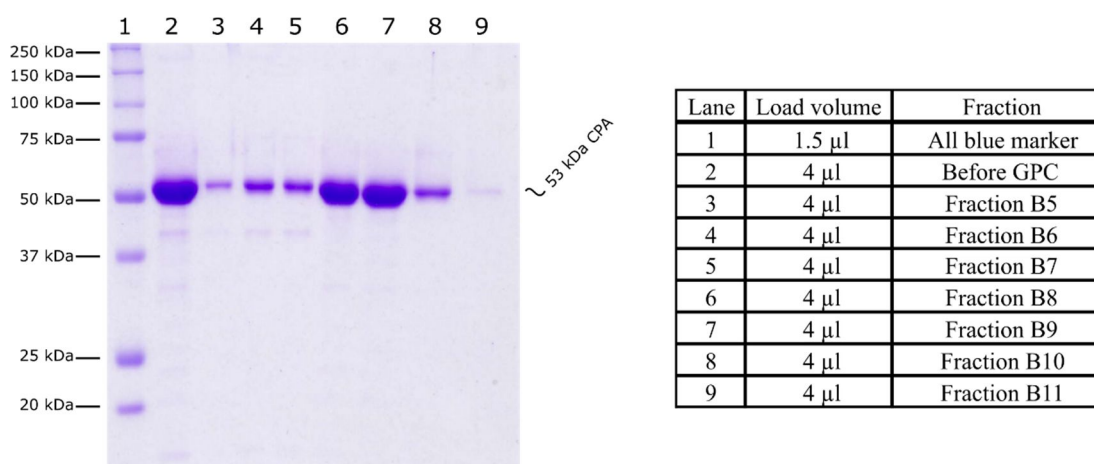| Lane	Load volume	Fraction
1	1.5 µl	All blue marker
2	4 µl	Before GPC
3	4 µl	Fraction B5
4	4 µl	Fraction B6
5	4 µl	Fraction B7
6	4 µl	Fraction B8
7	4 µl	Fraction B9
8	4 µl	Fraction B10
9	4 µl	Fraction B11

Figure 27: SDS-PAGE documentation of gel permeation chromatography experiment in which CPA protein was re-purified after chelation chromatography. On the left side, polyacrylamide gel is stained with Coomassie Brilliant Blue. All lanes correspond to table on the right side depicting individual sample loads. Molecular weight protein standard All blue marker (Bio-Rad) is in the first lane. Lanes 3-9 correspond to elution fractions from the column.

Figure 27 on previous page depicts purification of CPA protein by GPC. Fractions B8 and B9 (lane 6, 7) were selected as the purest ones. When compared to the second lane on SDS-PAGE gel, decrease in contaminant proteins is observable. B8-B9 fractions from all 9 GPC experiments were pooled together, concentrated and further aliquoted for future crystallization experiments.

4.3. Analysis of ligand binding by SPR

SPR experiments were performed to test whether GST-CPA protein is able to bind PB1 peptide. Furthermore, next experiment tested the effect of detergent concentration on binding to the peptide. In the third experiment, K_d value of PB1 peptide binding to GST-CPA protein was determined. Lastly, control experiment of GST binding to PB1 layer was performed.
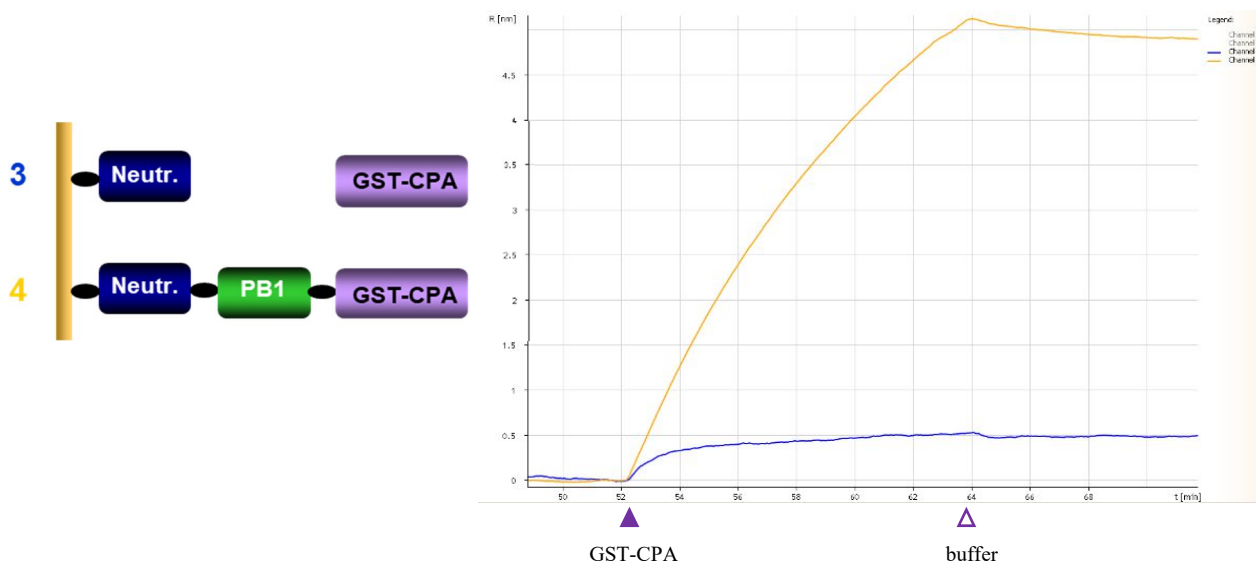

Figure 28: First SPR experiment. Scheme on the left side depicts an array of two tested channels. The sensorgram on the right side shows timely accumulation of proteins on the golden chip as a change in refractive index. R unit on the vertical axis is relative unit corresponding to the accumulation of 1pg/mm².

In the first experiment (Figure 28), the evaluation of ability of GST-CPA protein bind to PB1 peptide was tested. Nonspecific binding of GST-CPA protein to neutravidin layer was tested. The first experiment shows that GST-CPA is able to bind to PB1 peptide. However it also nonspecifically binds to neutravidin layer. For determination of K_d value of this interaction, testing of detergents to abolish this nonspecific binding had

to be performed. Binding to the PB1 peptide in the presence of detergent in the running buffer was tested.
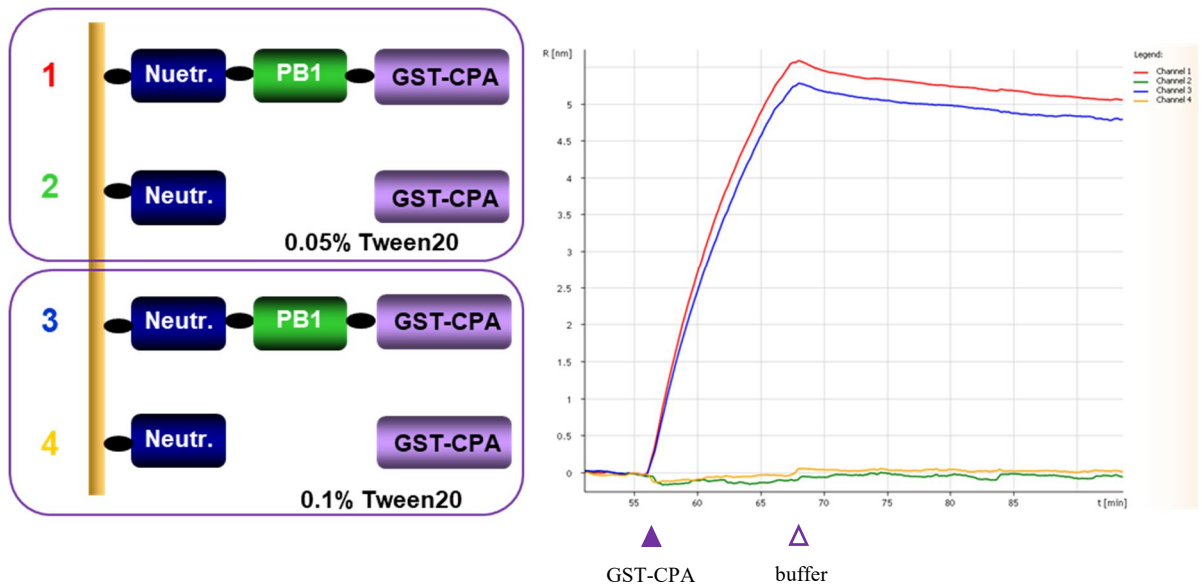

Figure 29: Second SPR experiment was performed to determine the effect of detergent on GST-CPA binding to the PB1 peptide as well as to test nonspecific binding to neutravidin. Scheme on the left side depicts an array of two tested channels. Sensorgram on the right side shows timely accumulation of proteins on the chip as a change in refractive index. R unit on the vertical axis is relative unit corresponding to the accumulation of 1pg/mm².

As you can see in Figure 29, introduction of detergent (Tween20) resulted in the elimination of non-specific binding of GST-CPA to neutravidin. Furthermore, no significant effect of higher concentration of this detergent on PB1/GST-CPA binding was observed.

Finding this optimal buffer allowed further running experiments determining K_d value of interaction between the PB1 peptide and GST-CPA protein. In the third experiment, four concentrations of GST-CPA were added to each channel.
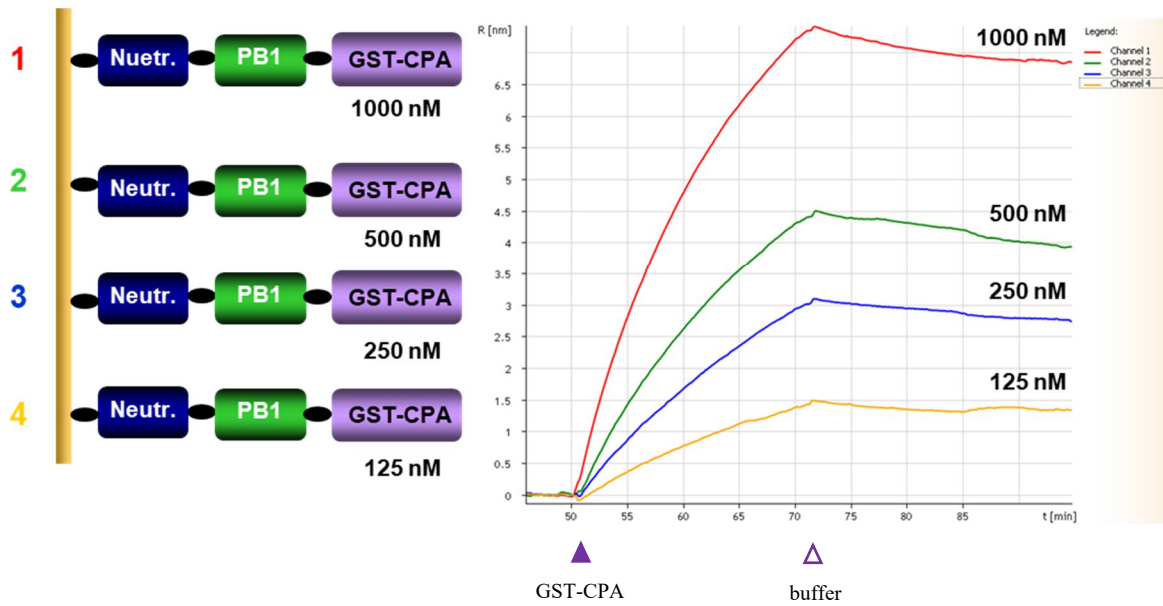
Figure 30: SPR experiment providing a determination of dissociation constant K_d of interaction between the PB1 peptide and GST-CPA protein. The left scheme shows an array of components in individual channels. Sensorgram on the right side shows the time course of SPR experiment and growth of layer upon binding of GST-CPA to PB1 peptide. Channels 1, 2, 3, 4 correspond to concentrations of GST-CPA 1,000 nM, 500 nM, 250 nM, and 125 nM, respectively. R unit on the vertical axis is relative unit corresponding to the accumulation of 1pg/mm².

All four curves from experiment shown in Figure 30 were imported into TraceDrawer software. Curves were fitted according to standard 1:1 model and k_{on}, k_{off} values were subtracted. Using these values dissociation constant K_d was calculated.

$$K_d = 23 \pm 6 \text{ nM}$$
$$k_{on} = (1.9 \pm 0.04).10^{3} \text{ M}^{-1}.\text{s}^{-1}$$
$$k_{off} = (4.4 \pm 1.0).10^{-5} \text{ s}^{-1}$$

Figure 31: Kinetic parametres of binding of PB1 peptide to GST-CPA protein.

In the last fourth experiment, we just wanted to explore possible binding of isolated GST to PB1-peptide. We tested both binding of GST to neutravidin layer and to PB1 peptide layer. No unspecific binding of GST protein to both these components was observed (data not shown).

4.4. AlphaScreen assay optimization, evaluation

High-throughput screening assay for the search of inhibitors targeting interaction between the C-terminal domain of influenza polymerase acidic subunit and the N-terminal peptide of polymerase basic subunit 1 was designed and evaluated. In pilot AlphaScreen experiment, we tested several concentrations of both biotinylated PB1 peptide and GST-CPA protein coated onto streptavidin donor and GSH acceptor beads, respectively.
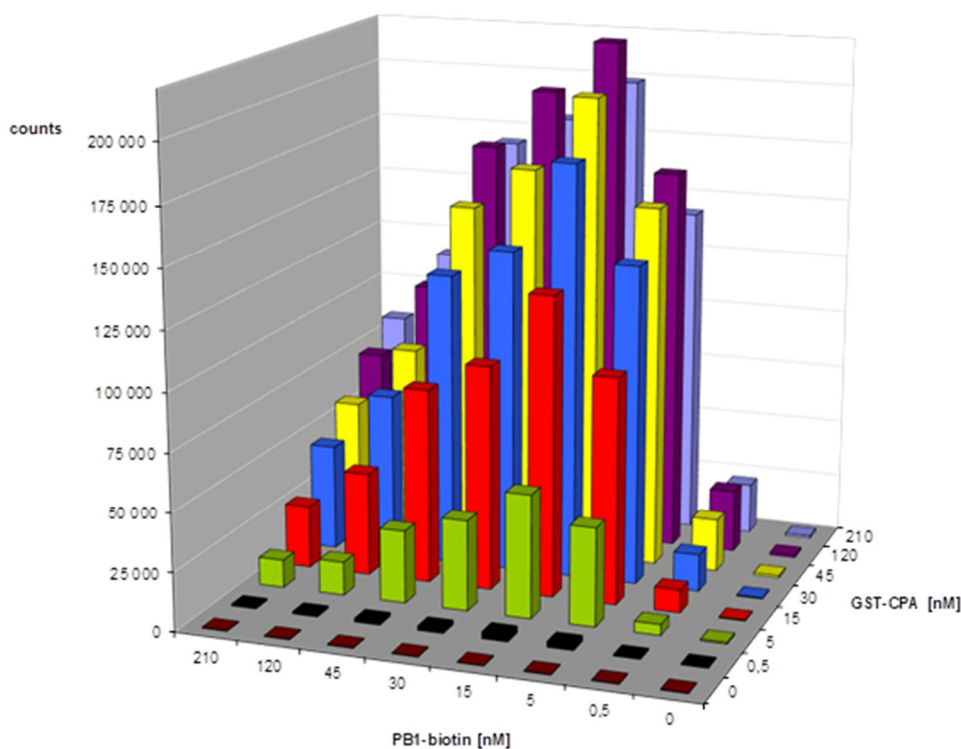
Figure 32: AlphaScreen cross-titration experiment. Fluorescent signal (counts) shown on the vertical axis. Two horizontal axes show used concentrations of both biotinylated PB1 peptide and GST-CPA protein.

The optimal concentrations of reactants were identified as 15 nM PB1 peptide and 60 nM GST-CPA protein where signal response was around the highest values (more than 200,000 counts). This condition was further selected for high-throughput screening

experiments. Statistical evaluation of this assay was performed as described previsouly [149].

signal-to-noise ratio S / N = 19,000

signal-to-background ratio S / B = 630

Z′ = 0.97

Figure 33: Statistical evaluation of of AlphaScreen assay.

When screening for minimal peptide binding motif, truncated PB1 peptides were tested for their inhibitory activity. The IC_{50} values were determined by AlphaScreen optimized assay from the peptide titrations. Furthermore, non-inhibitory peptide was also tested as a negative control.
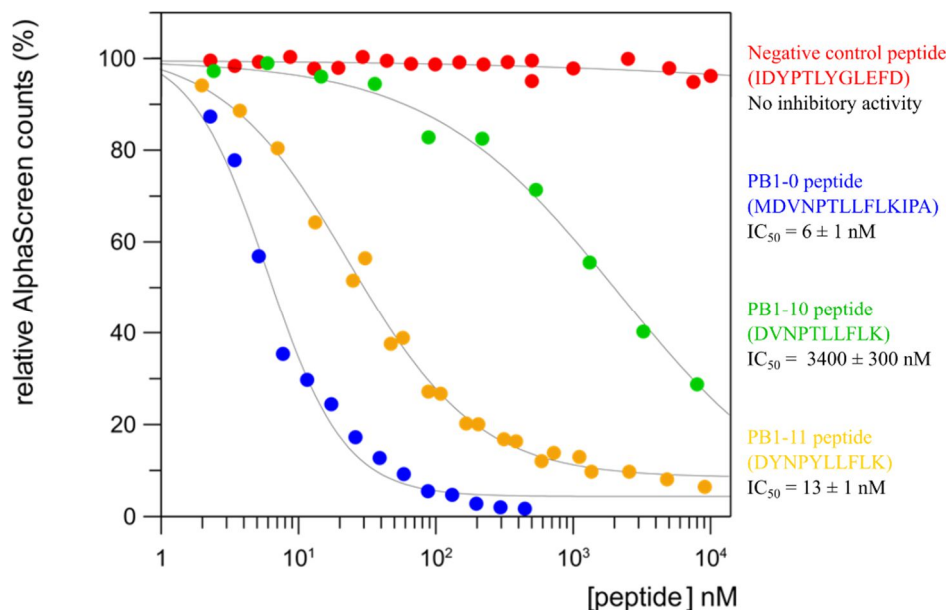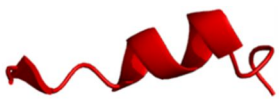
Figure 34: The titration curves for four peptide titrations in the AlphaScreen experiment. Positive control PB1-0 peptide (blue) with significant inhibitory activity, non-inhibitory negative control peptide (red) and two versions of truncated decapeptides (yellow, green) are shown. The diagram shows the dependency of relative AlphaScreen signal percentage on compound concentration. The relative signal is gradually decreased with increasing potent inhibitor concentration.

First peptide (positive control PB1-0 peptide) showed inhibition of interaction resulting in a significant reduction of AlphaScreen signal. Titration points were fitted with curves in GraFit software (Erithacus Software LTD.) and IC_{50} values for these peptides were determined. Furthermore, 11 truncated peptides were tested in the same manner.

		IC_{50}
PB1-0 :	MDVNPTLLFLKIPA	6 ± 1 nM
PB1-1 :	MDVNPTLLFLKIP	22 ± 1 nM
PB1-2 :	MDVNPTLLFLKI	39 ± 4 nM
PB1-3 :	MDVNPTLLFLK	130 ± 20 nM
PB1-4 :	MDVNPTLLFL	440 ± 40 nM
PB1-5 :	MDVNPTLLF	7,200 ± 1,000 nM
PB1-6 :	DVNPTLLFLKIPA	33 ± 3 nM
PB1-7 :	VNPTLLFLKIPA	390 ± 50 nM
PB1-8 :	NPTLLFLKIPA	20,200 ± 1,700 nM
PB1-9 :	PTLLFLKIPA	36,000 ± 4,000 nM
PB1-10 :	DVNPTLLFLK	3,400 ± 300 nM
PB1-11 :	DYNPYLLFLK	13 ± 1 nM

Figure 35: IC_{50} values of the interaction of truncated PB1 peptides with CPA. The red ribbon picture shows collocation of structural elements within the peptide. Peptide helix is further highlighted in each peptide sequence in grey. For each peptide tested, the final IC_{50} value of binding of the peptide to the C-terminal domain of polymerase acidic subunit is shown on the right side.

4.5. Crystallization experiments

C-terminal domain of polymerase acidic subunit was co-crystalized in the presence of decapeptide PB1-11 (DYNPYLLFLK). Initially, 6 commercial screens (Morpheus® – Molecular Dimensions, JBScreen JCSG++ - Jena Bioscience, PEGs Suite – QIAGEN, Crystallization Basic Kit for Proteins – Sigma Aldrich, Crystallization Extension Kit for Proteins – Sigma Aldrich and Index Hampton – Hampton Research) counting for 480 chemically different conditions were tested. Crystallization plates were observed in time for potentially grown crystals. Condition B5 from the Morpheus screen (0.1 M MOPS/HEPES pH 7.5, 10% PEG 20000, 20% PEG MME 550, 30 mM NaBr, 30 mM NaF, 30 mM NaI) was optimal for crystals growth.
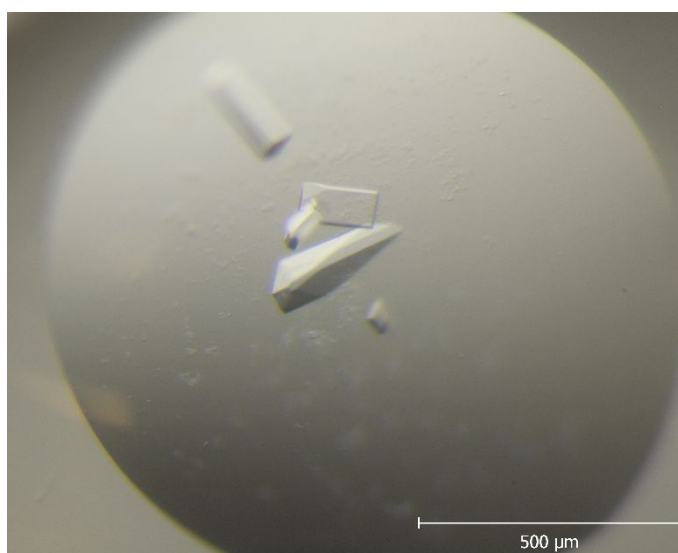
Figure 36: Crystals from the initial screening of conditions. These crystals were later used for seeding experiments.

These crystals were used in seeding experiments in which 4 aforementioned screen kits were tested (Morpheus® – Molecular Dimensions, Crystallization Basic Kit for Proteins – Sigma Aldrich, Crystallization Extension Kit for Proteins – Sigma Aldrich and Index Hampton – Hampton Research). In seeding experiments, multiple conditions provided crystals. Few conditions were cherry-picked as most suitable and hand optimized. Finally the best crystals were collected (Index Hampton D8 condition – 0,1 M HEPES pH 7.5, 35% PEG 3350) and tested for optimal cryo-condition. Addition of 30% ethylene glycol was chosen as the best. One crystal was flash frozen in liquid

nitrogen and measured by Petr Pachl Ph.D. from the IOCB of the CAS at Bessy II beamline at synchrotron radiation center at Helmholtz Zentrum Berlin. Data were collected with diffraction up to 1.6 Å. The structure was later solved also by Petr Pachl Ph.D.
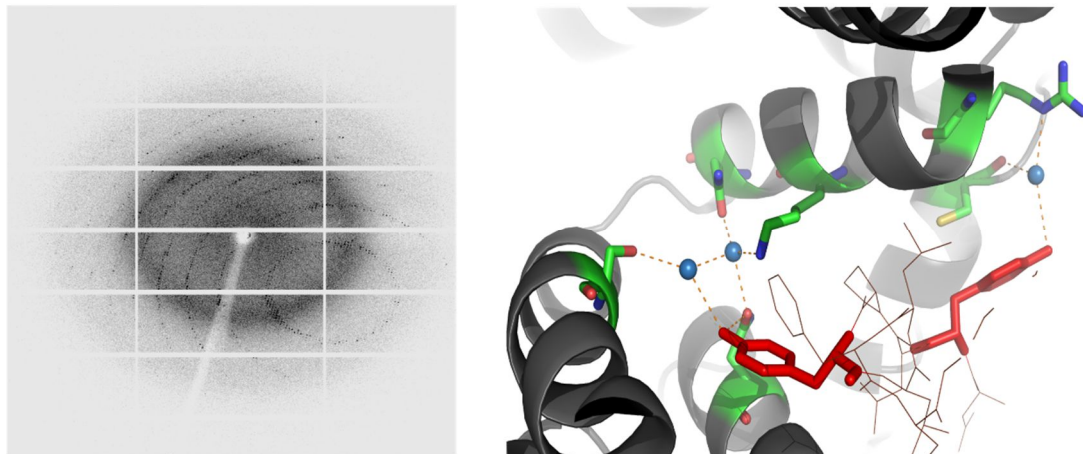
Figure 37: Left picture shows diffraction pattern of the crystal of complex between the C-terminal domain of polymerase acidic subunit and peptide PB1-11 measured on synchrotron. The picture on the right created in PyMOL depicts the interaction between the peptide (Red) and CPA (grey/green). Two mutated tyrosines inside inhibitory peptide are highlighted. The hydrogen bonds formed with side chains of these tyrosines are showed as dashed red lines.

4.6. DIANA optimization experiment

DIANA was designed for potential high-throughput screening of compounds that would inhibit the protein-protein interaction between influenza polymerase subunits PA and PB1. The basic principle of DIANA is a quantification of binding of a specific ligand of a protein conjugated to a reporter DNA oligonucleotide by real-time PCR (for more detail see Figure 19 on page 59). In this particular setup, we determine the inhibition of the interaction of the specific probe (peptide ligand conugated to an oligonucleotide) by the tested compound. This assay was chosen for its easy technical execution as well as for excellent hit discovery rate supported by getting a low number of unspecific hits. Previous experiments determining basic viable buffer conditions and principle setup were done by my consultant Milan Kožíšek Ph.D.

This experiment was performed in order to investigate dissociation constant of DIANA probe (PB1-0 peptide conjugated to DNA). K_d value for a probe at 1,000 pg of protein per well was 1 nM. Moreover, background signal of the assay was determined to be approximately 6 cycles.
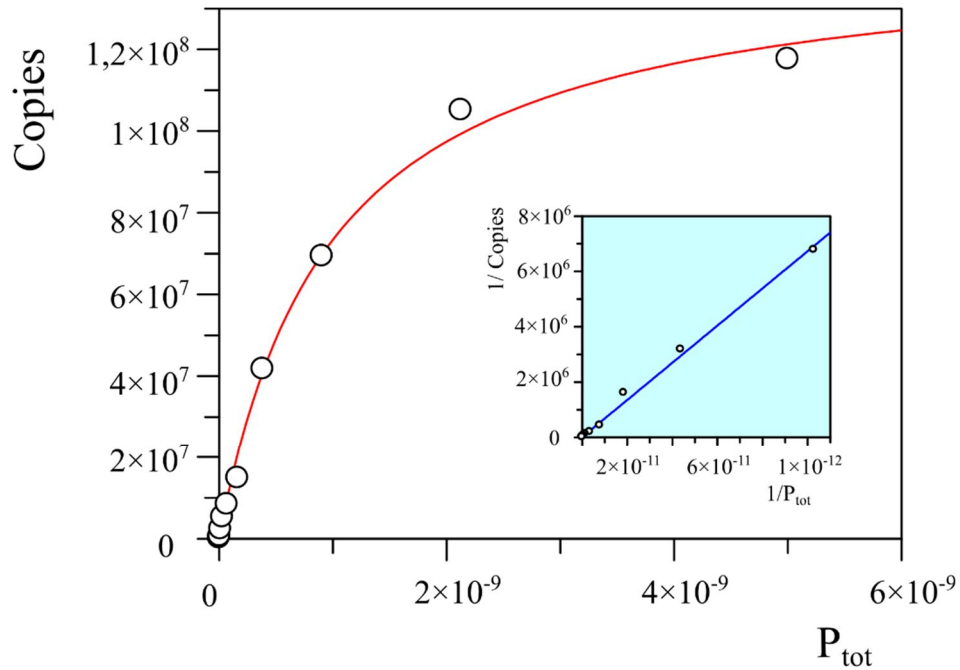
Parameter	Value [M]	Std. Error
K_d	9.74×10^{-10}	9.6×10^{-11}

Figure 38: Graph showing fitting of titration of His$_6$-SUMO-CPA protein with DIANA probe for determination of K_d. Y axis depicts calculated copies of probe molecules which are counted from standard calibration experiment from previous DIANA experiments. These copies correspond to meassuered C_q values determined from real-time PCR experiment. X axis corresponds to probe concentration. Smaller graph shows double reciprocal plot illustrating linearized function.

Finally, the inhibition activity of unconjugated peptide PB1-0 (also tested in AlphaScreen assay, see Figure 35 on page 75) was investigated. Peptide was incubated with the captured CPA protein and DIANA probe at 3 different concentrations of DMSO to analyze possible effect of DMSO on peptide inhibition efficiency. Inhibition constants K_i of PB1-0 peptide were calculated at these 3 DMSO concentrations from obtaind ΔC_q values.

DMSO [%]	K_i PB1-0 [nM]
0	0.71 ± 0.64
2	0.49 ± 0.31
10	0.30 ± 0.18

Figure 39: Determined K_i values for PB1-0 peptide at different DMSO concentrations.

5. Discussion

Influenza virus is one of the most common pathogens of humans. This virus is responsible for flu disease that is specific for its rapid development of symptoms such as fever muscle/body aches, a headache etc. In most severe cases this illness can lead to patient death. Age groups consisting of young children and elderly are the ones most prone to complications development. Moreover, virus ability of rapid transmission between hosts makes it effectual in development of epidemics and less frequently pandemics. The virus is also unique since it can infect different host species resulting in the frequent emergence of antigenically different viral strains. This effect is responsible for the unexpected introduction of new epidemic strains to the human population.

Only two drug targets within influenza virus have been exploited. First, M2 proton channel inhibitors displayed rapid development of mutations leading to resistance shortly after the introduction of these drug to the public. Nowdays, the application of these drugs is not recommended. The secondary target is influenza neuraminidase. Two main drugs (oseltamivir, zanamivir) were developed against this target. They have been effectively used both in treatment and prophylaxis. However, after the introduction of 2009 pandemic influenza resistant mutations within neuraminidase emerged.

A search for new targets within influenza virus has been of great interest in past few years. A lot of treatment possibilities have been recently proposed and studied. One of the most promising molecules influenza RNA dependent RNA polymerase has been found, showing few potential drug targets. Three main sites were recognized for their interesting mechanistic properties that could be possible targets. First two are the cap-binding activity of polymerase basic subunit 2 highly important in viral genomic replication, as well as endonuclease activity of the N-terminal domain of polymerase acidic subunit important during the transcriptional stage of viral life cycle. One more target has been studied – the C-terminal domain of influenza polymerase acidic subunit. This domain forms a protein-protein interaction with the N-terminal part of PB1 subunit and inhibition of this interaction has been shown to impair virulence. Interestingly, considering challenges of protein-protein interaction inhibition, this interface possesses properties such as relatively small buried interaction interface as well as the high affinity of these binding partners which makes this target of great importance.

In this master thesis, I focused on designing recombinant protein constructs for later utilization in development of high-throughput screening of inhibitors of the protein-protein interaction between influenza C-terminal domain of polymerase acidic subunit and the N-terminal part of polymerase basic subunit 1. I designed two constructs of the C-terminal domain of PA subunit. First, the construct of GST fused with CPA was designed with intention to use it in the development of high-throughput screening. The second construct was hexahistidine-SUMO tag fused to CPA that could be used later in another high-throughput screening assay and for protein crystallography. Both proteins were expressed in bacterial expression systems and both were expressed with substantial yield. GST-CPA construct expression was significantly lower compared to His_6-SUMO-CPA. That was probably influenced by SUMO protein that enhances the solubility of fused proteins. However, considering low consumption for experiments carried out with GST-CPA construct this lower expression profile was not an issue. Both proteins were purified using affinity chromatography as the first step in purification. In case of His_6-SUMO-CPA major fraction of protein was cleaved by ULP1 protease to obtain CPA for protein crystallization experiments. Cleavage was successful, but still some small portion of protein remained uncleaved. After all, this contaminant was later eliminated as next round of affinity chromatography was done, where both cleaved His_6-SUMO tag as well as uncelaved portion of protein were captured onto Ni-NTA resin. This process also purified CPA protein itself from other impurities.

As the second step in protein purification, gel permeation chromatography was selected as final method. In case of GST-CPA protein, main goal was to eliminate small impurities as well as buffer exchange from an eluting agent from the affinity chromatography (reduced L-glutathione). The obtained final fraction was of satisfactory purity for later purposes. In case of CPA after His_6-SUMO cleavage, some protein portion was assembled in the multimeric state (small first peak in Figure 26 on page 69). This effect wasn't anticipated since multimerization of heterotrimeric influenza polymerase was not observed. It may have been due to lower stability of CPA[132] that protein compensate for this effect by multimerization. However, introduction of reducing agent (TCEP) in running buffer used in GPC was enough to suppress further multimerization after isolation of multimeric fraction of protein. Furthermore, protein purity after GPC was >95% which should be sufficient for crystallization experiments.

Before using GST-CPA recombinant protein construct, we wanted to evaluate if the protein is able to bind native PB1 peptide. We chose SPR as a reasonable quick analytic method. First, we wanted to simply test this binding. However, non-specific binding to anchoring neutravidin layer was evident so further precautions had to be made. Detergent was introduced into sample buffer and this effect was diminished. This information was also valuable for later AlphaScreen experiments as detergent helps to decrease non-specific binding of tested compounds during HTS. Moreover, dissociation constant K_d of 23 ± 6 nM was determined by SPR. This value was consistent with previously published results[134] and confirmed possible use of protein in HTS. Furthermore, testing of binding of GST protein was performed and neither binding to PB1 peptide or neutravidin was monitored refuting influence on CPA-PB1 interaction.

Evaluation of GST-CPA binding allowed us to start AlphaScreen HTS development. We had initially chosen this assay as it was previously frequently used in our laboratory for screening the inhibitors of protein-protein interaction driving HIV capsid assembly[153]. We proceeded to standard experiment in which wide set of concentrations of both binding partners was tested in the cross-titration experiment. As you can see in Figure 32 on page 73 gradually more fluorescent signal was received with an increasing concentrations of both interacting partners. At one point signal started declining with the further increase in concentration. That is due to saturation when further increase of concentration interferes with unbound interacting molecules. Finally cross-point, where the highest signal was obtained, was selected as the optimal condition. At this best condition, we obtained a signal of more than 200,000 counts which is excellent signal for this assay. Background of both type of beads in the absence of binding partners had average signal of 360 counts. That means that wide range between positive signal and background signal was achieved. Moreover, statistical evaluation of this assay was determined and Z′ value (standard value for determination of HTS quality[149]) was calculated as 0.97 which is rated as an excellent assay. Signal to noise ratio of 19,000 is a satisfactory result. Lastly, signal to background ratio of more than 600 fold difference is more than pleasing. Taking all these results into account, the assay was very well optimized and evaluated, and was ready for further screening.

One of our main aims was to identify a minimal PB1 binding peptide that would still inhibit in low nanomolar range as wild-type 14 amino acid PB1 peptide. We obtained

11 peptides truncated sequentially from both N-terminus and C-terminus. We titrated all these peptides in the AlphaScreen assay. Results are shown in Figure 35 on page 75 documenting that when the peptide is truncated from C-terminus no significant effect on the binding occurs until reaching isolueucin next to the helical part of the peptide (highlighted in grey). Elimination of isoleucine (PB1-3) results in more than 10-fold decrease in binding affinity. Further elimination of AAs implemented in helix results in significant decrease in binding as an example of the PB1-5 peptide has more than 1,000-fold decreased binding affinity. Truncation from the N-terminus shows more dramatic effect than from C-terminus. Exclusion of first two AAs (methionine, aspartate) effects binding by 65-fold decrease. Interestingly removing valine in possition 3 of the peptide results in extreme reduction of binding affinity (more than 3000-fold). This suggest high importance of this residue to binding. Lastly, exclusion of all preceding AAs of helix results in distinct impairment of binding affinity towards CPA (PB1-9). Based on that, we have chosen decapeptide PB1-10 as one that was both truncated substantially by the elimination of four amino acids as well as it retained binding affinity although still decreased by 560-times. Based on one publication where two possible mutations within peptide for tyrosines after protein array were proposed to increase binding to CPA[135], we decided to test this decapeptide containing these mutations (PB1-11) and surprisingly this peptide showed only two-fold decreased binding affinity compared to the original PB1-0 tetradecapeptide. In addition, this decapeptide in comparison to PB1-10 peptide with WT sequence has 260-fold difference in binding affinity suggesting a key role of introduced tyrosines in the binding mechanism.

We wanted to further investigate the effect of mutated tyrosines in PB1-11 peptide on binding and thus we started the crystallization experiments. I was lucky to identify one condition where crystals grew. Since multiple crystals were present in this condition we decided to test the best one and use others for seeding experiments. Tested crystal diffracted poorly so the detection of better condition was necessary. Seeding experiments were of great success as multiple new conditions appeared with grown crystals. I hand-optimized few conditions where overall best crystals were found. Finally, multiple crystals were measured at synchrotron facility and best crystal diffracted up to the resolution of 1.6 Å. With help of Petr Pachl Ph.D. from the IOCB of the CAS, the crystal structure was solved with the method of molecular replacement. As shown in Figure 37

on page 77, implemented unnatural tyrosines create additional hydrogen bond network between water molecules and amino acid side chains of CPA. This newly formed bonds are most probably the reason for significantly improved binding affinity compared to wt PB1-10 decapeptide. This information was of significant value for us as we want to use such information in rational drug design.

Last part of this work focused on development and evaluation of another HTS assay based on DIANA technology[150]. DIANA technology is more favourable with its high hit discovery rate, lower number of unspecific hits and is easier to upscale compared to AlphaScreen method. The pivotal stage of DIANA HTS development was set by my consultant Milan Kožíšek Ph.D. and then I assisted in final assay evaluation experiments as described in the Methods section. To possibly utilize DIANA assay in HTS, we needed to determine K_d value of our DIANA probe. Form previous experiments, 1,000 pg of targeted protein per well seemed as optimal as it was anticipated to be still in the linear range of the assay. We confirmed this hypothesis and subsequently, we have chosen 1,000 pg of protein per well to be optimal amount for the assay as it was still in linear range and also had a better signal to background ratio than the lower amounts previously tested. Lastly, we wanted to evaluate if our results of obtained binding affinities from the AlphaScreen assay are in agreement with the data received from DIANA assay. We tested PB1-0 (MDVNPTLLFLKIPA) peptide as a primary inhibitor. We focused mainly on the determination of the effect of DMSO concentration on inhibitor binding since most small molecule compound libraries are desolved in DMSO and a percentage of DMSO up to 10% may be necessary in the HTS to prevent precipitation of the compounds in the assay. We needed to evaluate if this concentration of DMSO effects peptide binding. As shown in Figure 39 page 79 no significant effect was observed as all K_i values were in the same range. Furthermore, K_i values for PB1-0 peptide were in approximately 10-fold difference (AlphaScreen peptide titration Figure 35 on page 75) confirming consistiency with the IC_{50} value for PB1-0 peptide determined by the AlphaScreen assay. Furthermore K_d of probe was consistent with K_i values of unconjugated PB1-0 peptide confirming good design of DIANA probe. Signal to background difference is approximately 6 cycles which is an acceptable value for further HTS. Final DIANA assay evaluation is still in progress by the time of writing this thesis and it will be soon used for library screening.

6. Conclusion

Successful cloning and later purification of two recombinant protein constructs of the C-terminal domain of influenza polymerase acidic subunit were performed. A substantial amounts of proteins of required purity were obtained in two-step purification process. Ability of GST-CPA to bind PB1 peptide was tested and confirmed. First GST-CPA fusion construct was used in the development of high-throughput screening assay based on the AlphaScreeen technology. The assay was optimized and evaluated as an excellent assay for screening of inhibitors. We tested truncated versions of polymerase basic subunit 1 N-terminal part on binding to CPA protein. We identified decapeptide, a low micromolar inhibitor as well as decapeptide bearing two mutations that showed same binding affinity as original tetradecapeptide. Finally, we obtained crystal structure of CPA binding this mutated decapeptide and confirmed the additional effect of introduced tyrosines in peptide binding mechanism. Lastly, we wanted to design second HTS as a method for final screening of compound libraries. I was able to perform few final test experiments directing assay towards final evaluation.

References

1. World Health Organization. Influenza - seasonal. (2015). Available at: http://www.who.int/mediacentre/factsheets/fs211/en/. (Accessed: 15th January 2018)

2. Webster, G. R., Monto, S. A., Braciale, J. T. & Lamb, A. R. *Textbook of Influenza*. (Wiley Blackwell, 2013).

3. Rossman, J. & Lamb, R. Influenza virus assembly and budding. *Virology* **411,** 229–236 (2011).

4. Centers for Disease Control and Prevention. Influenza (Flu). Available at: https://www.cdc.gov/flu/about/index.html. (Accessed: 15th January 2018)

5. Nelson, M. I. & Holmes, E. C. The evolution of epidemic influenza. *Nat. Rev. Genet.* **8,** 196–205 (2007).

6. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56,** 152–79 (1992).

7. Ferguson, N. M., Galvani, A. P. & Bush, R. M. Ecological and immunological determinants of influenza evolution. *Nature* **422,** 428–433 (2003).

8. Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7712–8 (1997).

9. Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J. & Fitch, W. M. Predicting the evolution of human influenza A. *Science* **286,** 1921–5 (1999).

10. Lindstrom, S. E., Cox, N. J. & Klimov, A. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* **328,** 101–119 (2004).

11. Potter, C. W. A history of influenza. *J. Appl. Microbiol.* 572–579 (2001).

12. Taubenberger, J. K., Reid, A. H., Krafft, A. E., Bijwaard, K. E. & Fanning, T. G. Initial genetic characterization of the 1918 'Spanish' influenza virus. *Science* **275,** 1793–6 (1997).

13. Olson, D. R., Simonsen, L., Edelson, P. J. & Morse, S. S. Epidemiological evidence of an early wave of the 1918 influenza pandemic in New York City. *Proc. Natl. Acad. Sci.* **102,** 11059–11063 (2005).

14. Farrar, J. *et al.* Avian Influenza A (H5N1) Infection in Humans. *N. Engl. J. Med.* **353,** 1374–1385 (2005).

15. Morens, D. M., Taubenberger, J. K. & Fauci, A. S. Predominant Role of Bacterial Pneumonia as a Cause of Death in Pandemic Influenza: Implications for Pandemic Influenza Preparedness. *J. Infect. Dis.* **198,** 962–970 (2008).

16. de Jong, J. C., Claas, E. C., Osterhaus, A. D., Webster, R. G. & Lim, W. L. A pandemic warning? *Nature* **389,** 554 (1997).

17. Bridges, C. B. *et al.* Risk of Influenza A (H5N1) Infection among Poultry Workers, Hong Kong, 1997–1998. *J. Infect. Dis.* **185,** 1005–1010 (2002).

18. Li, K. S. *et al.* Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* **430,** 209–213 (2004).

19. Hien, T. T., de Jong, M. & Farrar, J. Avian Influenza — A Challenge to Global Health Care Structures. *N. Engl. J. Med.* **351,** 2363–2365 (2004).

20. Wang, H. *et al.* Probable limited person-to-person transmission of highly pathogenic avian influenza A (H5N1) virus in China. *Lancet* **371,** 1427–1434 (2008).

21. Centers for Disease Control and Prevention. Community Mitigation Guidelines to Prevent Pandemic Influenza — United States, 2017. Available at: https://www.cdc.gov/mmwr/volumes/66/rr/rr6601a1.htm. (Accessed: 16th January 2018)

22. Viboud, C., Miller, M., Olson, D., Osterholm, M. & Simonsen, L. Preliminary Estimates of Mortality and Years of Life Lost Associated with the 2009 A/H1N1 Pandemic in the US and Comparison with Past Influenza Seasons. *PLoS Curr.* **2,** (2010).

23. Roberts, P. C. & Compans, R. W. Host cell dependence of viral morphology. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 5746–5751 (1998).

24. Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H. & Barman, S. Influenza virus morphogenesis and budding. *Virus Res.* **143,** 147–161 (2009).

25. Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci.* **103,** 19123–19127 (2006).

26. Nayak, D. P., Hui, E. K.-W. & Barman, S. Assembly and budding of influenza virus. *Virus Res.* **106,** 147–165 (2004).

27.   Noda, T. & Kawaoka, Y. Structure of influenza virus ribonucleoprotein complexes and their packaging into virions. *Rev. Med. Virol.* **20,** 380–391 (2010).

28.   Shi, Y., Wu, Y., Zhang, W., Qi, J. & Gao, G. F. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* **12,** 822–831 (2014).

29.   Wise, H. M. *et al.* A Complicated Message: Identification of a Novel PB1-Related Protein Translated from Influenza A Virus Segment 2 mRNA. *J. Virol.* **83,** 8021–8031 (2009).

30.   Compans, R. W., Content, J. & Duesberg, P. H. Structure of the ribonucleoprotein of influenza virus. *J. Virol.* **10,** 795–800 (1972).

31.   Jagger, B. W. *et al.* An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response. *Science (80-.).* **337,** 199–204 (2012).

32.   Lamb, R. A. & Choppin, P. W. Segment 8 of the influenza virus genome is unique in coding for two polypeptides. *Proc. Natl. Acad. Sci.* **76,** 4908–4912 (1979).

33.   Lamb, R. a, Choppin, P. W., Chanock, R. M. & Lai, C. J. Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. *Proc. Natl. Acad. Sci.* **77,** 1857–1861 (1980).

34.   Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* **26 Suppl 4,** D49-53 (2008).

35.   Ha, Y. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. *EMBO J.* **21,** 865–875 (2002).

36.   Gamblin, S. J. & Skehel, J. J. Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. *J. Biol. Chem.* **285,** 28403–28409 (2010).

37.   Wiley, D. C. & Skehel, J. J. The Structure and Function of the Hemagglutinin Membrane Glycoprotein of Influenza Virus. *Annu. Rev. Biochem.* **56,** 365–394 (1987).

38.   Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl. Acad. Sci.* **98,** 11181–11186 (2001).

39.   Skehel, J. J. & Wiley, D. C. Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin. *Annu. Rev. Biochem.* **69,** 531–569 (2000).

40. Rogers, G. N. & Paulson, J. C. Receptor determinants of human and animal influenza virus isolates: Differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology* **127,** 361–373 (1983).

41. Russell, R. J. *et al.* The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **443,** 45–49 (2006).

42. Varghese, J. N. & Colman, P. M. Three-dimensional structure of the neuraminidase of influenza virus A/Tokyo/3/67 at 2.2 A resolution. *J. Mol. Biol.* **221,** 473–86 (1991).

43. Shtyrya, Y. A., Mochalova, L. V & Bovin, N. V. Influenza virus neuraminidase: structure and function. *Acta Naturae* **1,** 26–32 (2009).

44. Bossart-Whitaker, P. *et al.* Three-dimensional Structure of Influenza A N9 Neuraminidase and Its Complex with the Inhibitor 2-Deoxy 2,3-Dehydro-N-Acetyl Neuraminic Acid. *J. Mol. Biol.* **232,** 1069–1083 (1993).

45. Colman, P. M., Hoyne, P. A. & Lawrence, M. C. Sequence and structure alignment of paramyxovirus hemagglutinin-neuraminidase with influenza virus neuraminidase. *J. Virol.* **67,** 2972–80 (1993).

46. Colman, P. M., Varghese, J. N. & Laver, W. G. Structure of the catalytic and antigenic sites in influenza virus neuraminidase. *Nature* **303,** 41–4 (1983).

47. Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. a & Klenk, H. Neuraminidase Is Important for the Initiation of Influenza Virus Infection in Human Airway Epithelium. *Society* **78,** 12665–12667 (2004).

48. Wagner, R., Wolff, T., Herwig, A., Pleschka, S. & Klenk, H.-D. Interdependence of Hemagglutinin Glycosylation and Neuraminidase as Regulators of Influenza Virus Growth: a Study by Reverse Genetics. *J. Virol.* **74,** 6316–6323 (2000).

49. Lamb, R. A., Zebedee, S. L. & Richardson, C. D. Influenza virus M2 protein is an integral membrane protein expressed on the infected-cell surface. *Cell* **40,** 627–33 (1985).

50. Schroeder, C., Heider, H., Moncke-Buchner, E. & Lin, T.-I. The influenza virus ion channel and maturation cofactor M2 is a cholesterol-binding protein. *Eur. Biophys. J.* **34,** 52–66 (2005).

51. Pielak, R. M. & Chou, J. J. Influenza M2 proton channels. *Biochim. Biophys. Acta* **1808,** 522–9 (2011).

52. Thomaston, J. L. *et al.* High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc. Natl. Acad. Sci.* **112,** 14260–14265 (2015).

53. te Velthuis, A. J. W. & Fodor, E. Influenza virus RNA polymerase: insights into the mechanisms of viral RNA synthesis. *Nat. Rev. Microbiol.* **14,** 479–493 (2016).

54. Kapoor, S. & Dhama, K. *Insight into Influenza Viruses of Animals and Humans.* (Springer International Publishing, 2014).

55. Ming, L. *Viral Molecular Machines. Advences in Experimental Medicine and Biology* **726,** (Springer US, 2012).

56. Kemble, G. W., Danieli, T. & White, J. M. Lipid-anchored influenza hemagglutinin promotes hemifusion, not complete fusion. *Cell* **76,** 383–391 (1994).

57. Bonnafous, P. & Stegmann, T. Membrane Perturbation and Fusion Pore Formation in Influenza Hemagglutinin-mediated Membrane Fusion. *J. Biol. Chem.* **275,** 6160–6166 (2000).

58. Pinto, L. H. & Lamb, R. A. The M2 Proton Channels of Influenza A and B Viruses. *J. Biol. Chem.* **281,** 8997–9000 (2006).

59. Bouloy, M., Plotch, S. J. & Krug, R. M. Globin mRNAs are primers for the transcription of influenza viral RNA in vitro. *Proc. Natl. Acad. Sci.* **75,** 4886–4890 (1978).

60. Plotch, S. J., Bouloy, M., Ulmanen, I. & Krug, R. M. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell* **23,** 847–858 (1981).

61. Poon, L. L., Pritlove, D. C., Fodor, E. & Brownlee, G. G. 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,** 3473–6 (1999).

62. Ulmanen, I., Broni, B. A. & Krug, R. M. 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.* **78,** 7355–7359 (1981).

63. Engelhardt, O., Smith, M. & Fodor, E. Association of the influenza A virus RNA-dependent RNA polymerase with cellular RNA polymerase II. *J. Virol.* **79,** 5812–5818 (2005).

64. Yuan, P. *et al.* Crystal structure of an avian influenza polymerase PAN reveals an

endonuclease active site. *Nature* **458,** 909–913 (2009).

65. Dias, A. *et al.* The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* **458,** 914–918 (2009).

66. Guilligay, D. *et al.* The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.* **15,** 500–506 (2008).

67. Smith, G. L. & Hay, A. J. Replication of the influenza virus genome. *Virology* **118,** 96–108 (1982).

68. Shapiro, G. I., Gurney, T. & Krug, R. M. Influenza virus gene expression: control mechanisms at early and late times of infection and nuclear-cytoplasmic transport of virus-specific RNAs. *J. Virol.* **61,** 764–73 (1987).

69. Kundu, A., Avalos, R. T., Sanderson, C. M. & Nayak, D. P. Transmembrane domain of influenza virus neuraminidase, a type II protein, possesses an apical sorting signal in polarized MDCK cells. *J. Virol.* **70,** 6508–15 (1996).

70. Lin, S., Naim, H. Y., Chapin Rodriguez, A. & Roth, M. G. Mutations in the Middle of the Transmembrane Domain Reverse the Polarity of Transport of the Influenza Virus Hemagglutinin in MDCK Epithelial Cells. *J. Cell Biol.* **142,** 51–57 (1998).

71. Elton, D. *et al.* Interaction of the Influenza Virus Nucleoprotein with the Cellular CRM1-Mediated Nuclear Export Pathway. *J. Virol.* **75,** 408–419 (2001).

72. Neumann, G. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. *EMBO J.* **19,** 6751–6758 (2000).

73. Gomez-Puertas, P., Albo, C., Perez-Pastrana, E., Vivo, A. & Portela, A. Influenza Virus Matrix Protein Is the Major Driving Force in Virus Budding. *J. Virol.* **74,** 11538–11547 (2000).

74. Huang, X., Liu, T., Muller, J., Levandowski, R. A. & Ye, Z. Effect of Influenza Virus Matrix Protein and Viral RNA on Ribonucleoprotein Formation and Nuclear Export. *Virology* **287,** 405–416 (2001).

75. Akarsu, H. Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2). *EMBO J.* **22,** 4646–4655 (2003).

76. Barman, S., Ali, A., Hui, E. K. W., Adhikary, L. & Nayak, D. P. Transport of viral proteins to the apical membranes and interaction of matrix protein with glycoproteins in the assembly

of influenza viruses. *Virus Res.* **77,** 61–69 (2001).

77. Avalos, R. T., Yu, Z. & Nayak, D. P. Association of influenza virus NP and M1 proteins with cellular cytoskeletal elements in influenza virus-infected cells. *J. Virol.* **71,** 2947–58 (1997).

78. Bui, M., Wills, E. G., Helenius, A. & Whittaker, G. R. Role of the Influenza Virus M1 Protein in Nuclear Export of Viral Ribonucleoproteins. *J. Virol.* **74,** 1781–1786 (2000).

79. Barman, S. *et al.* Role of transmembrane domain and cytoplasmic tail amino acid sequences of influenza a virus neuraminidase in raft association and virus budding. *J. Virol.* **78,** 5258–69 (2004).

80. Ison, M. G. & Hayden, F. G. Therapeutic options for the management of influenza. *Curr. Opin. Pharmacol.* **1,** 482–90 (2001).

81. Mould, J. A. *et al.* Influenza B virus BM2 protein has ion channel activity that conducts protons across membranes. *Dev. Cell* **5,** 175–84 (2003).

82. Hayden, F. G. Antivirals for influenza: Historical perspectives and lessons learned. *Antiviral Res.* **71,** 372–378 (2006).

83. Deyde, V. M. *et al.* Surveillance of Resistance to Adamantanes among Influenza A(H3N2) and A(H1N1) Viruses Isolated Worldwide. *J. Infect. Dis.* **196,** 249–257 (2007).

84. von Itzstein, M. The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discov.* **6,** 967–974 (2007).

85. Chong, A. K., Pegg, M. S., Taylor, N. R. & von Itzstein, M. Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* **207,** 335–43 (1992).

86. Varghese, J. N., Laver, W. G. & Colman, P. M. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature* **303,** 35–40 (1983).

87. von Itzstein, M. *et al.* Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **363,** 418–423 (1993).

88. De Clercq, E. Antiviral agents active against influenza A viruses. *Nat. Rev. Drug Discov.* **5,** 1015–1025 (2006).

89. Kim, C. U. *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.* **119,** 681–90 (1997).

90. Moscona, A. Neuraminidase Inhibitors for Influenza. *N. Engl. J. Med.* **353,** 1363–1373 (2005).

91. Hurt, A. C. *et al.* Antiviral resistance during the 2009 influenza A H1N1 pandemic: public health, laboratory, and clinical perspectives. *Lancet Infect. Dis.* **12,** 240–248 (2012).

92. Le, Q. M. *et al.* Isolation of drug-resistant H5N1 virus. *Nature* **437,** 1108–1108 (2005).

93. Herlocher, M. L. *et al.* Influenza Viruses Resistant to the Antiviral Drug Oseltamivir: Transmission Studies in Ferrets. *J. Infect. Dis.* **190,** 1627–1630 (2004).

94. Collins, P. J. *et al.* Structural basis for oseltamivir resistance of influenza viruses. *Vaccine* **27,** 6317–6323 (2009).

95. Yen, H.-L. *et al.* Neuraminidase inhibitor-resistant influenza viruses may differ substantially in fitness and transmissibility. *Antimicrob. Agents Chemother.* **49,** 4075–84 (2005).

96. Gupta, P. *et al.* Preclinical pharmacokinetics of MHAA4549A, a human monoclonal antibody to influenza A virus, and the prediction of its efficacious clinical dose for the treatment of patients hospitalized with influenza A. *MAbs* **8,** 991–997 (2016).

97. Dreyfus, C. *et al.* Highly Conserved Protective Epitopes on Influenza B Viruses. *Science (80- .).* **337,** 1343–1348 (2012).

98. McKimm-Breschkin, J. L. & Fry, A. M. Meeting report: 4th ISIRV antiviral group conference: Novel antiviral therapies for influenza and other respiratory viruses. *Antiviral Res.* **129,** 21–38 (2016).

99. Naesens, L., Stevaert, A. & Vanderlinden, E. Antiviral therapies on the horizon for influenza. *Curr. Opin. Pharmacol.* **30,** 106–115 (2016).

100. Haffizulla, J. *et al.* Effect of nitazoxanide in adults and adolescents with acute uncomplicated influenza: a double-blind, randomised, placebo-controlled, phase 2b/3 trial. *Lancet Infect. Dis.* **14,** 609–618 (2014).

101. Badani, H., Garry, R. F. & Wimley, W. C. Peptide entry inhibitors of enveloped viruses: The importance of interfacial hydrophobicity. *Biochim. Biophys. Acta - Biomembr.* **1838,** 2180–2197 (2014).

102. Chan-Tack, K. M., Murray, J. S. & Birnkrant, D. B. Use of Ribavirin to Treat Influenza. *N. Engl. J. Med.* **361,** 1713–1714 (2009).

103. Tompkins, S. M., Lo, C.-Y., Tumpey, T. M. & Epstein, S. L. Protection against lethal influenza virus challenge by RNA interference in vivo. *Proc. Natl. Acad. Sci.* **101,** 8682–8686 (2004).

104. Ge, Q. *et al.* Inhibition of influenza virus production in virus-infected mice by RNA interference. *Proc. Natl. Acad. Sci.* **101,** 8676–8681 (2004).

105. Malakhov, M. P. *et al.* Sialidase Fusion Protein as a Novel Broad-Spectrum Inhibitor of Influenza Virus Infection. *Antimicrob. Agents Chemother.* **50,** 1470–1479 (2006).

106. Gambaryan, A. S. *et al.* Polymeric inhibitor of influenza virus attachment protects mice from experimental influenza infection. *Antiviral Res.* **55,** 201–5 (2002).

107. Pflug, A., Guilligay, D., Reich, S. & Cusack, S. Structure of influenza A polymerase bound to the viral RNA promoter. *Nature* **516,** 355–360 (2014).

108. Furuta, Y. *et al.* Mechanism of Action of T-705 against Influenza Virus. *Antimicrob. Agents Chemother.* **49,** 981–986 (2005).

109. Furuta, Y. *et al.* Favipiravir (T-705), a novel viral RNA polymerase inhibitor. *Antiviral Res.* **100,** 446–454 (2013).

110. Pautus, S. *et al.* New 7-Methylguanine Derivatives Targeting the Influenza Polymerase PB2 Cap-Binding Domain. *J. Med. Chem.* **56,** 8915–8930 (2013).

111. Clark, M. P. *et al.* Discovery of a Novel, First-in-Class, Orally Bioavailable Azaindole Inhibitor (VX-787) of Influenza PB2. *J. Med. Chem.* **57,** 6668–6678 (2014).

112. Byrn, R. A. *et al.* Preclinical Activity of VX-787, a First-in-Class, Orally Bioavailable Inhibitor of the Influenza Virus Polymerase PB2 Subunit. *Antimicrob. Agents Chemother.* **59,** 1569–1582 (2015).

113. Tomassini, J. *et al.* Inhibition of cap (m7GpppXm)-dependent endonuclease of influenza virus by 4-substituted 2,4-dioxobutanoic acid compounds. *Antimicrob. Agents Chemother.* **38,** 2827–37 (1994).

114. Stevaert, A. *et al.* Mutational Analysis of the Binding Pockets of the Diketo Acid Inhibitor L-742,001 in the Influenza Virus PA Endonuclease. *J. Virol.* **87,** 10524–10538 (2013).

115.   Yuan, S. *et al.* Cross-Protection of Influenza A Virus Infection by a DNA Aptamer Targeting the PA Endonuclease Domain. *Antimicrob. Agents Chemother.* **59,** 4082–4093 (2015).

116.   Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. *Nat. Rev. Drug Discov.* **15,** 533–550 (2016).

117.   Skwarczynska, M. & Ottmann, C. Protein–protein interactions as drug targets. *Future Med. Chem.* **7,** 2195–2219 (2015).

118.   Bogan, A. A. & Thorn, K. S. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280,** 1–9 (1998).

119.   Clackson, T. & Wells, J. A hot spot of binding energy in a horomone-receptor interface. *Science (80-.).* **267,** 383–386 (1995).

120.   Jones, S. & Thornton, J. M. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **93,** 13–20 (1996).

121.   Miller, S. The structure of interfaces between subunits of dimeric and tetrameric proteins. *"Protein Eng. Des. Sel.* **3,** 77–83 (1989).

122.   Blundell, T. L. *et al.* Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. *Philos. Trans. R. Soc. B Biol. Sci.* **361,** 413–423 (2006).

123.   de Vega, M. J. P., Martín-Martínez, M. & González-Muñiz, R. Modulation of protein-protein interactions by stabilizing/mimicking protein secondary structure elements. *Curr. Top. Med. Chem.* **7,** 33–62 (2007).

124.   Henchey, L. K., Jochim, A. L. & Arora, P. S. Contemporary strategies for the stabilization of peptides in the α-helical conformation. *Curr. Opin. Chem. Biol.* **12,** 692–697 (2008).

125.   Fletcher, S. & Hamilton, A. D. Protein surface recognition and proteomimetics: mimics of protein surface structure and function. *Curr. Opin. Chem. Biol.* **9,** 632–638 (2005).

126.   Pérez, D. R. & Donis, R. O. A 48-amino-acid region of influenza A virus PB1 protein is sufficient for complex formation with PA. *J. Virol.* **69,** 6932–9 (1995).

127.   Perez, D. R. & Donis, R. O. Functional analysis of PA binding by influenza a virus PB1: effects on polymerase activity and viral infectivity. *J. Virol.* **75,** 8127–36 (2001).

128. Ghanem, A. *et al.* Peptide-Mediated Interference with Influenza A Virus Polymerase. *J. Virol.* **81,** 7801–7804 (2007).

129. He, X. *et al.* Crystal structure of the polymerase PAC–PB1N complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).

130. Obayashi, E. *et al.* The structural basis for an essential subunit interaction in influenza virus RNA polymerase. *Nature* **454,** 1127–1131 (2008).

131. He, X. *et al.* Crystal structure of the polymerase PAC–PB1N complex from an avian influenza H5N1 virus. *Nature* **454,** 1123–1126 (2008).

132. Moen, S. O. *et al.* Structural analysis of H1N1 and H7N9 influenza A virus PA in the absence of PB1. *Sci. Rep.* **4,** 5944 (2015).

133. Rognan, D. Rational design of protein–protein interaction inhibitors. *Medchemcomm* **6,** 51–60 (2015).

134. Wunderlich, K. *et al.* Identification of a PA-Binding Peptide with Inhibitory Activity against Influenza A and B Virus Replication. *PLoS One* **4,** e7517 (2009).

135. Wunderlich, K. *et al.* Identification of High-Affinity PB1-Derived Peptides with Enhanced Affinity to the PA Protein of Influenza A Virus Polymerase. *Antimicrob. Agents Chemother.* **55,** 696–702 (2011).

136. Muratore, G. *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.* **109,** 6247–6252 (2012).

137. Fukuoka, M. *et al.* Structure-based discovery of anti-influenza virus A compounds among medicines. *Biochim. Biophys. Acta - Gen. Subj.* **1820,** 90–95 (2012).

138. Kessler, U. *et al.* Discovery and synthesis of novel benzofurazan derivatives as inhibitors of influenza A virus. *Bioorg. Med. Chem. Lett.* **23,** 5575–5577 (2013).

139. Massari, S. *et al.* Structural Investigation of Cycloheptathiophene-3-carboxamide Derivatives Targeting Influenza Virus Polymerase Assembly. *J. Med. Chem.* **56,** 10118–10131 (2013).

140. Pagano, 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* **9,** 129–150 (2014).

141.   Trist, I. M. L. *et al.* 4,6-Diphenylpyridines as Promising Novel Anti-Influenza Agents Targeting the PA–PB1 Protein–Protein Interaction: Structure–Activity Relationships Exploration with the Aid of Molecular Modeling. *J. Med. Chem.* **59,** 2688–2703 (2016).

142.   Tintori, C. *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.* **24,** 280–282 (2014).

143.   Massari, S. *et al.* A Broad Anti-influenza Hybrid Small Molecule That Potently Disrupts the Interaction of Polymerase Acidic Protein–Basic Protein 1 (PA-PB1) Subunits. *J. Med. Chem.* **58,** 3830–3842 (2015).

144.   Hegnerová, K. *et al.* Surface plasmon resonance biosensors for detection of Alzheimer disease biomarker. *Sensors Actuators B Chem.* **139,** 69–73 (2009).

145.   pGEX-1λT vector. Available at: http://www.snapgene.com/resources/plasmid__files/pgex_vectors_(ge_healthcare)/pGEX-1_lambda_T/. (Accessed: 14th April 2018)

146.   pETM11-SUMO3GFP vector. Available at: https://www.embl.de/pepcore/pepcore_services/cloning/sumo/. (Accessed: 14th April 2018)

147.   NEBioCalculator. Available at: http://nebiocalculator.neb.com/#!/ligation. (Accessed: 20th November 2016)

148.   PerkinElmer Inc. From Small Molecules To Live Cells - The Assay That Measures It All. (2013).

149.   Zhang, J.-H. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. *J. Biomol. Screen.* **4,** 67–73 (1999).

150.   Navrátil, V. *et al.* DNA-linked Inhibitor Antibody Assay (DIANA) for sensitive and selective enzyme detection and inhibitor screening. *Nucleic Acids Res.* **45,** e10–e10 (2017).

151.   Šácha, P. *et al.* iBodies: Modular Synthetic Antibody Mimetics Based on Hydrophilic Polymers Decorated with Functional Moieties. *Angew. Chem. Int. Ed. Engl.* **55,** 2356–60 (2016).

152.   Presolski, S. I., Hong, V. P. & Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Curr. Protoc. Chem. Biol.* **3,** 153–162 (2011).

153.   Machara, A. *et al.* Specific Inhibitors of HIV Capsid Assembly Binding to the C-Terminal

Domain of the Capsid Protein: Evaluation of 2-Arylquinazolines as Potential Antiviral Compounds. *J. Med. Chem.* **59,** 545–558 (2016).