RaPID selection of macrocyclic peptide inhibitors for influenza A hemagglutinin proteins

By Stein Vermeir

Examiner: Seino A. K. Jongkees
Second reviewer: Xander C. A. M. de Haan

Abstract

Of the influenza viruses, it is solely influenza A viruses (IAVs) that are responsible for flu pandemics. They are quickly gaining resistances against approved drugs targeting the endonuclease subdomain of the RNA-dependent RNA polymerase, neuraminidase or matrix protein. Hemagglutinin is a challenging target and has only one approved drug, arbidol. Arbidol and some monoclonal antibodies are broad-acting inhibitors of viral infection, inhibiting multiple hemagglutinin subtypes. Therefore, hemagglutinin was selected as target for a combinational approach of the mRNA display and flexible *in vitro* translation system, called Random nonstandard Peptide Integrated Discovery (RaPID) to find inhibitors. This system allows for the creation of mRNA-linked peptide libraries of over 10^12 sequences, by including non-canonical amino acids. By starting each peptide with an N-terminal chloroacetylated tyrosine and incorporating at least one downstream cysteine residue, spontaneous macrocyclization occurs, improving peptide stability and target-binding. The library is used for iterative selection cycles where only peptides binding the target protein continue to the next round. In an earlier RaPID selection the hemagglutinin (H1) protein of the pandemic H1N1 IAV isolate Neth09 was used. This process yielded a potent inhibitory peptide, which was later used to drive the virus to resist this peptide, finding a mutation in H1 from isoleucine 375 to phenylalanine (manuscript under review). In this project, a stem only construct of the mutant H1 protein was used for a second selection, aiming to find an inhibitor for the H1 mutant or a broad-acting inhibitor for both the wild type and mutant. After sequencing of the enriched libraries and motif identification through bioinformatic analysis, 17 peptides were selected for solid phase peptide synthesis and tested for activity in a luciferase assay of influenza virus infection. 2 peptides were found that inhibit the H1 mutant, and 5 that inhibit H1 wild type. Only 1 peptide was found to show some cross reactivity, partially inhibiting H1 mutant in addition to full inhibition of H1 wild type. So far, a combination of two peptides is best suited to prevent Neth09 IAV infection.

Layman's summary

Influenza viruses cause the flu in humans. There have been multiple flu pandemics in recent history, initiating a search for drugs to prevent infection and thus the flu. Multiple drugs were created and approved for human use, however, evolution of the virus alongside the drugs quickly resulted in decreasing effectiveness. Only one drug has been approved for the hemagglutinin protein present on the surface of influenza, which is vital for infection of hosts. It is difficult to find drugs for this protein because the outside, which rapidly mutates, shields the functional core of the protein, preventing drugs from working. Therefore, hemagglutinin is the target of this project where the RaPID system was used to find small, randomly generated amino acid sequences (called peptides) that prevent its function. In the RaPID system over 10 trillion peptides can be created at once and all are exposed to hemagglutinin simultaneously. Every peptide is directly attached to the RNA that carries its genetic information, so when peptides binding the hemagglutinin protein are found, the RNA can be used to identify the peptides. Peptides were then tested for their ability to prevent influenza virus infection. 7 peptides were found that fully prevent infection.

List of abbreviations

Beads	MagStrep “type 3” XT beads
BPB	Bromo phenol blue
BSA	Bovine serum albumin
ClAc	Chloroacetyl
Cq	Quantification cycle
DIPEA	*N*,*N*-Diisopropylethylamine
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDDET	2,2’-(ethylenedioxy)diethanethiol
EDTA	Ethylenediamine Tetraacetate Acid
eFX	Enhanced flexizyme
ELISA	Enzyme-linked immunosorbent assay
FIT	Flexible *In vitro* Translation
Fmoc	Fluorenylmethyloxycarbonyl
IPTG	Isopropyl β-D-1-thiogalactopyranoside
H1s	Stem only H1
H5s	Stem only H5
HA or H	Hemagglutinin
HATU	1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluorophosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
I375F	Isoleucine to phenylalanine mutation at amino acid 375 of H1
IAV	Influenza A virus
LB	Luria-Bertani
M2	Membrane protein
MMLV-RT	Reverse transcriptase
MS	Mass spectrometry
NA or N	Neuraminidase
NanoDSF	Nano differential scanning fluorimetry
NEB	New England biolabs
Neth09	H1N1 pandemic influenza isolate A/Netherlands/602/2009
OD_{600}	Optical density at 600 nm light
PAGE	Polyacrylamide gel electrophoresis
PBS(-T)	Phosphate-buffered saline (-tween-20)
PEG	Poly-ethylene glycol
PURE	Protein system Using Recombinant Elements
PyBOP	Benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate
RaPID	Random nonstandard Peptide Integrated Discovery
(q)PCR	(Quantitative) Polymerase Chain Reaction
rcf	Relative centrifugal force
RFU	Relative fluorescence units
SDS	Sodium dodecyl sulfate
SPPS	Solid Phase Peptide Synthesis
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
tRNA	Transfer RNA
WT	Wild type, in this report from the “Neth09” isolate

Table of Contents

Introduction

Influenza

Influenza viruses are in the family of Orthomyxoviridae and are classified in four genera: alpha-, beta-, gamma- and delta influenza viruses, respectively influenza A, B, C and D[1]. A and B are most harmful to humans and cause annual seasonal outbreaks[2], while influenza C is thought to be less severe, although it can lead to hospitalization of children, often under the age of six[3,4]. Influenza D is unable to infect humans, instead infecting pigs and cattle[5].

Every year, influenza A and B viruses cause 3 to 5 million severe flu cases and 290,000 to 650,000 deaths[2]. During non-pandemic years, the number of cases caused by the A and B types is about equal, however, influenza A virus (IAV) causes all pandemics[6,7] and will therefore be the focus of this work. IAV has a reservoir in aquatic birds and affects humans, pigs, dogs, cats, horses, domestic birds, tigers, seals, whales and minks[8]. It has two glycoproteins on its surface, hemagglutinin (HA or H) and neuraminidase (NA or N)[9]. A viral strain is named after the present HA and NA types, of which H1 to H16 and N1 to N9 are relevant for humans, giving a theoretical variance of 144 strains[10]. In aquatic birds 116 of these different HA/NA combinations of have been isolated[11]. To prevent human infection, yearly vaccines are available for people above the age of six months[12] and many domestic birds are culled[11]. People suffering from influenza can develop symptoms including coughs, myalgias, headaches and fevers[12].

Overview of viral structure

The influenza A virion is about 100 nm in diameter and is either spherical or filamentous in shape (figure 1)[9]. It consists of eight single stranded, antisense RNA segments, a lipid-bilayer envelope and nine different proteins: polymerase-basic protein 1 and 2, polymerase-acidic protein, nucleoprotein, HA, NA, matrix protein and membrane protein (M2) (figure 2)[4]. The RNA also encodes nonstructural protein 1 and nuclear export protein, which are not included in the viral particle[4]. The RNA segments are wrapped around nucleoproteins and in complex with polymerase-acidic protein and polymerase-basic protein 1 and 2[13,14]. Together, the polymerase proteins form the RNA-dependent RNA polymerase which is required for both the production of messenger RNA (sense RNA) and RNA segment replication[13]. Matrix protein is largely responsible for the shape[15] and integrity of the virion[16], M2 and HA find their function in entry into a host cell[17,18] and NA plays a role in migration and release of viral particles[19].

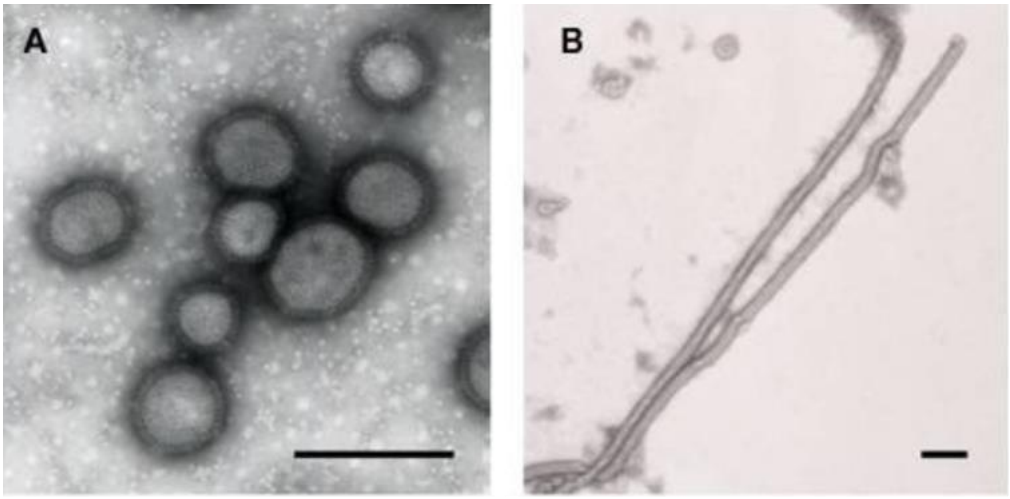


Figure 1: ***Electron microscope pictures of different morphologies of the influenza A virus****. 200 nm bars. (A) H1N1 spherical virions and (B) H3N2 filamentous virions. Figure from Noda 2011*[9]*.*

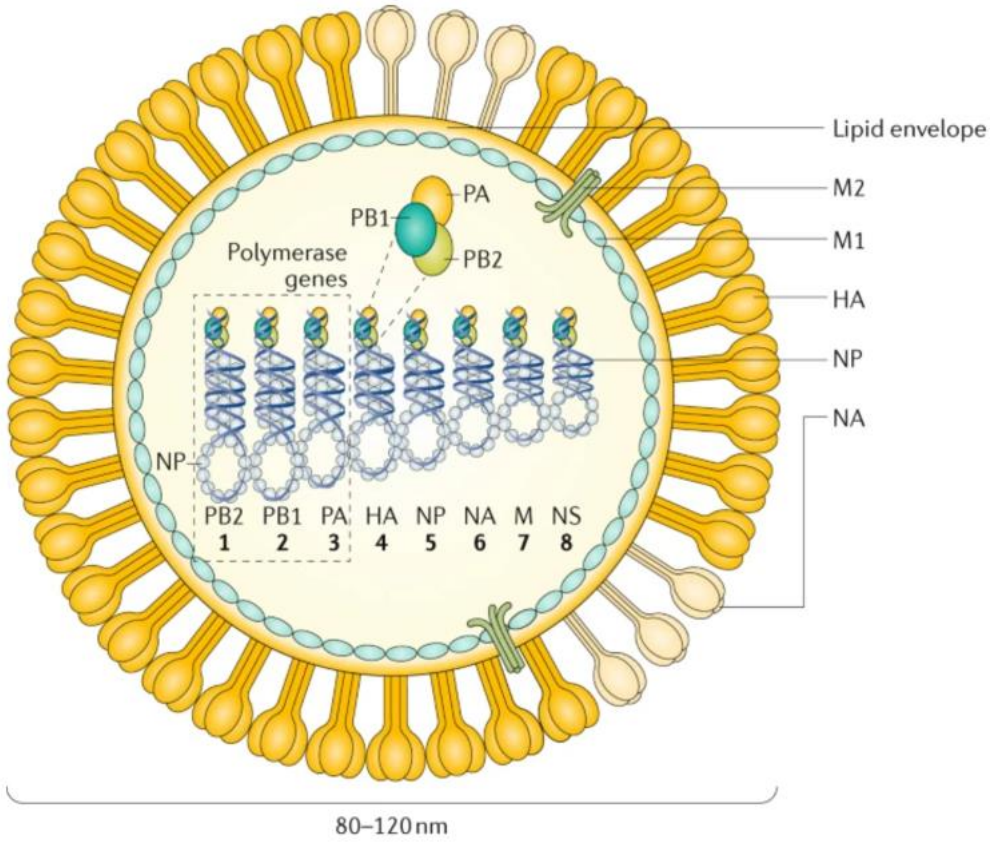


Figure 2: ***Schematic structure of an influenza A virion****. PB1, 2 and PA bind at the 5' and 3' end of a hairpin-shaped NP-coated RNA segment on the inside of the virion, which is protected by a layer of M1 proteins with occasional M2 protein protrusions. Outside the M1 layer is the envelope, derived from the hosts lipid bilayer, where M2, HA and NA pass through to the outer environment. Underneath the RNA segments is shown which genes they encode, 7 encodes both M1 and M2, 8 encodes both NEP and NS1 (PB1: polymerase-basic protein 1; PB2: polymerase-basic protein 2; PA: polymerase-acidic protein; NP: viral nucleoprotein; M1: matrix protein; M2: ion channel or membrane protein; HA: hemagglutinin; NA: neuraminidase). Figure from Krammer et al., 2018*[4]*.*

HA, NA and M2 function

IAV, when adapted to humans, mainly infect epithelial cells in the upper respiratory tract[20]. These cells express glycosylated proteins with terminal galactose α2,6-linked sialic acid sugar groups which are bound by homotrimer HA[21]. These sugars are also extensively present in the mucus layer that sits on top of the epithelial cells; however, influenza virions still cross the mucus and reach target cells[19]. This ability is attributed to tetrameric receptor destroying enzyme neuraminidase, which degrades sialic acids in the mucus layer[19]. When arrived at the target cell HA binds an unidentified receptor and the virion enters via clathrin dependent endocytosis[20] or macropinocytosis[22]. Then, the created vesicle will become or fuse with an endosome, where ion channel M2 is activated by low pH and acidifies the virion[17]. Through acidification, receptor-binding subunit HA1 partially reveals fusion-peptide containing subunit HA2, allowing the insertion of fusion peptide into the host membrane[18]. HA2 then folds back, bringing the viral and host membrane in close proximity and fusing them into pores, allowing the viral RNA to escape the endosome[23] (figure 3). The multiple conformations of HA are possible through a number of pH sensors, one of which is shown in figure 4[23]. After release of the RNA, new virions are created by the host and assemble at the cell membrane, where NA mediates viral release[21]. There are approved drugs to combat IAV, most notably M2 blocker amantadine, neuraminidase inhibitors such as oseltamivir and zanamivir, and endonuclease (a subdomain of the viral RNA polymerase) inhibitor baloxavir[12], though influenza is quickly gaining resistances against these drugs[1,24–26]. At the time of writing, there is one approved drug for the inhibition of hemagglutinin, arbidol, which is only approved in Russia and China (Leneva et al 2009). Both arbidol and monoclonal antibodies have shown broad-acting inhibition[27,28] which is why this work will focus on discovering inhibitors for HA.

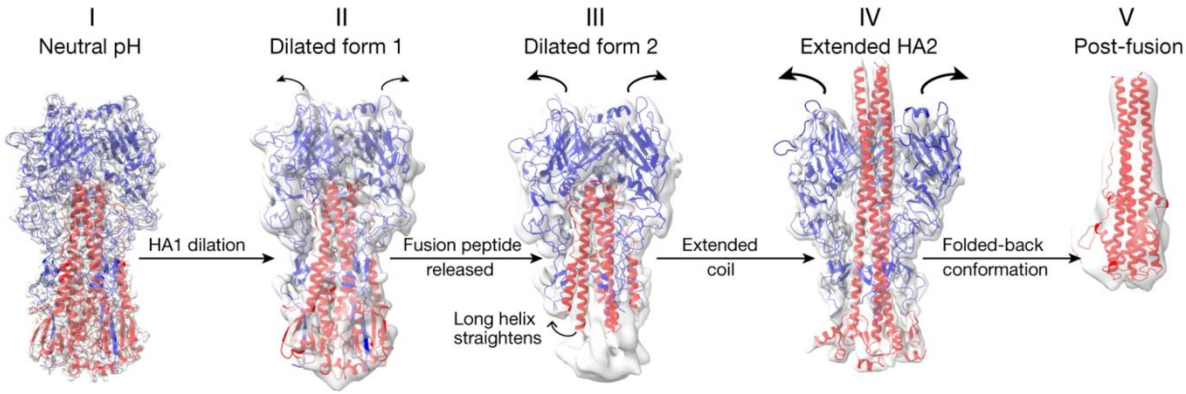


*Figure 3: **HA trimer conformational changes**. HA is anchored to the viral membrane on the bottom, the host membrane is located above the protein. Upon acidification, subunit HA1 (blue) shifts outward, making space for subunit HA2 (red) (I-III). HA2 assembles in long coils and inserts its fusion peptide into the host membrane (IV), before folding back to create pores in the endosome, freeing the viral RNA (V). Figure from Benton et al., 2020[18].*

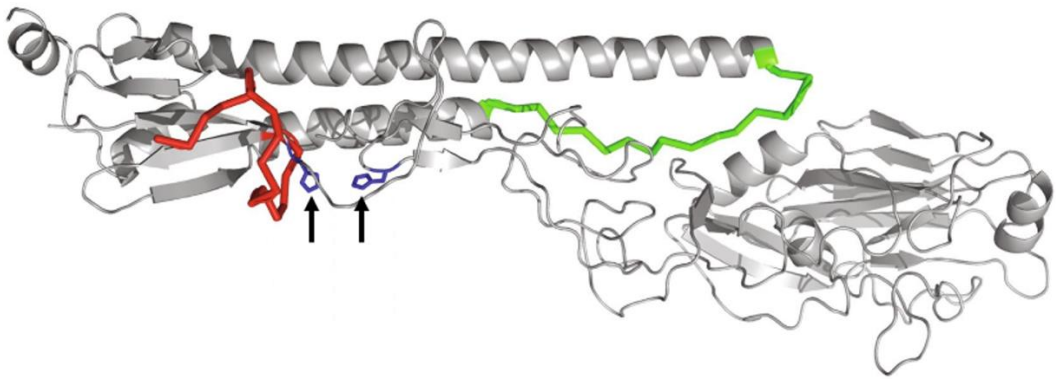


*Figure 4: **Crystallized H5 monomer**. Upon acidification pH sensors such as the residues shown in blue (with black arrows) cause a conformational change in HA, exposing the fusion peptide (red). The fusion peptide can then insert itself into the endosomal membrane. Through more conformational changes the stem loop (green) becomes helical and extends, bringing the endosomal and viral membrane in close proximity, initiating membrane fusion and releasing viral RNA. Figure from Antanasijevic et al., 2020[23].*

Viral escape

IAV viruses have shown antigenic drift in the HA1 subunit (outside of the receptor binding site), with a more conserved HA2 subunit, especially the fusion peptide[29]. Strain-specific and broad-acting neutralizing antibodies are usually overcome by a single amino acid mutation when targeting the HA1 subunit, while broad-acting antibodies targeting the HA2 subunit often exhibit a small loss of activity after a point mutation[30]. This suggests it will be more useful to find inhibitors for the HA2 subunit than for HA1 or full protein.

Random nonstandard peptide integrated discovery (RaPID)

RaPID is a powerful system combining Flexible *In vitro* Translation (FIT) and mRNA display to screen over 10^12 random peptides at once[31]. The FIT system uses catalytic RNAs called flexizymes to charge transfer RNAs (tRNA) with non-canonical amino acids, effectively reprogramming translation by changing codon-amino acid pairs. This allows the creation of peptidase-resistant and improved target-binding peptides through incorporation of macrocycles[32]. The macrocycles are created by starting each peptide with a tyrosine containing an N-terminal chloroacetyl moiety (ClAc-Tyr) which will react to form a thioether bond with the sulfur of a downstream cysteine spontaneously[33]. Using the FIT-system, ClAc-Tyr can be charged on an initiator tRNA with anticodon CAU ($tRNA^{fMet}_{CAU}$), allowing the start of translation with ClAc-Tyr on the regular AUG start codon[33]. Because the NH_2 group of ClAc-Tyr is replaced by ClAc it cannot be incorporated in the random sequence of the peptides. Instead, AUG codons in the random sequence truncate translation. Two libraries are constructed, one with levorotatory (L) and one with dextrorotatory (D) tyrosine stereochemical isomer (only the first amino acid is D, all others are natural L-amino acids), further increasing library diversity. The RNA coding for the library is obtained through transcription of a DNA template, containing from 5' to 3' a T7 polymerase promotor, transcription enhancer, AU-rich translation enhancer epsilon and Shine-Dalgarno sequence (ribosomal binding site), then a start codon, random amino acid coding sequence ($(NNK)_{15}$, table 1), cysteine-coding sequence, GSGSGS-spacer coding sequence and a stop codon (figure 5).

Table 1: ***NNK codon and reprogramming****. The codons consist of two times any (N) and then once U or G (K) nucleotides, made by mixing building blocks in solid phase synthesis of oligonucleotides. The table shows codons with their respective amino acids in NNK libraries. Methionine (AUG) has been omitted from the mixture so that it can be replaced with ClAc-Tyr charged on initiator tRNA$^{fMet}_{CAU}$. This ensures every peptide starts with ClAc-Tyr. All other natural amino acids are present. During selection, either ClAc-L-Tyr or ClAc-D-Tyr will be added to the translation mix, resulting in two different libraries.*

First base	Second base: U	Second base: C	Second base: A	Second base: G	Third base
U	Phe	Ser	Tyr	Cys	U
					C
					A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
					C
					A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
					C
					A
	ClAc-Tyr	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
					C
					A
	Val	Ala	Glu	Gly	G

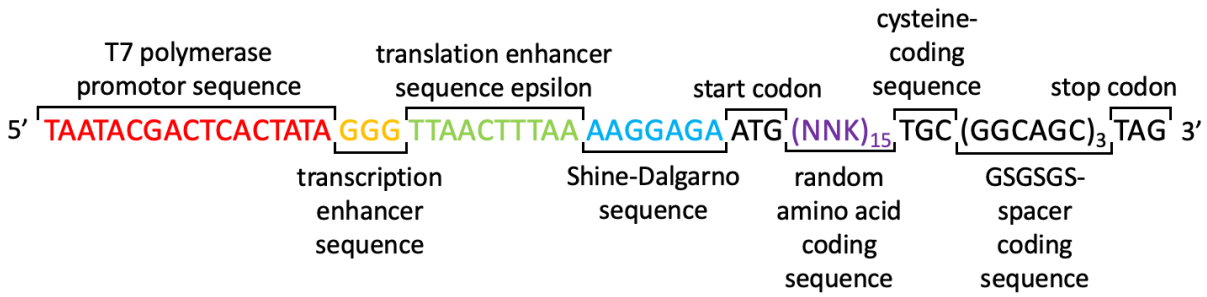


Figure 5: ***DNA template for RNA library****. At the 3' end, primer annealing sequence "GACGGGGGGCGGAAA" is attached.*

mRNA display is a high throughput screening technique featuring iterative rounds of selection to a target. First, an RNA library is made with high diversity in the random sequence, while retaining conserved 3' and 5' ends for amplification and sequencing. The library is then coupled to puromycin and translated using the commercial PURE (Protein synthesis Using Recombinant Elements) *in vitro* translation system in combination with FIT reprogramming. At the end of translation, release factors disengage the ribosome from the RNA, resulting in separate RNA and protein. In the PURE system however, release factors can be omitted. Instead, puromycin attached to mRNA, resembling aminoacyl tRNA, enters the ribosome and is attached to the amino acid chain. This terminates translation and attaches RNA to

peptide[34]. The RNA is then reverse transcribed to DNA and the tagged peptide library exposed to H1 mutant protein immobilized by magnetic beads. A fraction of the peptides is used for quantification by quantitative polymerase chain reaction (qPCR). The cDNA is then multiplied by regular PCR and finally transcribed to RNA to start the next cycle (figure 6). When qPCR shows high recovery of DNA, hits will be sequenced and synthesized on fluorenylmethyloxycarbonyl (Fmoc)-protected Solid Phase Peptide Synthesis (SPPS), purified and then tested for function.

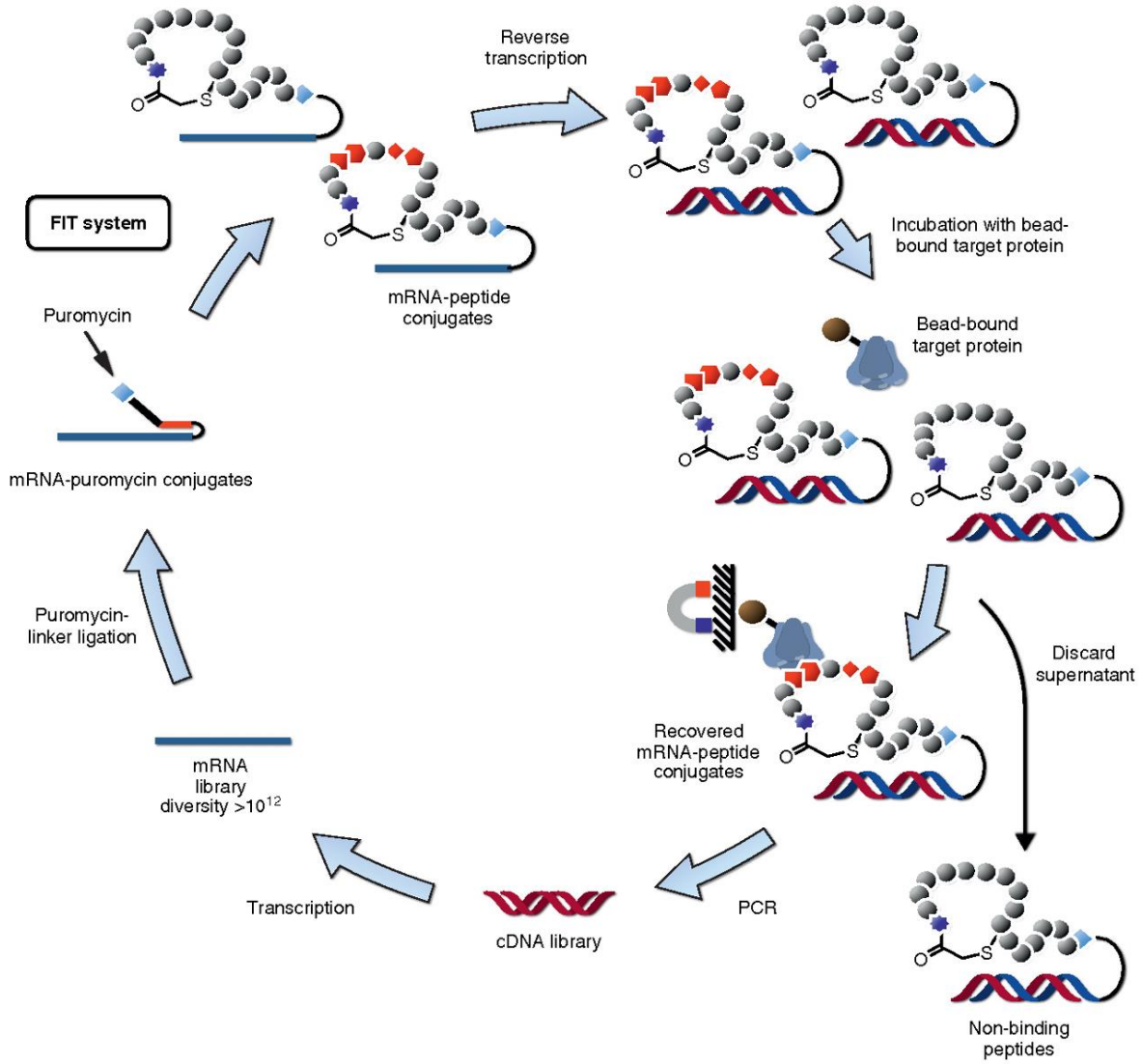


*Figure 6: **Selection cycle**. The cycle starts with the mRNA library (bottom-left). It is linked to puromycin and afterwards translated using the FIT system, attaching RNA and peptide. At this point, the peptide spontaneously cyclizes between the acetylated tyrosine and sulfide of the cysteine. Then, reverse transcription is used to create cDNA and peptides are incubated with target bound on beads, resulting in binding peptides on beads and non-binding peptides in the supernatant, which is then discarded. DNA is multiplied by PCR and transcribed, so a new round can be started. Figure from Hipolito & Suga, 2012[35].*

Previous work

Previously, our group has used the RaPID system to identify small, 17-residue peptides that inhibit H1 of the 2009 H1N1 IAV A/California/04/2009, A/Netherlands/602/2009 (Neth09) and A/PR8/8/34/Mount Sinai (manuscript under review). One of these peptides, named S5, was particularly effective, binding the well-conserved fusion peptide directly and thus preventing viral fusion (sequence: **yVLFRWDHGTLATHWVC** (D-Tyr)), also when tested against the HA of a H5N1 virus (A/duck/Hunan/795/2002). However, the peptide lacks activity against the HA of a H3N1 virus (A/Bilthoven/1761/76). Upon discovery of this inhibitor, our collaborators from the Utrecht University virology department drove the virus to resistance against this peptide by passaging the virus with a low concentration of inhibitory peptide, finding an escape variant with a mutation from isoleucine to phenylalanine in residue 375 (I375F).

Current project outline

In this project, the RaPID will be used to find more hemagglutinin inhibitors. The main goal of this project is twofold: to treat resistant virus or prevent resistance by co-administration of S5 and a new peptide. Additionally, a long-term goal is to find inhibitors for every escape variant, so there can be a definitive treatment. The mutant was expressed as homotrimeric stem only, which is subunit HA2, because the protein was not stable as full-length monomer and to direct the selection towards fusion peptide inhibition. If the recovery in the stem only H1 mutant I375F selection increases, a secondary goal is set for the identification of a broad-acting inhibitor for H1, H3, H5 and H9. Of these proteins, H1, H5 and H9 are in the same phylogenetic group, while H3 is more distantly related[36]. H3 and H9 were selected for their pandemic potential[37,38], while H5 was also selected for its severe pathogenicity[38].

Methods

Reagent sourcing

All chemicals and reagents were purchased and used without further purification from Avantor, Biosolve, Manchester organics, Merck, New England biolabs (NEB) or Thermo Fisher Scientific unless mentioned otherwise. All primers were purchased from Integrated DNA Technologies. All hemagglutinin proteins were created and purified by the Utrecht University Virology department.

Thermal shift

qPCR

Duplicate mixtures were made with 0.1 µg/µL stem only H1 (H1s) wild type (WT, from Neth09) or H1s mutant I375F (H1s I375F), 2x SYPRO orange protein gel stain in 20 µL 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)/150 mM sodium chloride. This was incubated for 5 seconds for each 0.5 °C difference in a melt curve from 25 to 75 °C on a Bio-Rad CFX96 Touch Real-Time PCR Detection System, using the built-in FRET excitation and emission settings. The mean of both duplicates was plotted as the first derivative of relative fluorescence units over temperature.

NanoDSF

10 µL of 10 µM H1s WT or H1s I375F in 100 mM HEPES/150 mM sodium chloride was injected into a nanoDSF machine (2Bind). A melt curve was generated from 35 to 95 °C with steps of 0.2 °C over 5 seconds. Fluorescence was measured at 330 and 350 nm. The first derivative of the ratio of 350nm/330 nm was plotted against temperature.

H1s I375F bead saturation assay

1, 2 or 5 µL 5% MagStrep “type 3” XT beads (IBA Lifesciences) (hereafter: beads) were washed 3 times with double the original volume of phosphate buffered saline (PBS) containing 0.1% tween-20 (PBS-T). Then, 0.38 µg H1s I375F in PBS-T (0.055 g/L) was added and incubated at 4 °C with inversion for 45 minutes. Supernatants were reserved and beads washed 3 times with double volume PBS-T. An additional sample of 0.4 µg H1s I375F protein and all bead and supernatant samples were diluted to a volume of 10 µL, then further diluted with 2 µL 6x sodium dodecyl sulfate (SDS) loading buffer (0.375M tris pH 6.8, 12% w/v SDS, 48% glycerol, 9% v/v 2-mercaptoethanol, 0.03% w/v bromophenol blue in ultrapure H_2O) and incubated at 95 °C for 5 minutes. Each entire sample was loaded onto an SDS-polyacrylamide gel electrophoresis (PAGE) gel (5 mL resolving gel: 0.375 M tris pH 8.8, 0.1% w/v SDS, 10% 19:1 acrylamide/bis (Bio-Rad) in ultrapure H_2O; 1 mL stacking gel: 0.125 M tris pH 6.8, 0.1% SDS, 5% 19:1 acrylamide/bis in ultrapure H_2O). Both gels were set with 0.1% w/v ammonium persulfate and 0.2% v/v tetramethylethylenediamine (TEMED, Roth)) next to a protein marker (unstained protein MW marker, Thermo Fisher Scientific) and ran at 120 volts for 90 minutes in SDS running buffer (25 mM tris, 192 mM glycine, 0.1% w/v SDS in ultrapure H_2O). Silver stain was performed as protocol A in Chevallet *et al*., 2006[39]. ImageJ software was used for densitometry analysis.

Making $tRNA^{fMet}_{CAU}$ and eFX

$tRNA^{fMet}_{CAU}$ and enhanced flexizyme (eFX) were produced as described in Suga *et al*., 2011[40]. DNA template was assembled by PCR, then transcribed into RNA with *in vitro* transcription. Product was purified by urea-PAGE, dissolved in ultrapure H_2O and stored at -20 °C.

Charging ClAc-L/D-Tyr on initiator tRNA

5 µL reactions were made by mixing 83 mM HEPES pH 7.2, 42 µM eFX and 42 µM $tRNA^{fMet}_{CAU}$ in 3 µL, which was incubated at 95 °C for 2 minutes, then allowed to cool at room temperature for 5 minutes. 1 µL 3 M magnesium chloride was added, and the mixture was again incubated for 5 minutes at room temperature, then briefly cooled on ice before adding 1 µL 25 mM ClAc-L-Tyr or ClAc-D-Tyr in dimethyl sulfoxide (DMSO) (produced in-house). The reactions were incubated on ice for 2 hours. Reactions were then quenched and precipitated by addition of 20 µL 0.3 M sodium acetate pH 5.2 and 30 µL ethanol, centrifuged at 15000 relative centrifugal force (rcf) for 15 minutes. Supernatant was removed and pellets were vortexed and washed with 50 µL 0.1 M sodium acetate pH 5.2 in 70% ethanol and centrifuged as above for 10 minutes, twice. Pellets were then washed without vortexing with 30 µL 70% ethanol and centrifuged for 5 minutes, then allowed to air-dry and stored as dry pellets at -20 °C.

Production and testing of reverse transcriptase

Transformation of plasmids

20 µL BL21 DE3 competent E. coli cells (NEB) were thawed on ice, 100 ng reverse transcriptase (MMLV-RT) plasmid 1 (Superscript II, from Crick institute's Walport lab) and 2-mercaptoethanol to a final concentration of 48 mM were added. The mixture was incubated on ice for 30 minutes, then a 42 °C heat shock for 30 seconds, then on ice for 2 minutes. 80 µL S.O.C. medium (NEB) was added before incubating the mixture at 37 °C, 300 rpm shaking for 1 hour. Cells were plated on sterile Luria-Bertani (LB granules dissolved in dH_2O following manufacturer's instructions, Thermo Fisher Scientific) agar plates containing 50 µg/mL kanamycin, grown at 37 °C overnight. The next day, one colony was picked and grown in LB medium containing 50 µg/mL kanamycin at 37 °C with 250 rpm shaking overnight. The cells were diluted 100 times in 25 mL YT medium (10 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride in dH2O, pH 7.2) and incubated at 37 °C with 250 rpm shaking until an optical density of 0.5 was reached at 600 nm (OD_{600}). The culture was cooled on ice before pelleting at 4000 rcf, 4 °C for 10 minutes. The pellet was gently resuspended in 2 mL 30 mM ice-cold calcium chloride and incubated on ice for 30 minutes. Cells were again pelleted and gently resuspended in 600 µL ice-cold 30 mM calcium chloride. To 40 µL cells, 200 ng of the second plasmid was added (pG-Tf2, containing chaperone proteins groES, groEL and tig, Takara bio). 2-mercaptoethanol was added as above and cells were incubated and grown as before. Cells were plated on LB agar plates with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol (Apollo Scientific) and incubated overnight at 37 °C.

Test expression

One colony was picked and grown in 2 mL LB medium at 37 °C, 300 rpm shaking overnight. 100 µL was taken and added to 10 mL LB medium with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol and grown at 37 °C, 250 rpm shaking until an OD_{600} of 0.2 was reached. Then, the chaperone plasmid was induced with 5 ng/mL tetracycline and cells were grown further until an OD_{600} of 0.6. The MMLV-RT plasmid was induced with 0.1 mM IPTG (isopropyl

β-D-1-thiogalactopyranoside) and cells were grown overnight as before. A sample was diluted with 6x SDS loading buffer to 1x loading buffer and incubated at 95 °C for 5 minutes. 10 µL sample was loaded onto a 10% SDS-PAGE gel as before next to a protein marker (unstained protein MW marker). The gel was run at 100 volts for 45 minutes, then washed with ultrapure H_2O, stained in Coomassie blue stain (Quick Coomassie stain, Protein Ark) for 50 minutes and de-stained in ultrapure H_2O for 80 minutes.

Purification

During the entire purification, samples, buffers and columns were kept on ice or in a climate-controlled room at 4 °C whenever possible and centrifuges spun at 4 °C unless mentioned otherwise. MMLV-RT expressing cells were grown in LB medium with 50 µg/mL kanamycin and 20 µg/mL chloramphenicol at 37 °C with 250 rpm shaking overnight. Cells were added 1:25 to 500 mL autoinducing medium (25 mM sodium phosphate, 25 mM potassium phosphate, 20 g/L tryptone, 5 g/L yeast extract, 5 g/L sodium chloride, 0.6% v/v glycerol, 0.5 g/L glucose and 2 g/L lactose in dH2O, pH 7.4) with both antibiotics and grown at 37 °C, 250 rpm shaking until OD_{600} of 0.2 was reached. Tetracycline was added to a concentration of 5 ng/mL and cells were grown at 18 °C, 250 rpm shaking overnight. Cells were pelleted by centrifugation at 4000 rcf for 10 minutes, resuspended in lysis buffer (20 mM tris, 5 mM imidazole and 0.5% triton X-100 in dH2O, pH 8) and lysed by micro fluidization (twice at 15,000 psi). The mixture was pelleted at 4000 rcf for 30 minutes and supernatant was loaded on a His GraviTrap TALON 1 mL column (Cytiva) equilibrated with lysis buffer. The column was washed with lysis buffer containing 40 mM imidazole, then eluted with lysis buffer containing 500 mM imidazole. Flowthrough was collected in a high and low protein concentration fraction (measured by A_{280} on nanodrop and extinction coefficient calculated from the sequence) and filtered separately on a 9 mL sephadex G-25 medium (particle size: 50-150 µ, Pharmacia) spin column with 2x MMLV-RT storage buffer (100 mM tris, 150 mM sodium chloride, 0.2 mM ethylenediamine tetraacetate acid (EDTA), 2 mM dithiothreitol (DTT) and 0.2% triton X-100 in ultrapure H_2O) by 4000 rcf centrifugation for 1 minute. Samples of the pellet, supernatant, washes and elute of the TALON column and filtrates of the sephadex column were loaded with protein marker (unstained protein MW marker) onto a 10% agarose SDS-PAGE gel as above and ran at 200 volts for 35 minutes in SDS running buffer. Silver stain was performed as above. MMLV-RT in 2x storage buffer was diluted 1:1 with 100% glycerol and stored at -20 °C.

RNase activity test

1 µL MMLV-RT high or low concentration (filtrate from the sephadex column) or 2x storage buffer was incubated with 2 µL 20 µM NNK15 RNA library at 42 °C overnight or for 1 hour. A 2 µL NNK15 RNA library control sample was also made, but not incubated. All samples were diluted 1:1 with 2x RNA loading buffer (to 4 M urea, 1 mM sodium EDTA, 1 mM tris pH 7.5 and 0.2% w/v bromo phenol blue (BPB)) and incubated at 95 °C for 5 minutes before loading onto a urea-PAGE gel (6 mL gel, 6 M urea, 8% acrylamide/bis 19:1 in ultrapure H_2O, set with 0.1% w/v ammonium persulfate and 0.2% TEMED) next to 2.5 µL DNA marker (FastRuler low range DNA marker, Thermo Fisher Scientific). High and low concentration fractions were combined to 12.5 mg/mL in a total volume of 28 mL before testing activity, of which 1 mL was stored at -20 °C for testing and further use, 27 mL was flash frozen with liquid nitrogen and stored at -80 °C.

Reverse transcriptase activity test (performed in-house)

Reagents were mixed to 10 µL 20 µM RNA template, 0.25 mM dNTPs, 5 µM fluorescein primer, 25 mM tris pH 8.3, 15 mM magnesium chloride, 10 mM potassium hydroxide and 0.25 µL buffer or 0.3 mg/mL homemade MMLV-RT. Mixtures were incubated at 42 °C for 1 hour. After incubation, 0.2 µL RNaseH (NEB) was added to 1 mixture and incubated at 37 °C for 20 minutes. 2 µL of sample was diluted 1:1 with 2x RNA loading buffer, ran on urea-PAGE gel as before and bands were visualized at 480 nm excitation and 525 nm emission.

RaPID selection against H1s I375F

Puromycin ligation

For both libraries, reagents were mixed to 20 µL 1x T4 RNA ligase buffer (NEB), 1 mM ATP, 1 mM DTT, 5% w/v poly-ethylene glycol (PEG)8000, 1.5 equivalents of puromycin linker, 30 units of T4 RNA ligase (NEB) and 1 µM NNK15 RNA library (produced in-house) (sequence: GGGUUAACUUUAAGAAGGAGAUAUACAUAUG$(NNK)_{15}$UGCGGCAGCGGCAGCGGCAGCUAGGA CGGGGGGCGGAAA) in ultrapure H_2O. The reaction was incubated at room temperature for 25 minutes before precipitation by addition of 20 µL 0.6 M sodium chloride with 10 mM EDTA and 80 µL ethanol, centrifuged at 15000 rcf for 15 minutes. Pellets were then washed with 70% ethanol by 3-minute centrifugation, allowed to air-dry and resuspended in 2 µL ultrapure H_2O.

Translation and reverse transcription

For each library, 5 µL translation reactions were prepared on ice: 30% v/v solution B without release factors (PURExpress Δ RF123 Kit, NEB), 14.2% v/v homemade solution ΔA (50 mM HEPES pH 7.6, 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 100 mM potassium acetate, 2 mM spermidine, 7 mM magnesium acetate (solution A heated to 37 °C before addition), 1.5 mg/mL tRNA from E. coli MRE 600 (Roche) and 15 mM DTT, final concentration in translation reaction), 25 µM ClAc-L-Tyr or ClAc-D-Tyr (one pellet, section "Charging ClAc-L/D-Tyr on initiator tRNA"), 2 µM puromycin-ligated RNA and 0.5 mM 19 amino acids mix (19 natural amino acids, omitting methionine). Reactions were incubated at 37 °C for 25 minutes, then at room temperature for 10 minutes before addition of 1 µL 100 mM EDTA and a second 37 °C incubation for 25 minutes. Then, a reverse transcription mix (prepared on ice) was added to 10 µL total volume with a final concentration of 0.25 mM dNTPs, 2 µM CGS3an13.R39 primer (supporting information, table 3), 25 mM tris pH 8.3, 15 mM magnesium chloride, 10 mM potassium hydroxide and 5 units of ProtoScript II RT (NEB) or 0.3 mg/mL homemade MMLV-RT (section "Production and testing of reverse transcriptase"). Libraries were incubated at 42 °C for 1 hour, before addition of 1.4 µL 1% acetylated bovine serum albumin (BSA) and 3.6 µL 5x PBS-T. A sample of 0.2 µL was then diluted to 500 µL in ultrapure H_2O to measure input.

HA pulldown

Every step of the bead selection was performed on ice or in a climate-controlled room at 4 °C. All cold incubations were done under gentle inversion. Beads were washed 3 times with double volume of PBS-T. On the last wash, aliquots were made of 2.52 µL (positive selection) or 2 µL (negative selection) and supernatants were removed. Positive selection beads were incubated with 3.78 µL 0.1 g/L H1s I375F for 20 minutes and washed 3x with double volume PBS-T again. Libraries were added to negative selection beads in series and incubated for 5 to 10 minutes, then transferred to positive selection beads and incubated for 30 minutes. 3 to 9

rounds of negative selection were included for each round, as indicated in results. The final negative selection and positive selection beads were washed 3x with 20 µL PBS-T, transferring the beads to a clean tube each wash. After the last wash, beads were resuspended in 50 µL ultrapure H_2O and incubated at 95 °C for 5 minutes, then supernatant was quickly transferred to a clean tube.

Analysis with qPCR

qPCR mix was prepared on ice, with a final volume 20 µL per sample. Either Taq DNA polymerase (Thermo Fisher Scientific) or KOD polymerase (produced in-house) were used, with different reaction mixes. Reaction mixtures using Taq polymerase contained 50 mM potassium chloride, 10 mM tris pH 9, 0.1% v/v triton X-100, 0.25 mM dNTPs, 2.5 mM magnesium chloride, 0.25 µM T7g10M.F48 (supporting information, table 3) and CGS3an13.R39 primers, 6 units Taq DNA polymerase (Thermo Fisher Scientific), 1x SYBR green I nucleic acid gel stain and 1 µL of sample (ultrapure H_2O, DNA standard, input sample, positive or negative selection supernatant). Reaction mixtures using KOD polymerase contained 120 mM tris pH 8, 10 mM potassium chloride, 6 mM ammonium sulfate, 0.1% v/v triton X-100, 0.001% BSA, 0.25 mM dNTPs, 1 mM magnesium chloride, 0.25 µM T7g10M.F48 and CGS3an13.R39 primers, 6 µg/mL KOD polymerase, 1x SYBR green I nucleic acid gel stain and 1 µL sample (as in Taq mix). DNA standards were mixed to concentrations ranging from 2 * 10^9 to 2 * 10^5 in 10x dilution steps before addition to the qPCR mixes (produced in-house). Samples were loaded onto an AriaMX real-time qPCR machine (Agilent) and incubated at 95 °C for 3 minutes, then 35 cycles of 95 °C for 10 seconds, 61 °C for 10 seconds and 72 °C for 30 seconds. Fluorescence was measured at 516 nm emission and 462 nm excitation following each 61 °C step.

Amplification with PCR

The remaining positive selection supernatants were prepared for PCR using reaction mixes as before, only omitting SYBR green I. 100 µL PCR reactions were amplified for quantification cycle (Cq, from qPCR) + 7 cycles of 95 °C, 61 °C and 72 °C, all steps lasting 40 seconds. 1 µL was mixed with 1 µL 2x DNA loading buffer (1 mM EDTA, 40 mM tris, 20 mM acetic acid, 15% w/v glycerol and 0.00025% w/v BPB final concentrations) and used to analyze the amplification product on a 3% agarose gel (10 mM sodium hydroxide pH 8 with boric acid, 1x SYBR safe DNA gel stain; Thermo Fisher Scientific) next to a DNA marker with DNA the size of 67, 98 and 123 base pairs (produced in-house, 123 base pairs corresponds to library) at 200 volts for 10 minutes in the same buffer. PCR product was then precipitated by addition of 10 µL 3 M sodium chloride and 220 µL ethanol and centrifuged at 15000 rcf for 15 minutes, air-dried and resuspended in 10 µL 50 mM potassium chloride. If the band was not intense enough, 2-5 more rounds of PCR were performed. If a large amount of truncated product was present, an RNA purification urea-PAGE gel was used to separate truncated product from desired product as previously described[40].

In vitro transcription

20 µL transcription reactions were prepared containing 40 mM tris pH 8, 1 mM spermidine, 0.01% v/v triton X-100, 10 mM DTT, 20 mM magnesium chloride, 3.75 mM unbuffered NTPs, 29 mM potassium hydroxide, 4 units of RNase inhibitor murine (NEB), 50 µg/mL T7 polymerase (produced in-house) and 4 µL precipitated DNA product. Reactions were incubated at 37 °C for 3 hours or overnight and then precipitated by addition of 20 µL 0.6 M

sodium chloride/50 mM EDTA and 32 µL isopropanol, centrifuged at 15000 rcf for 5 minutes. Pellets were then washed with 70% ethanol and then centrifuged for 3 minutes, air-dried and resuspended in ultrapure H_2O to 10 µM for the next round which starts at puromycin ligation.

RaPID selection against other constructs

Bead saturation with 50 µM strep-tag peptide

All stem only constructs used contain a triple strep-tag, while all full-length proteins contain a double strep-tag. H1s WT and full-length H9 were used to test displacement of protein on beads in the presence of strep-tag peptide. 2 µL beads were washed 3 times with double volume PBS-T. Then, 4 µL H1s WT or full-length H9 at 0.1 g/L was added and incubated at 4 °C with inversion for 30 minutes. Supernatants were reserved and beads were washed 3 times with double volume PBS-T. 15 µL 50 mM strep-tag peptide in PBS-T (sequence: **WSHPQFEK**, produced in-house) was added and beads were incubated as above. Supernatants were reserved again, and beads were washed as above. All samples were diluted to a volume of 15 µL, then further diluted with 3 µL 6x SDS loading buffer and incubated at 95 °C for 5 minutes. Each whole sample was loaded onto an 8% SDS-PAGE gel next to a protein marker (unstained protein MW marker) and run at 120 volts for 90 minutes in SDS running buffer as in section "H1s I375F bead saturation assay". Silver staining was performed as above.

Selection

The DNA output of H1s I375F round 3 was used with 50 µM strep-tag peptide in the positive selection on H1s I375F and other HAs, including H1s WT, full-length H3, stem only H5 and full-length H9. Also, the library of a previous selection against the H1s WT was used against H1s I375F. These selections were performed as in section "RaPID selection against H1s I375F", except 3.6 µL 0.1 g/L protein was immobilized on 2.52 µL beads and strep-tag peptide was added during the positive selection incubation.

Sequencing

The amplified DNA product of every selection round was prepared for sequencing by two rounds of extension PCR, using the same general mixture: 120 mM tris pH 8, 10 mM potassium chloride, 6 mM ammonium sulfate, 0.1% v/v triton X-100, 0.001% BSA, 0.25 mM dNTPs, 1 mM magnesium chloride and 6 µg/mL KOD polymerase. In the first extension PCR, 0.5 µM CGS3an13Rd2.R49 and Rd1T7g10M.F70 primers (supporting information, table 3) and 0.5 µL DNA sample were mixed with PCR mix to 50 µL, then PCR was performed with a 1-minute 95 °C incubation, then 5 cycles of 95 °C, 63 °C and 72 °C, all 1 minute, and then 72 °C for 2 minutes. 1 µL of this product was used for the second PCR round, using the same mix but a combination of different primers; S505 through S507 and N716 through N729 (supporting information, table 3 and 4) in a 100 µL reaction. 9 PCR cycles were performed as above, then product amplification was verified on a 3% agarose gel as in section "multiplication with PCR". Samples were pooled, then purified via NucleoSpin clean-up kit (Macherey-Nagel) and sequenced on ISeq (Illumina) following respective manufacturer's instructions.

Selection of sequences

The sequences were extracted using a Python script searching for exact matches to the T7 promotor and puromycin linker sequences and sorted based on amino acid sequence

prevalence. The top 100 hits of all rounds performed after H1s I375F round 2 were aligned in CLC sequence viewer 7 software and alignments were visualized as a tree created by neighbor joining with 100 replicates. A sequence logo was created for every cluster of interest (containing at least 4 sequences) on https://weblogo.berkeley.edu/logo.cgi (default settings), from which important binding motifs were identified. The top hits of each cluster with a clear binding motif were combined and aligned again, then trimmed visually based on sequence similarity between rounds and targets. Sequences were discarded if they were similar to strep-tag motif **WSHPQFEK**.

Solid phase peptide synthesis

Chosen peptide hits were synthesized using SPPS on a SYRO II (Biotage). All peptides were synthesized with a GS linker instead of GSGSGS. First, TentaGel S RAM resin (25 µmol scale, Rapp polymere) was incubated with 1.2 mL 25% piperidine in dimethylformamide (DMF) for 10 minutes with shaking to remove the Fmoc protecting group and then washed 6 times with 1 mL DMF. The following steps were repeated until the full length of the peptide was achieved: 750 µL coupling of 0.19 M amino acid with 0.19 M HATU (1-[bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxide hexafluoro-phosphate) in DMF for 40 minutes, then a second 750 µL coupling step with 0.19 M amino acid, 0.19 M PyBOP (benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate) in DMF, then a 1 mL capping step with 10% acetic anhydride, 0.38 M DIPEA (*N*,*N*-diisopropylethylamine) and 0.19 M PyBOP in DMF for 30 minutes and finally Fmoc deprotection as above. Then, resins were incubated with 0.2 M chloroacetyl-*N*-hydroxysuccinimide in DMF for 30 minutes, twice. Resins were washed with DMF 3 times.

Cleavage, deprotection and cyclization

Resin was washed with dichloromethane 3 times, dried under vacuum and then incubated with 1 mL 90% trifluoroacetic acid (TFA), 5% ultrapure H_2O 2.5% triisopropylsilane and 2.5% EDDET (2,2'-(ethylenedioxy)diethanethiol) under vortex at room temperature for 1 hour (or 3 hours if the peptide contains arginine). Then, the reaction mixture was filtered into 20 mL cold diethyl ether and centrifuged at 4500 rcf for 5 minutes. Supernatant was discarded and pellets were washed 3 times by vortexing with 10 mL cold diethyl ether then centrifuging as above, then allowed to air-dry. Pellets were dissolved in 2 mL DMSO and incubated under basic conditions by addition of ~10 drops of DIPEA with mild shaking at room temperature overnight, then quenched with TFA.

Purification by HPLC

Peptides in DMSO were diluted 1:1 with 0.1% TFA in ultrapure H_2O and filtered through a 0.45 µm filter before injection on a preparative high-performance liquid chromatography (HPLC) system (Agilent) with UV/VIS detection at 215 and 280 nm. Contents were separated by C18 column (Gemini C18, 250 x 21.2 mm, 10 µm particle size, Phenomenex), with a 12.5 milliliters per minute flow of a 5 to 95% acetonitrile gradient in ultrapure H_2O with 0.1% TFA over 30 or 60 minutes. Collected fractions were lyophilized and dissolved in 100 µL 10 mM hydrogen chloride and lyophilized again, twice. Peptides were dissolved in 20 or 40 µL DMSO and purity and identity verified on analytical HPLC-mass spectrometry with ESI-MS detection (LC-MS) (supporting information).

Peptide ELISA

First, 96-well plates (flat bottom Nunc-immuno maxisorp) were coated with 50 µL 50, 10, 2, 0.4 or 0 µg/mL (concentration range 1) or 12.5, 2.5, 0.5, 0.1 or 0 µg/mL (concentration range 2) of stem only H5 in 0.05 M (bi)carbonate, pH 9.6 at 37 °C for 1 hour, then washed 3x with 200 µL PBS-T (0.05% tween-20 unless mentioned otherwise). Control experiments were performed without coating and the subsequent washes. Plates were blocked with 100 µL blocking agent (0.1 or 0.2 M spermidine, 10 or 20% BSA or 10% soy protein) and incubated at 37 °C for 1 or 2 hours. Plates were washed as above with 1% BSA in PBS-T. 50 µL 2, 0.2, 0.02 or 0 µM biotinylated L2 (sequence: YWHKNKYVLTYSCLFAAGGKG, where the penultimate K has a PEG4-biotin moiety attached, produced in-house) in PBS-T was added, and plates were incubated at room temperature for 1 hour, before washing 3, 6 or 10x with 1% BSA in PBS-T, with 0.05 or 0.2% tween-20. 50 µL of 1/5000 streptavidin-conjugated horseradish peroxidase (Cytiva) in 1% BSA in PBS-T was added to each well and incubated at room temperature for 20 minutes. Wells were washed 3x with 1% BSA in PBS-T, then 50 µL of substrate mix (0.005% w/v 3,3',5,5'-tetramethylbenzidine in DMSO and 0.003% v/v hydrogen peroxide in 0.2 M citric acid, pH 4, vortexed) was added and color change was measured at 605 nm. All experiments were performed in duplicate with controls omitting protein, peptide or both.

Peptide inhibitory assay (performed by virology)

Henrietta Lacks cells transfected with the negative-sense RNA Gaussia luciferase gene were incubated with 100 µL 2 µM peptide/virus mix (H1 WT full-length or H1 I375F full-length, with a viral concentration resulting in 80-90% infection without inhibitors) at 37 °C with 5% CO_2 for 16 hours. 10 µL of cells were lysed with 40 µL passive lysis buffer on a 96-well plate and incubated with 40 µL coelenterazine, then measured by luminometer (GloMax, Promega).

Results and discussion

Thermal shift (qPCR)

Previously, the full-length H1 I375F mutant could not be expressed as stable construct, so homotrimeric stem only H1 (H1s) I375F was made instead. To determine stability of the protein and biological relevance of a RaPID selection, this project starts with thermal shift. In thermal shift, a change in relative fluorescence units (RFU) signifies protein unfolding by increased dye-protein binding. By plotting graphs as derivative of RFU over temperature, a peak will be visible at the temperature where 50% of the protein is unfolded, called the inflection point. H1s WT and H1s I375F resulted in peaks at 35.5 and 65.5 °C, with a third peak at 53 °C for H1s WT and one at 51.5 °C for H1s I375F (figure 7). The first peak is dissociation of the trimer domain, while the second confirms the stability of the protein. The third peak may be the unfolding of a secondary element elsewhere in the protein. Signal in thermal shift is directly correlated to the amount of protein present. H1s I375F provides both a larger signal in the trimer domain dissociation and a smaller signal in the overall stability of the protein. H1s WT may be present in dimers instead of trimers, resulting in a smaller signal at trimer domain dissociation, but be at a higher concentration than H1s I375F, resulting in a larger signal at the overall stability of the protein. Based on the thermal shift resulting in similar inflection points for both H1s WT and H1s I375F, H1s I375F was considered stable enough to be biologically relevant and thus warranted a RaPID selection.

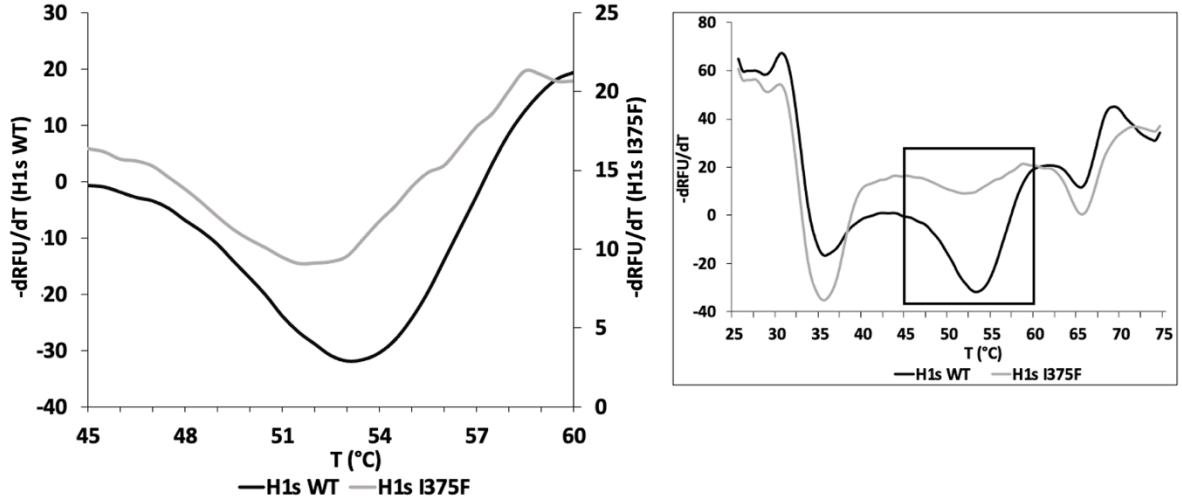


Figure 7: ***H1s WT and H1s I375F thermal shift melt curves****. Duplicates measured by qPCR were averaged. The derivative of relative fluorescence units (RFU) of H1s WT (black) and H1s I375F (gray) is plotted against temperature. The left graph is an excerpt of the full graph on the right, with H1s WT on the left y-axis and H1s I375F on the right y-axis.*

H1s I375F bead saturation assay

For a RaPID selection, in general 200 nM protein is preferred during library pulldown. To determine protein-bead binding, a bead saturation assay was performed where protein was bound to different volumes of beads (figure 8). Densitometry was used to determine protein band intensity of bead-supernatant pairs and values were used to find the conditions where 67% of protein is bound by beads, to ensure no beads had free binding sites in the positive selection. A ratio of 150 nanograms of protein per microliter beads was found.

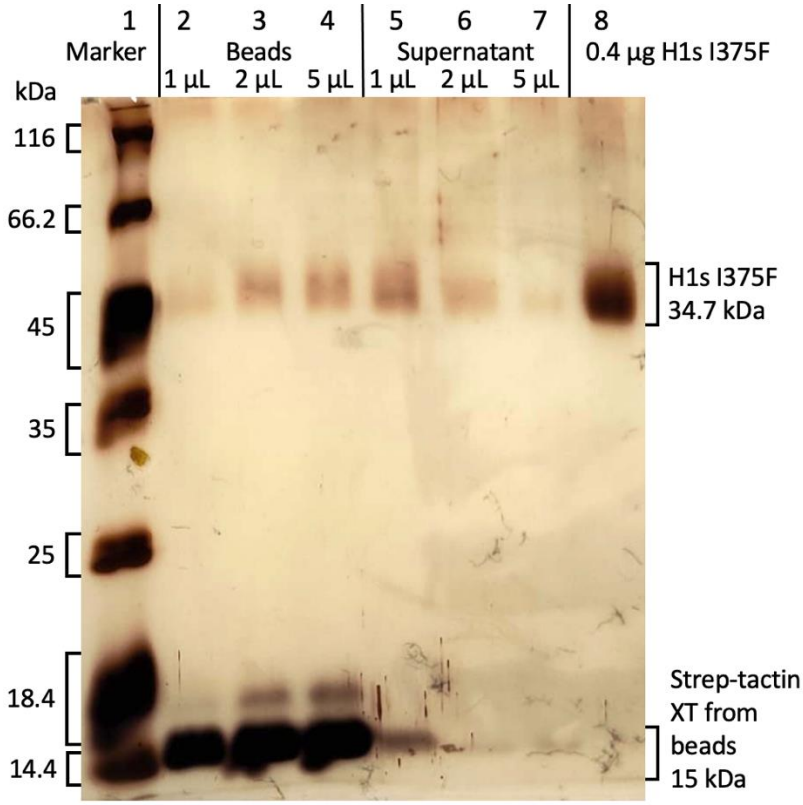


Figure 8: ***SDS-PAGE gel (silver stain) showing H1s I375F binding to beads****. 1: protein marker; 2-4: washed bead fraction of 1, 2 or 5 µL of beads, respectively; 5-7: supernatant fraction of 1, 2 or 5 µL of beads, respectively; 8: 0.4 µg H1s I375F. H1s I375F is 34.7 kDa without glycosylation, the strep-tactin XT from the beads is 15 kDa. Relative band intensity was calculated with densitometry and used to calculate the volume of beads and protein to yield fully occupied beads and 200 nM protein in a 15 µL library selection.*

Reverse transcriptase

Test expression

For correct folding of MMLV-RT (75 kDa), chaperone proteins GroEL, Tig and GroES (60, 48 and 10 kDa, respectively) are required. To determine correct transformation of both plasmids into the bacterial cells, a protein test expression was performed (figure 9). MMLV-RT, GroEL and Tig were found on gel, but GroES was not. As protein marker bands for 18.4 and 14.4 kDa ran off the gel, GroES is only 10 kDa and all three chaperones are found on one plasmid, expression of GroES was not tested further. Rather, a large-scale expression was set up for purification and testing of the MMLV-RT.

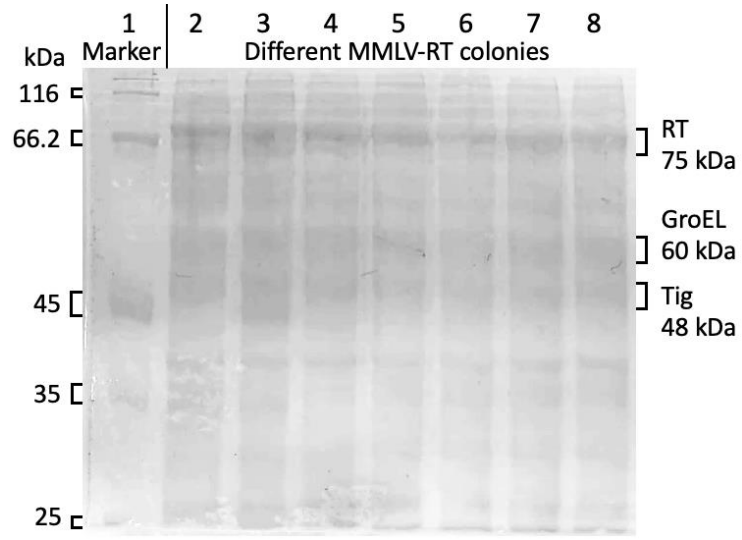


Figure 9: ***Reverse transcriptase and chaperones test expression on SDS-PAGE (Coomassie stain, +77% contrast)****. 1: protein marker; 2-8: different MMLV-RT colonies. MMLV-RT and chaperones GroEL, Tig and GroES are 75, 60, 48 and 10 kDa, respectively. The colony in well 4 was used for large-scale expression and purification.*

RNase activity test

As the library is made up of RNA and E. coli express RNases, contamination must be avoided. The MMLV-RT is therefore tested for RNase activity to determine if it is pure enough for use. High and low concentration fractions of MMLV-RT came off the column at different times, so may have different contamination levels and are tested separately. Both high and low concentration mixtures of MMLV-RT degrade the mRNA NNK15 library similarly. After an overnight incubation, the gel shows degradation of RNA (figure 10), but degradation is limited after a 1-hour incubation. Both are sufficiently free of RNase to use for RaPID selection and were combined to a final volume of 28 mL at 12.5 mg/mL before testing the activity.

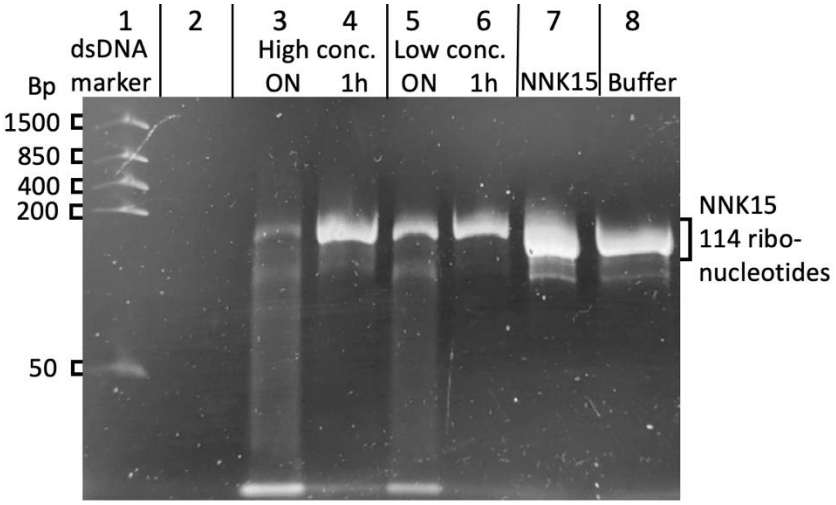


Figure 10: ***Reverse transcriptase RNase test on urea-PAGE (SYBR safe DNA gel stain)****. 1: dsDNA marker; 2: empty; 3-4: high concentration fraction of MMLV-RT overnight (ON) and 1 hour incubation, respectively; 5-6: low concentration fraction of MMLV-RT, as 3-4; 7: NNK15 only, not incubated; 8: buffer incubated overnight. NNK15 library mRNA was incubated with MMLV-RT fractions or buffer to assess RNase activity from leftover E. coli RNases.*

Reverse transcriptase activity test (performed in-house)

The MML-RT was tested for reverse transcriptase activity to check if the protein is functional. When MMLV-RT is not present, a band of RNA template-fluorescein primer complex and a band of only fluorescein primer are visible. With MMLV-RT added, another band forms which is product. To confirm that it is DNA, the third reaction was treated with RNaseH. This removes the top band of RNA template-fluorescein primer complex, while the complementary DNA band and DNA primer remain visible (figure 11). This confirms the desired reverse transcriptase activity and RaPID selection can be started.

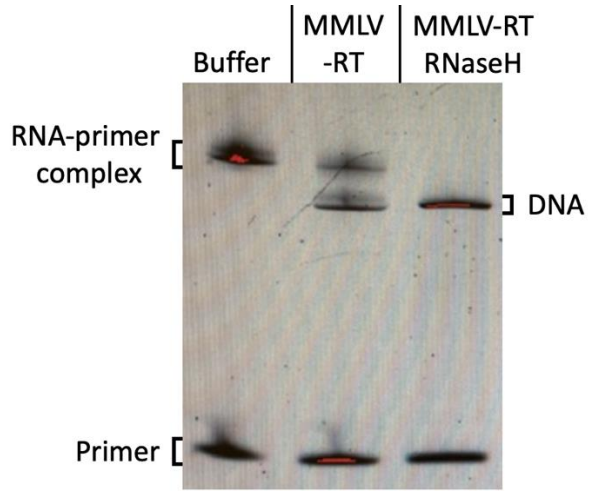


Figure 11: ***MMLV-RT activity gel (480 nm excitation/525 nm emission, detecting fluorescein)****. Only the DNA primer has fluorescein, so all bands contain DNA primer. The top band contains a complex of RNA template and DNA primer, the middle band represents DNA product of reverse transcriptase and the bottom band is DNA primer.*

RaPID selection against H1s I375F

To find peptides binding the H1s I375F, 5 rounds of RaPID selection were performed. During selection, truncated product was seen once, after round 2. The RNA output of round 1 was then purified to obtain only full-length RNA, which was used to repeat round 2 (first round 2 not shown). All rounds were performed with 3 rounds of negative selection, except for round 5, which had 9. DNA input and output levels of rounds 1 through 5 were quantified using a qPCR standard curve, then recovery was calculated by taking the ratio while accounting for dilution factors (figure 12). Starting at round 4, both libraries' positive recovery increase to above 0.1% while the negative recovery remains low, which indicates a successful selection. To confirm the end of this selection, another round was performed. In round 5, library L increases further while library D decreases, but remained sufficiently high to stop.

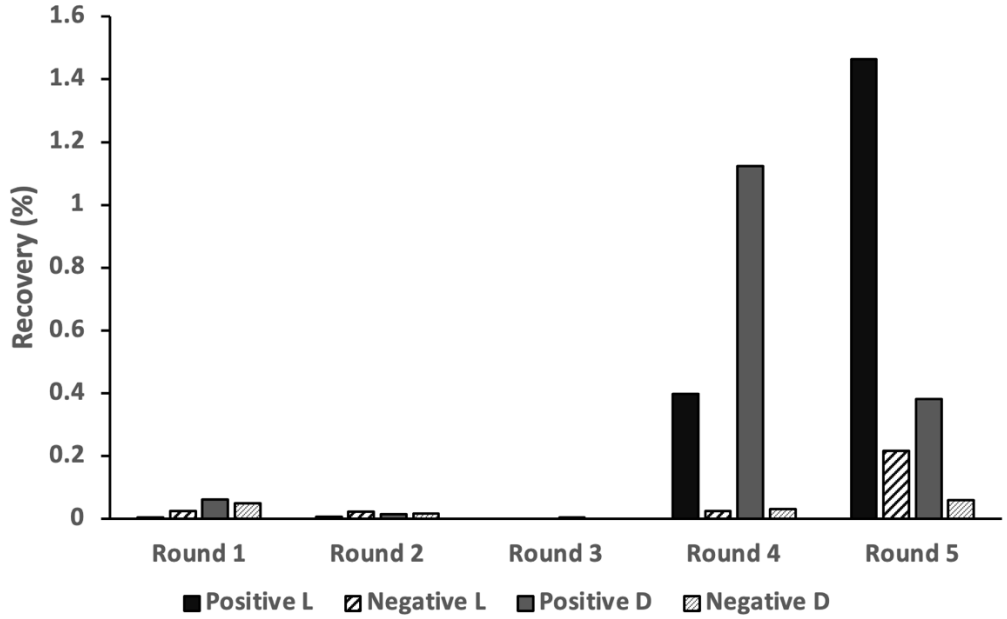


Figure 12: ***Recovery against H1s I375F in 5 consecutive selection rounds****. Recovery is the percentage of DNA present after the positive or negative selection rounds (output) as part of the total DNA before positive and negative selection (input). Positive L (black) or D (gray) is the positive selection recovery of the L and D libraries, while negative is the negative selection recovery of the L (wide black diagonal stripes) and D (narrow gray diagonal stripes) libraries.*

RaPID selection against other constructs

The output of round 3 of the H1s I375F selection, the last round before the recovery increase, was used to pan the library against other HA proteins (H1s WT, full-length H3, stem only H5 (H5s) and full-length H9) to find possible broad-acting inhibitors. Of these additional selections, 2 out of 8 libraries (H1s WT D and H5s L) had a recovery of at least 0.1% in their first round and were continued for another round. To increase the chance of finding a broad-acting peptide, the libraries of full-length H3 D and H5s D were also included for their relatively high positive to negative selection ratio (data not shown). These 4 libraries were continued for more rounds to find more specific and established peptide motifs in sequencing. H1s WT D and H5s D both see an increase in positive selection recovery with a modest increase in negative selection recovery and were thus considered finished in the second round. Full-length H3 D also had an increase in positive recovery in the second round; however, negative selection recovery rose with it. Aiming to reduce negative selection recovery while remaining a high positive selection recovery, another round was performed with 9 negative selection rounds instead of 3. Instead of only reducing the negative selection recovery, both recoveries decreased in round 3. H5s L positive recovery decreased while the negative recovery increased in round 2, so this construct was unsuccessful and stopped (figure 13).

In addition to the H1s WT, full-length H3, H5s and full-length H9 selections, the previously created H1s WT selection L and D libraries (where peptide S5 was found) were used to find peptides that inhibit both H1 WT and H1 I375F virus by using H1s I375F as target. Regardless of initial recovery, this selection would be performed for 2 rounds for both L and D libraries. When compared to round 1, the L library positive selection recovery remains stable, while the

D library positive recovery increases in round 2. Both libraries show a decrease in negative selection recovery (figure 13).

The HA proteins all use the amino acid strep-tag sequence **WSHPQFEK** to bind the magnetic beads, resulting in the immediate identification of any peptide binding to beads instead of target. Addition of strep-tag peptide binds the free bead binding sites, preventing library peptides from binding those sites and continuing to the next round without binding target protein. Thus, it theoretically prevents an increase in negative recovery such as in round 5 of the L library for H1 I375F (figure 12). Because positive selection follows after negative selection, a decrease in recovery of the negative selection is expected in the second round. To test if the proteins were displaced by adding strep-tag peptide, bead saturation of HA following a strep-tag peptide incubation was assessed. H1s I375F, H1s WT and H5s are stem only and have a triple strep-tag to bind the beads, while H3 and H9 are full-length proteins and only have a double strep-tag sequence. H1s WT and full-length H9 were chosen for initial bead saturation testing due to high availability of protein. H1s WT does not get displaced much by the peptide, while full-length H9 does (figure 14). The gel also shows displacement of H1s WT and full-length H9 without strep-tag peptide incubation because protein without strep-tag peptide incubation was incubated only when binding protein to beads, and supernatant was directly collected. As the beads are oversaturated with protein in the first incubation, protein will be present in the supernatant. To overcome this problem, beads should be washed, then incubated a second time with PBS-T only before collecting the supernatant. For selections, the stem only proteins were incubated with strep-tag peptide while the full-length proteins were not.
The results of strep-tag peptide incubation during the positive selection are inconclusive. For H1s I375F, negative selection recovery was either stable (L library) or increased (D library). For H1s WT, both L and D libraries' negative selection decreased, and for H5s it was again either stable (D library) or increased (L library) (figure 13). It may prove difficult to show an effect of strep-tag peptide incubation, as RaPID selections are highly variable and may not contain a similar degree of bead-binding peptides in each selection. Using the same target and the same library as in this project again may not yield the same results due to the large variation in sequences. Instead, a library of strep-tag and near-strep-tag peptides may be created in the RaPID library format (cyclized, 17-amino acid long peptides attached to their RNA) and subjected to a bead pulldown with or without strep-tag peptide during the incubation to determine an effect on negative selection. This would however require much setup.

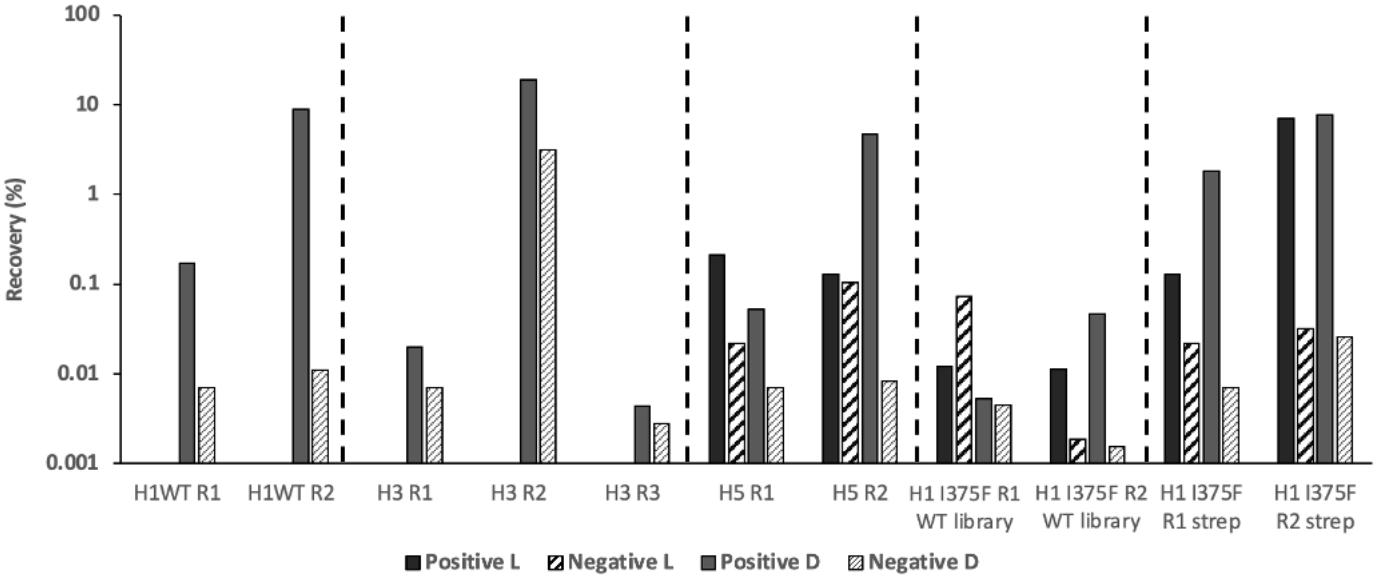


Figure 13: ***Recovery with strep-tag peptide, H1s WT library or against other targets****. Recovery is the percentage of DNA present after the positive or negative selection rounds (output) as part of the total DNA before positive and negative selection (input). Positive L (black) or D (gray) is the positive selection recovery of the L and D libraries, while negative is the negative selection recovery of the L (wide black diagonal stripes) and D (slim gray diagonal stripes) libraries. Of H1s WT, full-length H3, H5s and full-length H9, only the libraries with a positive selection recovery of at least 0.1% in the first round are shown. "H1 I375F strep" is H1s I375F with H1s I375F library and strep-tag peptide, "H1 I375F WT library" is H1s I375F incubated with the previous H1s WT selection library. All libraries for all targets were incubated with strep-tag peptide except for full-length H3.*

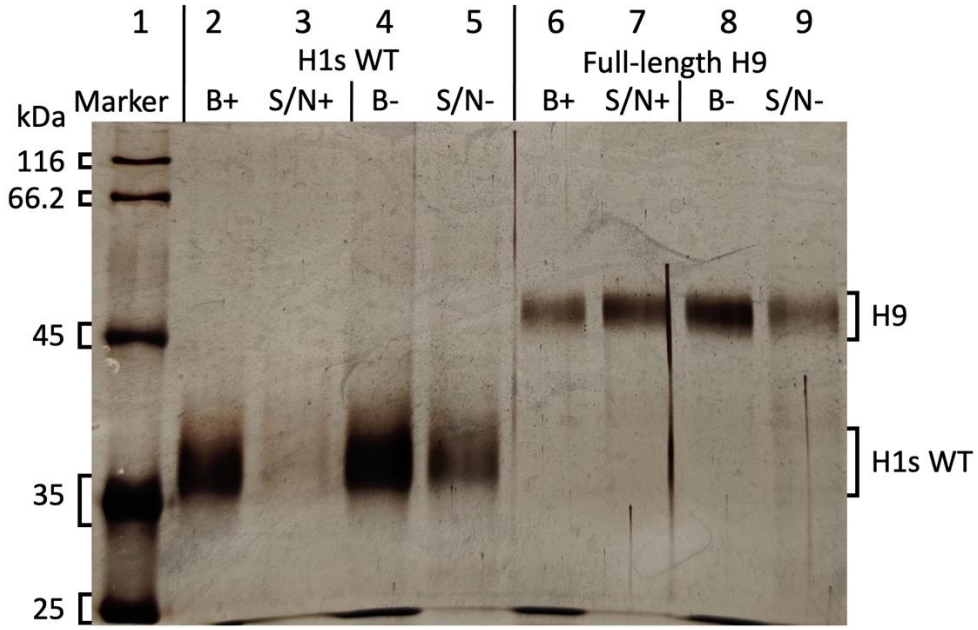


Figure 14: ***Bead saturation with strep-tag peptide on SDS-PAGE (silver stain)****. 1: protein marker; 2-5: H1s WT and 6-9: full-length H9, both in the following order: beads and then supernatant incubated with strep-tag peptide and beads then supernatant incubated without strep-tag peptide. (B = beads, S/N = supernatant.)*

Sequencing and selection of sequences

To recover peptide sequences from selections, ISeq sequencing was performed. Sequencing was successful, yielding between 2.5 x 10^4 and 1.6 x 10^5 sequences per library. The aligned top 100 most prevalent sequences of H1s I375F round 3 to 5, strep-tag and WT library rounds 1 and 2, full-length H3 D round 3 and H5s round 2, with examples of selected and not selected cluster sequence logos, can be found in the supporting information (figure 18 and 19). 17 sequences were selected as promising-looking sequence 16 was found in a selection with an unrelated target protein and therefore discontinued, most likely due to a contaminant in sequencing (table 2). Peptide 3/4 was found in the D library of all targets (H1s I375F, H1s WT, full-length H3 and H5s) and contained 2 cysteines. To see which form binds the target proteins best, two versions were made, both with one of the cysteines replaced by alanine.

*Table 2: **Chosen peptide sequences**. Peptide number and origin are shown alongside their sequence. H1 and H5 proteins are stem only, H3 is full-length. Rounds are named starting with the target protein, then (for H1s I375F only) "strep" if the library was incubated with strep-tag peptide or "WTlib" if the target was incubated with the previous H1s WT selection library, then library tyrosine stereochemical isomer "L" or "D" followed by the round number. For very similar peptides present in multiple libraries, the sequence was taken from the underlined and italic library. Sequences are in one-letter amino acid code format, with uppercase "Y" for L-Tyr and lowercase "y" for D-Tyr. *Peptide 3/4 is made twice, with alanine substitutions for either cysteine residue to identify which macrocycle best binds the target. This was only done for this peptide as this bound the most different targets (H1s I375F, H1s WT, full-length H3 and H5s). **Peptide 11 may form a disulfide bridge between C12 and C17. ***Peptide 16 was found thrice with only 2 amino acid substitutions, but later found in an unrelated selection and thus discontinued.*

Peptide	Found cluster in round	Sequence
1	*H1 WT D2*, *H1 I375F D5*, H1 I375F WTlib D2	Cyclo(Ac-yLVGTWDRHYVIKSYFC)GSGSGS-NH_2
2	*H1 WT D2*, *H5 D2*	Cyclo(Ac-yFC)LFGKTSWALYETSCGSGSGS-NH_2
3	*H1 WT D2*, *H3 D3*, *H5 D2*, *H1 I375F strep D2**	Cyclo(Ac-yLC)LNNKLSWTTVPSGAGSGSGS-NH_2
4	*H1 WT D2*, *H3 D3*, *H5 D2*, *H1 I375F strep D2**	Cyclo(Ac-yLALNNKLSWTTVPSGC)GSGSGS-NH_2
5	H1 I375F WTlib L2	Cyclo(Ac-YWHKNKYVLTYSC)LFACGSGSGS-NH_2
6	*H1 I375F L5*, *H1 I375F strep L2*	Cyclo(Ac-YLKGSWNDHRIYWASKC)GSGSGS-NH_2
7	H1 I375F strep D2	Cyclo(Ac-yLKGSWSKHKLVINGNC)GSGSGS-NH_2
8	H5 L2	Cyclo(Ac-YLVLRKLWNGQLVIEKC)GSGSGS-NH_2
9	H1 I375F D5	Cyclo(Ac-yLSGSWNSHELTYLATC)GSGSGS-NH_2
10	*H1 I375F strep D2*, H1 I375F WTlib D2	Cyclo(Ac-yLIGTWDNHVLSEC)AYCGSGSGS-NH_2
11	H5 D2**	Cyclo(Ac-yRC)LINRVTWACPDTRCGSGSGS-NH_2
12	H1 I375F WTlib L2	Cyclo(Ac-YPFTNTYEHTRSLVLHC)GSGSGS-NH_2
13	H1 I375F D5, *H1 I375F L5*, *H1 I375F strep L2*	Cyclo(Ac-YFITIVPLYFPYLIYQC)GSGSGS-NH_2
14	H1 I375F WTlib L1	Cyclo(Ac-YISLRWQKNTIVHVWHC)GSGSGS-NH_2
15	H1 I375F WTlib L1	Cyclo(Ac-YVQWWNAENCFKIAWTC)GSGSGS-NH_2
16	H5 L2***	Cyclo(Ac-YYFWYC)IVRKSRTTDECGSGSGS-NH_2
17	H1 I375F WTlib D2	Cyclo(Ac-yVKSLYPSRLLIVFFC)CGSGSGS-NH_2
18	H1 I375F WTlib L2	Cyclo(Ac-YHWRFNFAIVSDNFSLC)GSGSGS-NH_2

SPPS

Peptides 1-18 (with the exception of peptide 16 as explained above), were synthesized and purified for binding and inhibition assays. Due to an error in purification, peptide 12 was mixed with part of peptide 14 before separation on HPLC and the peptides could not be separated. Purified peptide 12 therefore contains much more peptide 14 than peptide 12, but the product is present and so it was decided to test for binding and activity and deconvolute the result based on the activity of peptide 14 alone. All other peptides are the main component of their respective fractions. Peptide-TFA conjugate was found in peptide 3. Usually, TFA-conjugated peptide and desired peptide are separated very well with HPLC. It remains unclear why the masses were only separated on LC-MS. Conjugated tert-butyl (sidechain protecting group)-peptide was found in peptides 2, 6, 10 and 11. Peptide 11 can form a disulfide bridge as it has 3 cysteines (one for macrocyclization, 2 for a disulfide bridge; yR**C**LINRVTWA**C**PDTR**C**GS) which has been found in LC-MS. Peptides 10, 15 and 17 show multiple retention times for the same peptide product on LC-MS. This may be due to a difference in macrocyclization of the ClAc-Tyr to the proximal or distal cysteine leading to a different interaction on the column. For LC-MS data, see supporting information (figure 20 and table 5).

Peptide testing

Thermal shift (qPCR and NanoDSF)

To verify binding of peptide to protein, a quick method is to use thermal shift with peptide present, where an increase in stability corresponds to a binding peptide. This approach was tried by qPCR for H1s WT and H1s I375F, however, all plots had a high noise to signal ratio, making results uninterpretable (tested reproducibility without peptide, data not shown). To uncover the problem of the noise, protein stability was tested with nanoDSF. Both proteins produced one peak, at 59.4 °C for H1s WT and 53.8 °C for H1s I375F (figure 15). As this is similar to qPCR thermal shift, denatured protein was unlikely. The different inflection points of the thermal shift and nanoDSF assays may be a result of different measurement techniques. Fresh SYPRO dye or a different dye for qPCR thermal shift may resolve the problem. The nanoDSF approach can only be used to test the stability in the presence of peptides that do not contain tryptophan, as it interferes with the measurements. Hence, 3 peptides can theoretically be tested (12, 13 and 17), but this was not yet performed.

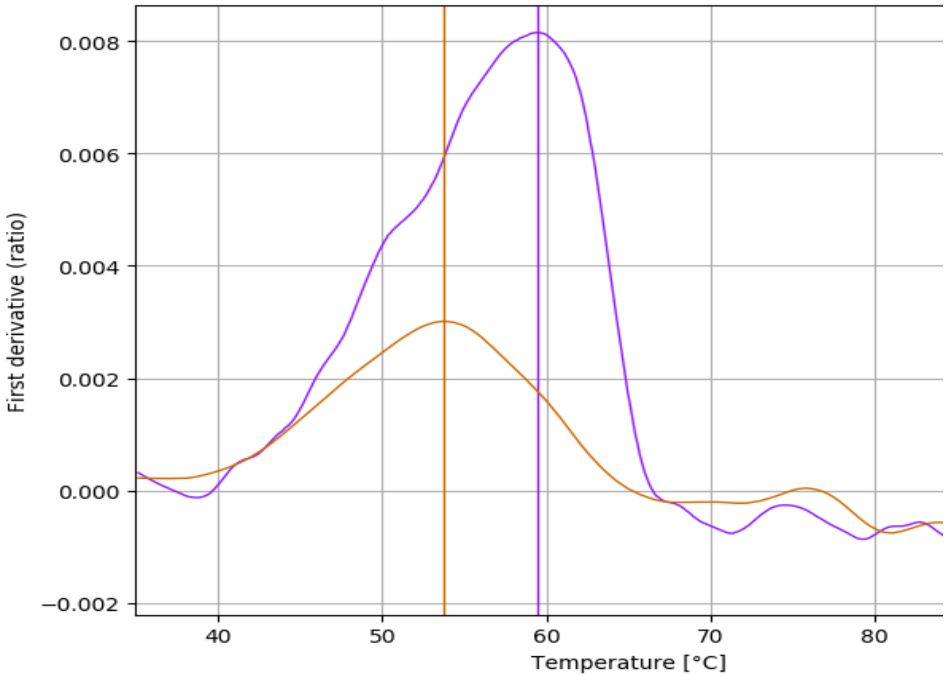


*Figure 15: **NanoDSF thermal shift**. The first derivative of the ratio of fluorescence at 350 nm/330 nm is plotted against temperature. Inflection temperatures of H1s WT (purple) and H1s I375F (orange) are 59.4 °C and 53.8 °C, respectively (calculated by nanoDSF).*

Peptide ELISA

A second assay was explored to determine peptide-protein binding, a peptide variation on ELISA where biotinylated peptide and streptavidin-conjugated horseradish peroxidase are used instead of antibodies[41,42]. This approach was tried with a previously identified peptide, H5s-binding biotinylated L2 (manuscript under review). In the initial test with peptide L2, blocking agents (spermidine, BSA or soy protein) were compared for the amount of signal they produce without target protein to determine nonspecific binding of the peptide. Soy protein resulted in a high background and was therefore immediately dismissed. Spermidine resulted in the lowest amount of background (data not shown) and was used in an assay with protein and peptide, combining concentration range 1 with 2 or 0.2 µM peptide and concentration range 2 with 0.02 or 0 µM peptide (exemplary data shown in figure 16). The measurements show a mostly peptide-dependent signal with some background from the H5s protein. Multiple strategies were performed to try to decrease nonspecific binding. Blocking agent (spermidine or BSA), blocking agent concentration, blocking incubation time, tween-20 percentage in peptide wash and amount of peptide washes in the methods were combined in all possible combinations with 2 µM peptide and without protein coating, except for 10x wash after peptide incubation for BSA and 6x wash for spermidine, which were not performed. Although some combinations were more successful than others (data not shown), background signal remained too high and thus none of the combinations were tested with protein. This project was halted because of time constraints. A different biotinylated peptide may result in lower background values and alleviate the problem, but ideally a protocol is also suitable for challenging peptides, as a sequence-independent assay. To achieve this, peptides could be synthesized with a polar peptide tag instead of biotin, such as FLAG tag (DYKDDDDK)[43] or HA tag (YPYDVPDYA)[44] to prevent nonspecific binding and then be detected by an anti-tag antibody. HA tag would not interfere with measurements in the stem only HA

proteins as the tag is found on the HA1 subunit of the protein and the stem is subunit HA2. To prevent direct protein detection the FLAG tag is preferred for the full-length proteins.

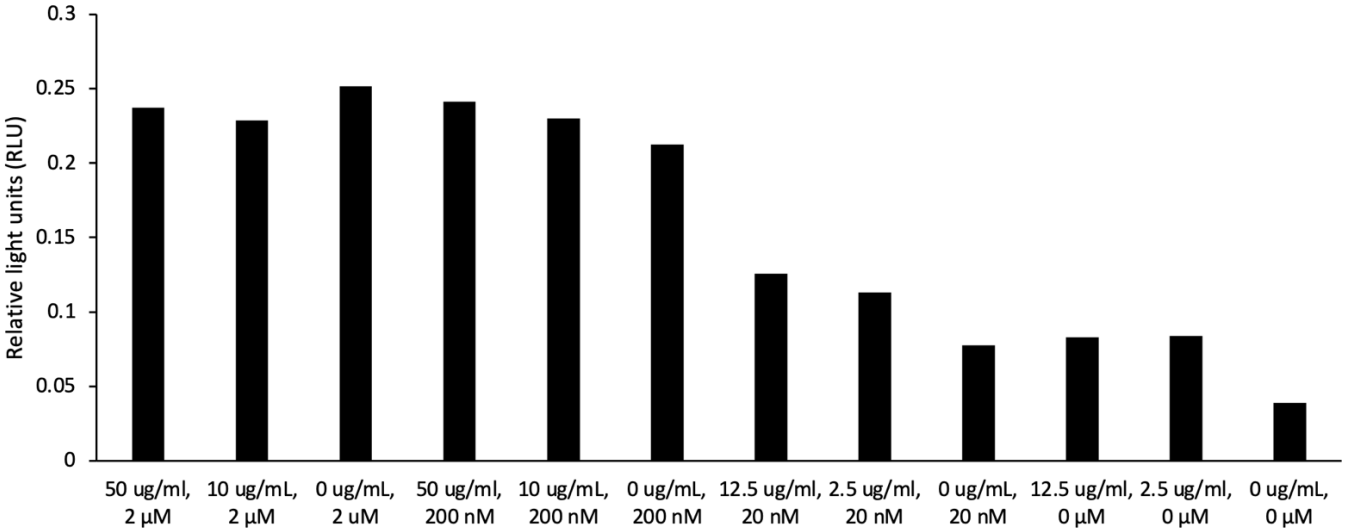


*Figure 16: **Exemplary peptide ELISA protein test data**. Protein is in µg/mL, then peptide in M. Duplicates were averaged. All combinations of protein and peptide tested follow this pattern.*

Peptide inhibitory assay

In this assay, peptides were tested for their inhibitory efficacy of the H1 WT and H1 I375F viruses. Cells were transfected with the negative-sense gene for Gaussia luciferase. When a cell is infected, the influenza RNA-dependent RNA polymerase will synthesize the complementary positive-sense gene mRNA and the host cell will create the enzyme. Upon addition of substrate, infected cells will produce light, thus more light equals more infected cells. Previously, full inhibitors were found by RaPID in a low nanomolar range. This infection assay was performed at a concentration of 2 micromolar peptide, so any peptide achieving full inhibition of full-length H1 WT or H1 I375F virus at nanomolar range should already be a full inhibitor at this high concentration. Partial inhibition was defined as a decrease in relative light units of 10 to 99% when compared to the no peptide control (DMSO), full inhibition as a decrease of over 99%. 5 full inhibitors were found for full-length H1 WT virus (5, 6, 12, 14 and 15) and 2 full inhibitors for full-length H1 I375F virus (1 and 10) (figure 17). Only peptide 5 inhibits both full-length H1 WT and H1 I375F virus, however, it only exhibits partial full-length H1 I375F virus inhibition. All but one full inhibitor (peptide 6) of full-length H1 WT virus were found in the H1s I375F selection with the previously generated H1s WT library. Peptide 12 was contaminated with peptide 14 and may thus require further testing as the results of both peptides are similar.

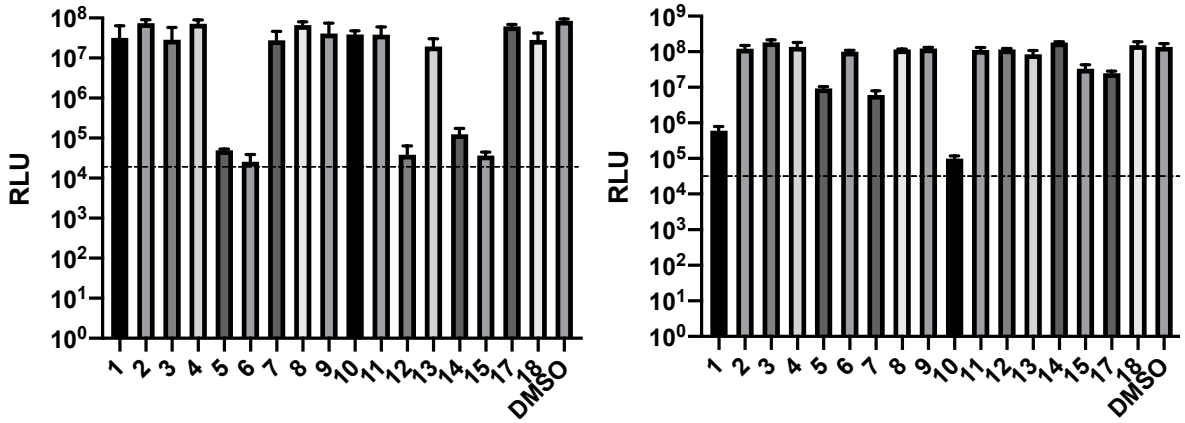


Figure 17: ***Full-length H1 WT and H1 I375F virus infection assay****. Left: virus with full-length H1 WT (Neth09), right: virus with full-length H1 I375F. Virus was first mixed with 2 µM peptide (1-15, 17 or 18) or DMSO (control) and then incubated with Henrietta Lacks cells. Cells only produce luciferase if they are infected with IAV, so more relative light units (RLU) corresponds to more infection. The dotted line shows the background RLU of non-infected cells.*

Conclusion and future outlook

In this work, the goal was to find an inhibitor which could either inhibit full-length H1 I375F, or both the full-length H1 WT and H1 I375F viruses. In this regard, the project was successful, as two full inhibitors were found for H1 I375F virus, peptide 1 and 10. In addition, a peptide was found that fully inhibits H1 WT virus and partially inhibits H1 I375F virus. Peptides 1 and 10 are full inhibitors of the H1 I375F virus, yet do not inhibit H1 WT virus, suggesting that they bind at the site of the mutation, I375F, which may be confirmed with X-ray crystallography, nuclear magnetic resonance or hydrogen deuterium exchange mass spectrometry. Crystallography is a time-consuming process while nuclear magnetic resonance requires large amounts of sample, making both costly approaches[45]. In contrast, hydrogen deuterium exchange combines fast throughput with a low detection limit[45]. Originally discovered in the H1s I375F selection, both peptide 1 and 10 were also found in the H1s WT library, suggesting they may bind H1 WT virus. The sequences of peptides 1 and 10 are similar (**yL**V**GTWD**R**H**Y**V**IKSYFCGS and **yL**I**GTWD**N**HV**LSECAYCGS, respectively) in contrast to H1 WT virus-inhibitor S5 (**y**VLFR**WDH**GTLATHWVCGS), possibly explaining why they both inhibit H1 I375F virus to a similar degree. This can be used to identify important binding residues with alanine scanning to optimize the peptides. Herein, every amino acid is replaced by alanine sequentially and tested for binding or inhibition of the target protein, similar to peptide 3/4 which had both of its cysteine residues replaced by alanine to identify the macrocycle important for binding. In addition, both peptides should be tested for cytotoxicity, to ensure they are viable for treatment or prevention of H1 I375F virus infection.

The secondary goal of this project was to find broad-acting inhibitors for H1 WT, H1 I375F, H3, H5 and H9 viruses. Infection assays with the H3, H5 and H9 viruses have not yet been performed. Table 2 suggests few of the chosen peptides will bind any of these targets, except for 2, 8 and 11 which were found in the H5s selection and 3/4, found in the full-length H3 and H5s selections. Peptides 8 and 11 were only found in the H5s selection, so may bind or inhibit H5s only. Unfortunately, the peptide found in most selections, peptide 3/4, does not inhibit H1 WT or H1 I375F viruses, which makes it unlikely to find a broad-acting inhibitor in this project. Before starting more infection assays, peptide-protein binding may be tested, for instance with thermal shift or a peptide variation on ELISA as tried in this report.

A third option is a pulldown assay similar to a regular RaPID selection round, where instead of a library, clonally pure mRNA for one peptide is used[46]. By comparison of recovery to a known binder, such as peptide S5 to H1s WT, binders can be identified. Preferably, this approach would be executed with a different type of magnetic beads to ensure target binding and prevent bead binding, which will require expression of the protein with another binding tag. However, as this assay is almost identical to the RaPID selection the peptides originated from, it includes the same biases so thermal shift and peptide ELISA were preferred.

To find a peptide which inhibits both H1 WT and H1 I375F viruses, another selection may be performed where the target protein is switched in subsequent rounds. If possible, this would include different beads for each target to keep negative selection enrichment low. As there is only one amino acid change between H1s WT and H1s I375F, but many more for the other targets, this approach may prove too stringent for a multiple-HA broad-acting inhibitor of a combination of stem only and full-length constructs. By creating the stem only protein of each HA, the selection will be targeted towards the conserved fusion peptide[29] which may allow

for protein switching between rounds. Alternatively, libraries may be enriched for all targets separately and then combined and panned against the targets with protein switching between rounds.

The long-term goal of this project is to find a peptide or a combination of peptides that inhibits the H1 of Neth09 and all of its escape mutants, such as I375F. In addition, this would be a broad-acting influenza A inhibitor affecting all viruses with potential pandemic characteristics. So far, it has only been possible to find full inhibitors for either H1 WT or H1 I375F virus, so a combination of peptides may be most feasible.

Acknowledgements

I want to thank Seino for inviting me back for my master's internship. As I remember from my bachelor's thesis, you were again very patient (especially for chemistry-related stuff), approachable and helpful. Also thank you to Vito, Minglong, Twan, Ryoji and Helena for teaching me the different lab techniques and for your advice on lab work. I thoroughly enjoyed my time in the group during lab work, at the borrels and at lab outings, of course also with Lin, David, Lisa and Wren.
I also want to thank Mirte and Xander from virology for showing me the inhibitory assay and Sebastiaan from Leiden University Medical Center for showing the practical basics of ELISA.

References

1. Hanpaibool C, Leelawiwat M, Takahashi K, Rungrotmongkol T. Source of oseltamivir resistance due to single E119D and double E119D/H274Y mutations in pdm09H1N1 influenza neuraminidase. *Journal of Computer-Aided Molecular Design*. 2020;34(1):27-37.
2. Iuliano AD, Roguski KM, Chang HH, et al. Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet*. 2018;391(10127):1285.
3. Sederdahl BK, Williams J v. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses*. 2020;12(1):89.
4. Krammer F, Smith GJD, Fouchier RAM, et al. Influenza. *Nature Reviews Disease Primers*. 2018;4(1):1-21.
5. Ferguson L, Olivier AK, Genova S, et al. Pathogenesis of Influenza D Virus in Cattle. *Journal of Virology*. 2016;90(12):5636-5642.
6. Fowlkes A, Biggerstaff M, Mchugh L, et al. Incidence of medically attended infl uenza during pandemic and post-pandemic seasons through the Influenza Incidence Surveillance Project, 2009-13. *Lancet Respir Med*. 2015;3:709-727.
7. Kilbourne ED. Influenza Pandemics of the 20th Century. *Emerging Infectious Diseases*. 2006;12(1):9.
8. Sreenivasan CC, Thomas M, Kaushik RS, Wang D, Li F. Influenza A in Bovine Species: A Narrative Literature Review. *Viruses*. 2019;11(6).
9. Noda T. Native Morphology of Influenza Virions. *Frontiers in Microbiology*. 2011;2(JAN).
10. Dou D, Revol R, Östbye H, Wang H, Daniels R. Influenza A virus cell entry, replication, virion assembly and movement. *Frontiers in Immunology*. 2018;9(JUL):1581.
11. Taubenberger JK, Kash JC. Influenza Virus Evolution, Host Adaptation and Pandemic Formation. *Cell host & microbe*. 2010;7(6):440.
12. Grohskopf LA, Sokolow LZ, Broder KR, Walter EB, Fry AM, Jernigan DB. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices—United States, 2018–19 Influenza Season. *MMWR Recommendations and Reports*. 2018;67(3):1.
13. Frensing T, Kupke SY, Bachmann M, Fritzsche S, Gallo-Ramirez LE, Reichl U. Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in MDCK cells. *Applied microbial and cell physiology*. 2016;100(16):7181-7192.
14. Stevaert A, Naesens L. The Influenza Virus Polymerase Complex: An Update on Its Structure, Functions, and Significance for Antiviral Drug Design. *Medicinal Research Reviews*. 2016;36(6):1127-1173.
15. Bourmakina S v., García-Sastre A. Reverse genetics studies on the filamentous morphology of influenza A virus. *Journal of General Virology*. 2003;84(3):517-527.
16. Kordyukova L V., Shtykova E V., Baratova LA, Svergun DI, Batishchev O V. Matrix proteins of enveloped viruses: a case study of Influenza A virus M1 protein. 2019;37(3):671-690.
17. Pinto LH, Lamb RA. The M2 Proton Channels of Influenza A and B Viruses. *Journal of Biological Chemistry*. 2006;281(14):8997-9000.
18. Benton DJ, Gamblin SJ, Rosenthal PB, Skehel JJ. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature*. 2020;583(7814):150-153.
19. Cohen M, Zhang XQ, Senaati HP, et al. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology Journal*. 2013;10(1):1-13.

20. Rust MJ, Lakadamyali M, Zhang F, Zhuang X. Assembly of endocytic machinery around individual influenza viruses during viral entry. *Nature Structural & Molecular Biology*. 2004;11(6):567-573.
21. Byrd-Leotis L, Cummings RD, Steinhauer DA. The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase. *International Journal of Molecular Sciences.* 2017;18(7):1541.
22. de Vries E, Tscherne DM, Wienholts MJ, et al. Dissection of the Influenza A Virus Endocytic Routes Reveals Macropinocytosis as an Alternative Entry Pathway. *PLOS Pathogens*. 2011;7(3):e1001329.
23. Antanasijevic A, Durst MA, Lavie A, Caffrey M. Identification of a pH sensor in Influenza hemagglutinin using X-ray crystallography. *Journal of Structural Biology*. 2020;209(1):107412.
24. Durrant MG, Eggett DL, Busath DD. Investigation of a recent rise of dual amantadineresistance mutations in the influenza A M2 sequence. *BMC Genetics*. 2015;16(2):1-9.
25. Trebbien R, Pedersen SS, Vorborg K, Franck KT, Fischer TK. Development of oseltamivir and zanamivir resistance in influenza a(H1N1)pdm09 virus, Denmark, 2014. *Eurosurveillance*. 2017;22(3):30445.
26. Uehara T, Hayden FG, Kawaguchi K, et al. Treatment-Emergent Influenza Variant Viruses With Reduced Baloxavir Susceptibility: Impact on Clinical and Virologic Outcomes in Uncomplicated Influenza. *The Journal of Infectious Diseases*. 2020;221(3):346-355.
27. Leneva IA, Russell RJ, Boriskin YS, Hay AJ. Characteristics of arbidol-resistant mutants of influenza virus: Implications for the mechanism of anti-influenza action of arbidol - ScienceDirect. *Antiviral Research*. 2009;81(2):132-140.
28. Tan GS, Krammer F, Eggink D, Kongchanagul A, Moran TM, Palese P. A Pan-H1 Anti-Hemagglutinin Monoclonal Antibody with Potent Broad-Spectrum Efficacy In Vivo. *Journal of Virology*. 2012;86(11):6179-6188.
29. Krammer F, Palese P, Steel J, Krammer F, Palese ÁP, Steel J. Advances in Universal Influenza Virus Vaccine Design and Antibody Mediated Therapies Based on Conserved Regions of the Hemagglutinin. *Current Topics in Microbiology and Immunology*. 2014;386:301-321.
30. Doud MB, Lee JM, Bloom JD. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nature Communications*. 2018;9(1):1-12.
31. Passioura T, Suga H. A RaPID way to discover nonstandard macrocyclic peptide modulators of drug targets. *Chemical Communications*. 2017;53(12):1931-1940.
32. Jiang Y, Long H, Zhu Y, Zeng Y. Macrocyclic peptides as regulators of protein-protein interactions. *Chinese Chemical Letters*. 2018;29(7):1067-1073.
33. Goto Y, Ohta A, Sako Y, Yamagishi Y, Murakami H, Suga H. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chemical Biology*. 2008;3(2):120-129.
34. Nemoto N, Miyamoto-Sato E, Husimi Y, Yanagawa H. In vitro virus: Bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Letters*. 1997;414(2):405-408.
35. Hipolito CJ, Suga H. Ribosomal production and in vitro selection of natural product-like peptidomimetics: The FIT and RaPID systems. *Current Opinion in Chemical Biology*. 2012;16(1-2):196-203.

36. Russell RJ, Kerry PS, Stevens DJ, et al. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *PNAS*. 2008;105(46):17736-17741.
37. Alexander PE, De P, Rave S. Is H9N2 Avian Influenza Virus a Pandemic Potential? *Canadian Journal of Infectious Diseases and Medical Microbiology*. 2009;20(2):e35-e36.
38. Neumann G, Noda T, Kawaoka Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*. 2009;459(7249):931-939.
39. Chevallet M, Luche S, Rabilloud T. Silver staining of proteins in polyacrylamide gels. *Nature Protocols*. 2006;1(4):1852.
40. Suga H, Goto Y, Katoh T. Preparation of materials for flexizyme reactions and genetic code reprogramming. 2011; doi:10.1038/protex.2011.209
41. Tran STPD, Hipolito CJ, Suzuki H, et al. Generation of non-standard macrocyclic peptides specifically binding TSC-22 homologous gene-1. *Biochemical and Biophysical Research Communications*. 2019;516(2):445-450.
42. Norman A, Franck C, Christie M, et al. Discovery of Cyclic Peptide Ligands to the SARS-CoV-2 Spike Protein Using mRNA Display. *ACS Central Science*. 2021;7(6):1001-1008.
43. Hamley IW. Small Bioactive Peptides for Biomaterials Design and Therapeutics. *Chemical Reviews*. 2017;117(24):14015-14041.
44. Bracha-Drori K, Shichrur K, Katz A, et al. Detection of protein–protein interactions in plants using bimolecular fluorescence complementation. *The Plant Journal*. 2004;40(3):419-427.
45. Zhang MM, Beno BR, Huang RYC, et al. An Integrated Approach for Determining a Protein-Protein Binding Interface in Solution and an Evaluation of Hydrogen-Deuterium Exchange Kinetics for Adjudicating Candidate Docking Models. *Analytical Chemistry*. 2019;91(24):15709-15717.
46. Jongkees SAK, Caner S, Tysoe C, Brayer GD, Withers SG, Suga H. Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. *Cell Chemical Biology*. 2017;24(3):381-390.

Supporting information

Primer sequences and primer-library pairs

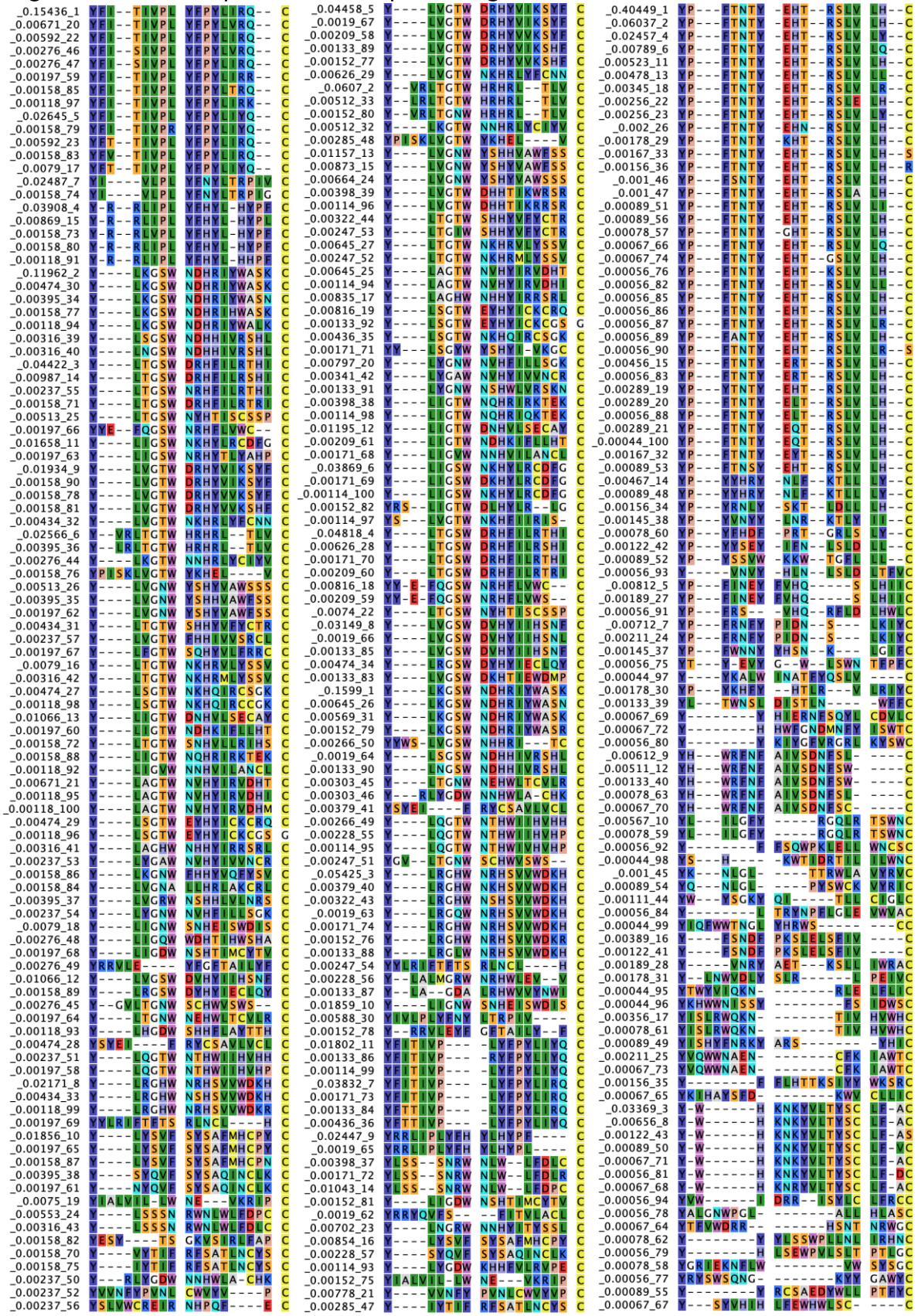
*Table 3: **Primer sequences**. For S505-S507 and N716-N729, unique parts of the sequences are underlined.*

Primer name	Sequence
CGS3an13.R39	TTTCCGCCCCCCGTCCTAGCTGCCGCTGCCGCTGCCGCA
T7g10M.F48	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATACATATG
CGS3an13Rd2.R49	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTTTTCCGCCCCCCGTCCT
Rd1T7g10M.F70	CACTCTTTCCCTACACGACGCTCTTCCGATCTTAATACGACTCACTATAGGG TTAACTTTAAGAAGGAGA
S505	AATGATACGGCGACCACCGAGATCTACACGTAAGGAGACACTCTTTCCCTA CACGAC
S506	AATGATACGGCGACCACCGAGATCTACACACTGCATAACACTCTTTCCCTA CACGAC
S507	AATGATACGGCGACCACCGAGATCTACACAAGGAGTAACACTCTTTCCCTA CACGAC
N716	CACGTCTGAACTCCAGTCACTAGCGAGTATCTCGTATGCCGTCTTCTGCTTG
N718	CACGTCTGAACTCCAGTCACGTAGCTCCATCTCGTATGCCGTCTTCTGCTTG
N719	CACGTCTGAACTCCAGTCACTACTACGCATCTCGTATGCCGTCTTCTGCTTG
N720	CACGTCTGAACTCCAGTCACAGGCTCCGATCTCGTATGCCGTCTTCTGCTTG
N721	CACGTCTGAACTCCAGTCACGCAGCGTAATCTCGTATGCCGTCTTCTGCTTG
N722	CACGTCTGAACTCCAGTCACCTGCGCATATCTCGTATGCCGTCTTCTGCTTG
N723	CACGTCTGAACTCCAGTCACGAGCGCTAATCTCGTATGCCGTCTTCTGCTTG
N724	CACGTCTGAACTCCAGTCACCGCTCAGTATCTCGTATGCCGTCTTCTGCTTG
N726	CACGTCTGAACTCCAGTCACGTCTTAGGATCTCGTATGCCGTCTTCTGCTTG
N727	CACGTCTGAACTCCAGTCACACTGATCGATCTCGTATGCCGTCTTCTGCTTG
N728	CACGTCTGAACTCCAGTCACTAGCTGCAATCTCGTATGCCGTCTTCTGCTTG
N729	CACGTCTGAACTCCAGTCACGACGTCGAATCTCGTATGCCGTCTTCTGCTTG

Table 4: ***Library index pairs for ISeq sequencing****.*

Target	Round	Forward primer	Reverse primer
H1s I375F	1 L	S505	N716
	1 D	S505	N718
	2 L	S505	N719
	2 D	S505	N720
	3 L	S505	N721
	3 D	S505	N722
	4 L	S505	N723
	4 D	S505	N724
	5 L	S505	N726
	5 D	S505	N727
	1 strep L	S505	N728
	1 strep D	S505	N729
	2 strep L	S507	N722
	2 strep D	S507	N723
	1 WT library L	S506	N716
	1 WT library D	S506	N718
	2 WT library L	S506	N719
	2 WT library D	S506	N720
H1s WT	1 L	S506	N721
	1 D	S506	N722
	2 D	S506	N723
Full-length H3	1 L	S506	N724
	1 D	S506	N726
	2 D	S506	N727
	3 D	S506	N728
H5s	1 L	S506	N729
	1 D	S507	N716
	2 L	S507	N718
	2 D	S507	N719

Aligned selection sequences and sequence logos



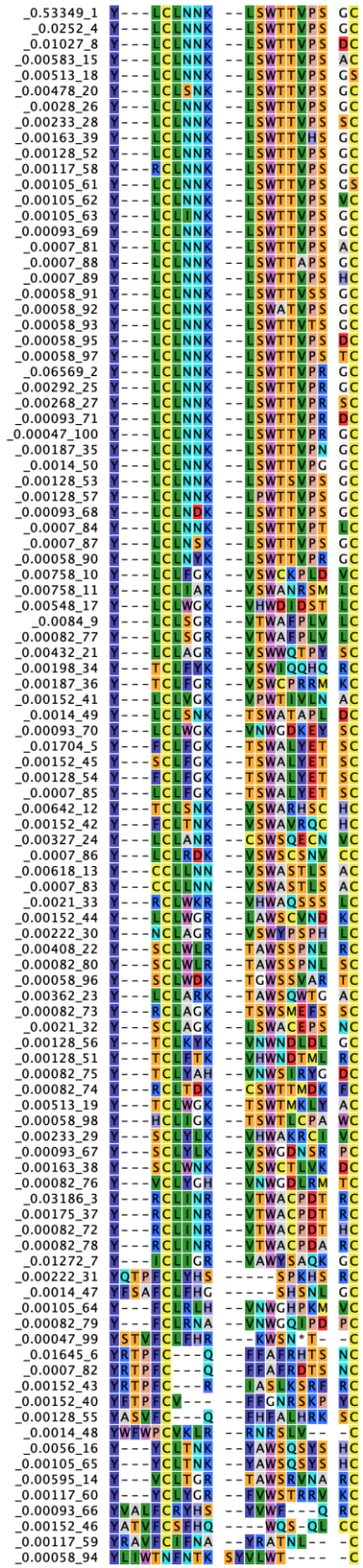
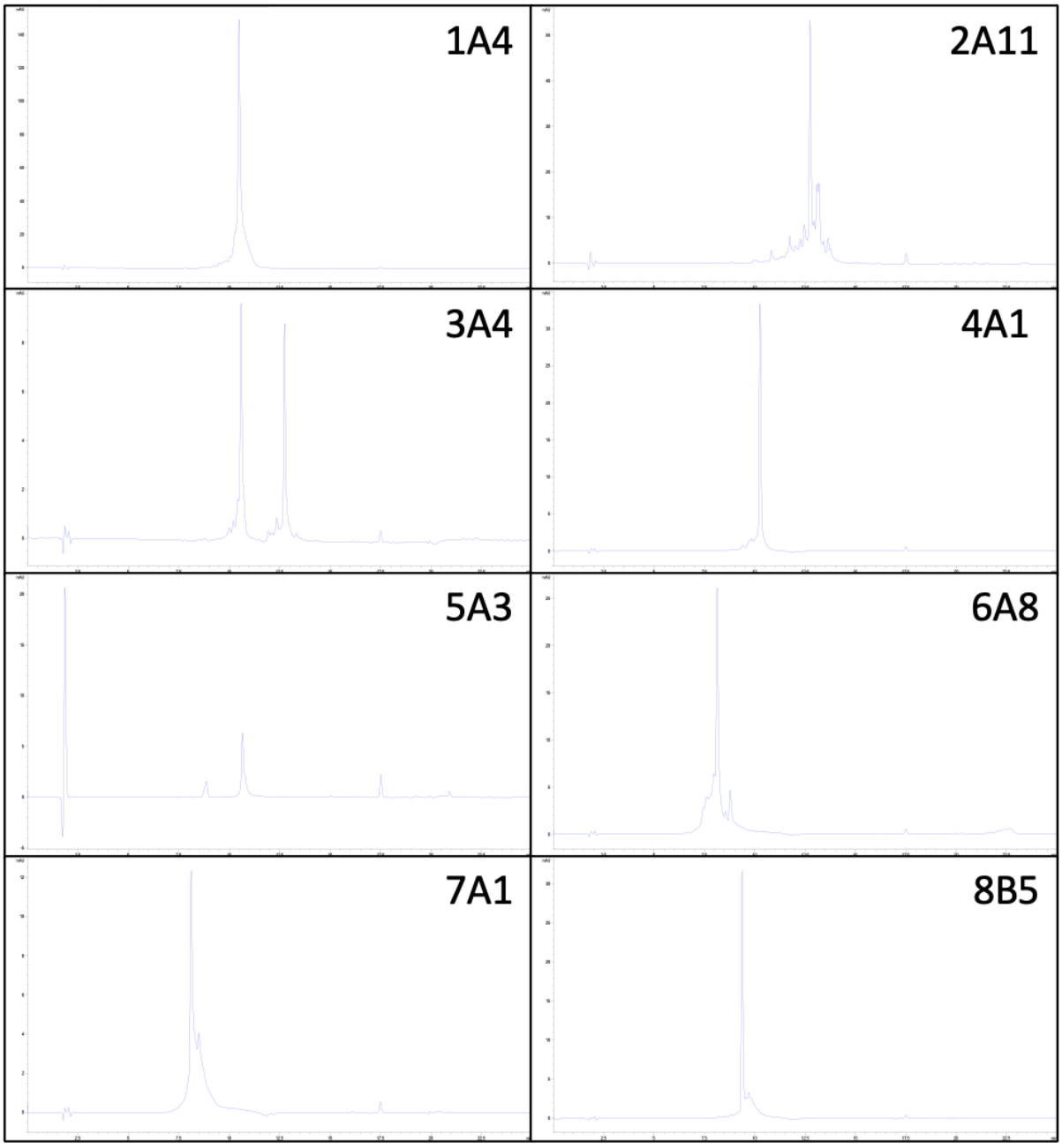


*Figure 18: **Aligned top 100 peptide sequences from Python**. In order: H1s I375F L round 5, H1s I375F L strep-tag round 2, H1s I375F L WT library round 2, H1s I375F D round 5, H1s I375F D strep-tag round 2, H1s I375F D WT library round 2, H1s WT D round 2, full-length H3 D round 3, H5s L round 2 and H5s D round 2. Sequences were named '_[abundance in round (out of 1)]_[abundance rank in round]'.*

Figure 19: ***Sequence logos examples****. Top: cluster of 12 sequences of round H1s I375F L5; middle: cluster of 15 sequences of round H1s I375F WT library D2; bottom: cluster of 4 sequences of round H1s I375F L3. The top logo was chosen for synthesis because of the large overlap in multiple sequences, ultimately being peptide 13. The middle logo was not chosen for synthesis because there is little to no overlap in sequence. The bottom logo is an example of how the strep-tag sequence shows in a real cluster.*

LC-MS analysis spectra of H1s I375F peptides

Spectra (figure 20) and masses (table 5) were obtained with an automated HPLC-MS system (Agilent LC-MS 1260 Infinity II) with a UV/VIS detector at 215 and 280 nm. Contents were separated by C18 column (poroshell 120, 4.6 x 100 mm, 2.7 µm particle size), with a 0.6 mL/minute flow creating a gradient of 10-95% solvent B over 25 minutes (solvent A: ultrapure H_2O with 0.1% formic acid; solvent B: acetonitrile with 0.1% formic acid). This was connected to an Agilent InfinityLab LC/MSD XT mass spectrometer.



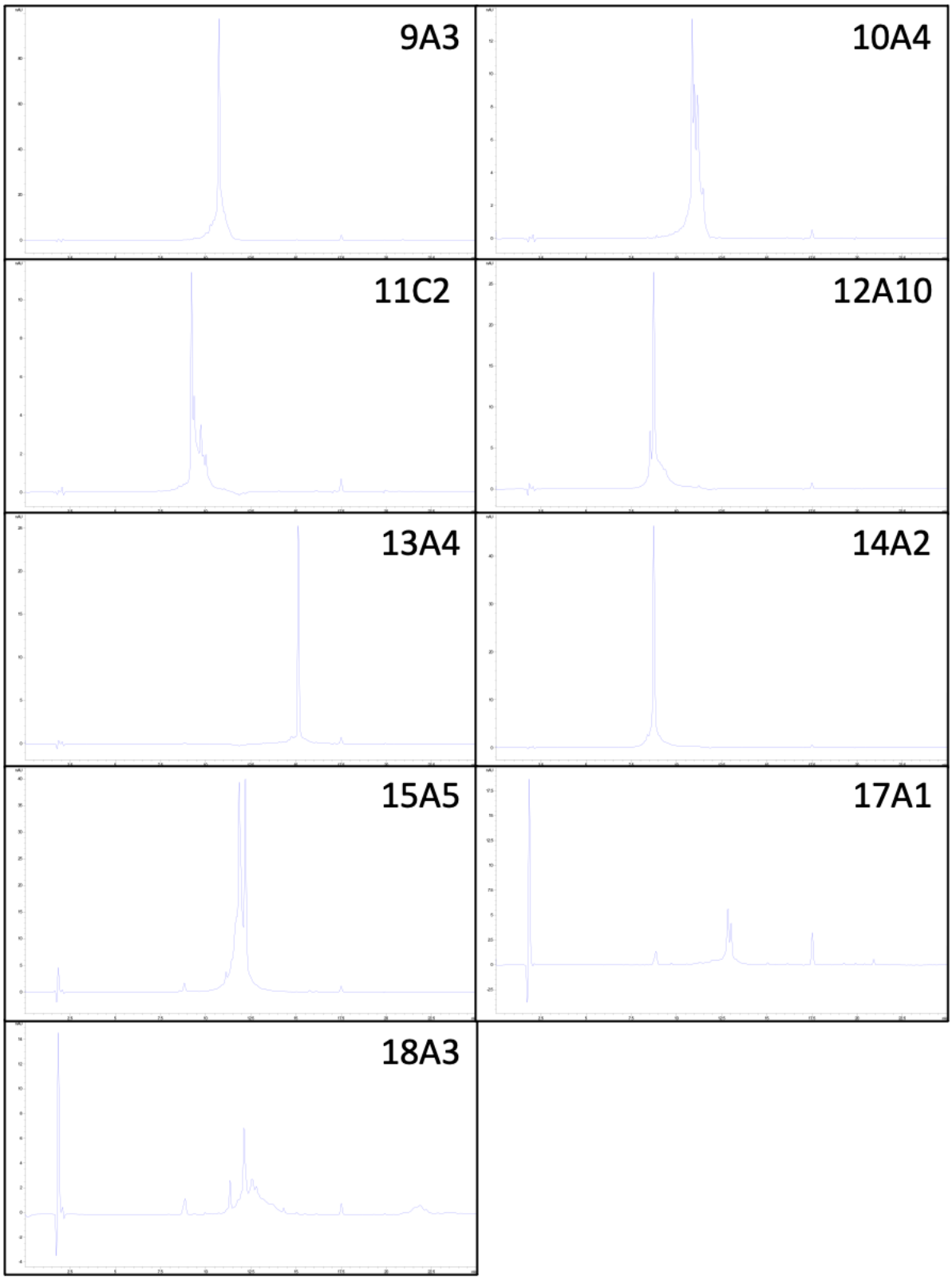


Figure 20: ***LC-MS 280 nm spectra****. For peptide 12A10, the main peak is peptide 14 contamination. For all other peptides, the main peak is the desired product. For mass spectrometry and product per peak, see table 5.*

*Table 5: **LC-MS spectra mass-per-peak**. Purity analysis results. Name/fraction is peptide number followed by HPLC collection tube location, sequence shows amino acid sequence of the peptide with lowercase y = D-tyrosine and uppercase Y = L-tyrosine, exact mass is theoretical mass of the product, calculated mass +2H is the expected readout from LC-MS, found mass is readout seen in a peak at the given retention time, retention time shows the time at which peaks came off the column, comments explain the found mass. In the comments, product: correct peptide mass was found; tert-butyl protecting group attached: a protection group used in synthesis is still conjugated to one of the amino acids after cleavage; unknown: the peak shows a uniform mass but compound is not identified; TFA-conjugate: some part of the peptide has formed an irreversible bond with TFA; various small masses: no uniform mass was found but a peak still shows; product with disulfide bridge: correct product was made, and it has used two cysteines to form a disulfide bridge in addition to macrocyclization; peptide 14: due to a purification error peptide 14 was found with peptide 12. Different retention times for the same product may be a result from cyclization to the proximal or distal cysteine, for example peptide 10 may cyclize using cysteine at position 14 or 17, altering the structure of the peptide and thus the retention time. This same principle can also be applied to peptides 15 and 17, and the tert-butyl-conjugated peptide 11. All spectra peaks correspond to those in figure 20, except peptide 16 as this peptide was discontinued after it was found in a sequencing round of a selection with an unrelated protein target and so likely a contaminant in sequencing.*

Peptide properties				LCMS spectra		
Name/ fraction	**Sequence**	**Exact mass**	**Calculated mass +2H**	**Found mass**	**Retention time (min)**	**Comments**
1A4	yLVGTWDRHYVIKSYFCGS	2333.09	1167.5	1167.4	10.5	Product
2A11	yFCLFGKTSWALYETSCGS	2201.94	1102.0	1101.8	12.8	Product
				1130.2	13.2	Tert-butyl protecting group attached
3A4	yLCLNNKLSWTTVPSGAGS	2049.98	1026.0	1026.8	10.6	Product
				1073.8	12.75	TFA-conjugate
					Other	Various small masses
4A1	yLALNNKLSWTTVPSGCGS	2049.98	1026.0	1026.8	10.3	Product
5A3	YWHKNKYVLTYSCLFACGS	2322.06	1162.0	1162.2	10.6	Product
					Other	Various small masses
6A8	YLKGSWNDHRIYWASKCGS	2310.06	1156.0		7.5	Various small masses
				1156.2	8.1	Product
				1184.3	8.8	Tert-butyl protecting group attached
7A1	yLKGSWSKHKLVINGNCGS	2130.07	1066.0	1065.9	8.1	Product
					8.55	Various small masses
8B5	YLVLRKLWNGQLVIEKCGS	2258.22	1130.1	1129.9	9.4	Product
				1130.6	9.75	Unknown
9A3	yLSGSWNSHELTYLATCGS	2127.97	1065.0	1064.7	10.8	Product
10A4	yLIGTWDNHVLSECAYCGS	2169.92	1086.0	1085.7	10.8	Product
				1085.6	11.2	Product

				1446.9	11.2	Unknown
				1114.1	11.5	Tert-butyl protecting group attached
11C2	yRCLINRVTWACPDTRCGS	2253.02	1127.5	1126.7	9.2	Product with disulfide bridge
				1127.3	9.3	Product without disulfide bridge
				1155.7	9.7	Tert-butyl protecting group attached
				1155.8	10.1	Tert-butyl protecting group attached
12A10	YPFTNTYEHTRSLVLHCGS	2264.03	1133.0	1133.2	8.6	Product
				1184.3	8.8	Peptide 14
13A4	YFITIVPLYFPYLIYQCGS	2339.16	1170.6	1170.8	15.2	Product
14A2	YISLRWQKNTIVHVWHCGS	2366.17	1184.1	1184.2	8.8	Product
15A5	YVQWWNAENCFKIAWTCGS	2345	1173.5	1173.7	11.9	Product
				1173.2	12.2	Product
16	YYFWYCIVRKSRTTDECGS	2416.06	1209.0	**Discontinued**		
17A1	yVKSLYPSRLLIVFFCCGS	2234.13	1118.1		8.9	Various small masses
				1118.2	12.9	Product
				1117.8	13.1	Product
					Other	Various small masses
18A3	YHWRFNFAIVSDNFSLCGS	2302.03	1152.0	1152.7	12.2	Product
					Other	Various small masses