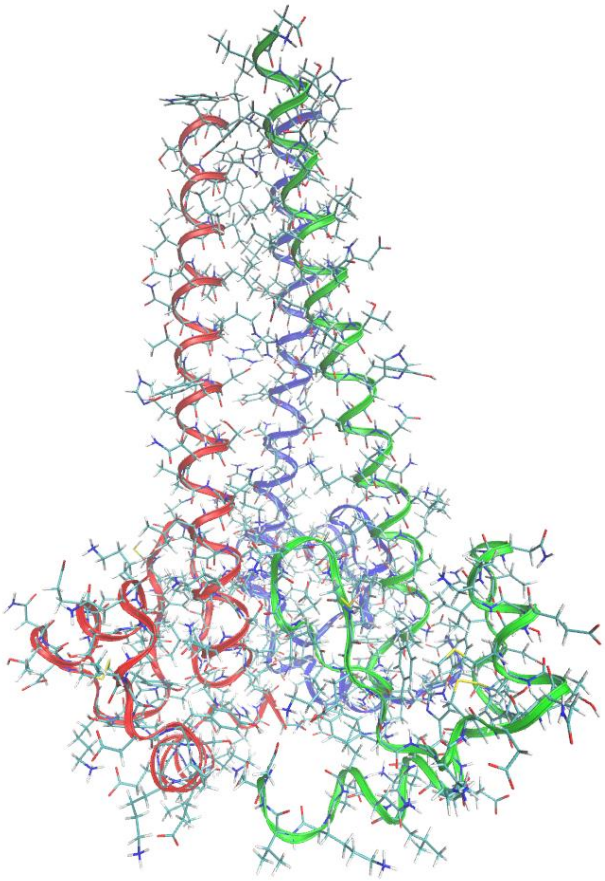
Conserved Interactions in Influenza Hemagglutinin

Bachelor Thesis



Author:	A. Jansen
Supervisors:	S. Boonstra MSc,
	prof. dr. ir. E. van der Giessen,
	prof. dr. ir. P. R. Onck
Date:	July 7, 2016

Micromechanics Group

Zernike Institute of Advanced Materials

Undergraduate School of Science, University of Groningen

university of groningen / faculty of mathematics and natural sciences / zernike institute for advanced materials

Abstract

Influenza viruses are responsible for around 250.000 fatalities every year [1]. Because influenza viruses mutate quickly, it has proved almost impossible to develop a universal antiviral influenza drug. A property shared by all influenza viruses is the presence of the protein HA (hemagglutinin) on the viral membrane. HA is a trimer, consisting of three identical proteins called monomers. HA consists of two parts: the HA1 subunit and the HA2 subunit. The HA2 subunit consists of 175 residues and plays a key role in the membrane fusion of the viral and endosomal membranes.

Using molecular dynamics simulations, mechanisms and interactions critical to HA2 stability have been identified in H1 and H5-type HA2. This was done by first removing residues 1 to 111 from HA2 to save computing time, position-restraining the N-terminal residues 112 and applying a downward constant force on the C-terminal residues 175. This way, the bottom of HA2 gets pulled apart, allowing us to look at the loss of tertiary structure and the unfolding of secondary structure of different segments. Additionally, it allows us to identify residues that provide key stabilizing interactions to the bottom of HA2. After the most suspect residues were identified, simulations were performed on mutations of these residues to confirm their influence.

The simulations on the wildtype showed that helix J dissociation is a critical event in both H1 and H5-type HA2. In both H1 and H5-type HA2, Tyr-159 and Tyr-157 seemed to be the two most critical residues. Additional simulations performed on mutations of Tyr-159 to Gly-159 showed a reduction of factor of 3.9 in macroscopic unfolding times in H1 and a reduction of a factor of 3.1 in H5. Simulations performed on mutations of Tyr-157 to Gly-157 only showed a reduction of a factor of 1.7 in macroscopic unfolding times in H1 and reduction of a factor of 1.5 in H5. This makes Tyr-159 the most critical residue in both H1 and H5-type HA2.

Tyr-159 is probably the most critical residue in H1 and H5 due to its unique location in the peptide chain. Due to its unique location, Tyr-159 is able to provide stability to both helix K and helix J. This is partially done by making H-bonds with residues such as His-142, but probably mostly by the hydrophobic interaction the aromatic side chain of Tyr-159 has with Tyr-162, Tyr-157, His-142 and Tyr-141. All of these residues are conserved in H1 and H5, explaining some of their similarity.

Concluding, Helix J dissociation as a critical event is conserved in H1 and H5-type HA2. The most critical residue, Tyr-159, is also conserved in H1 and H5-type HA2. Although the most critical residue of H3, Arg-163 [9], is not conserved in H1 and H5, the hydrophobic interaction largely responsible for helix J an helix K stability in H1 and H5 is also present in H3 and therefore conserved in all three subtypes.

Contents

Introduction

Influenza, usually known as the flu, is an infectious disease caused by an influenza virus. Influenza viruses infect over five million people each year, causing around 250.000 fatalities [1]. Although new influenza vaccines become available every year, these vaccines are often only effective for a short period of time. This is due to the high mutation rate of influenza viruses. So far, it has proved almost impossible to develop a universal antiviral influenza drug.

When searching for an antiviral drug, it is logical to look for properties of the virus that are critical to its function. Once such a property is identified, one can start to look for ways to disrupt this property. If it is possible to find the same critical property in different influenza mutations (i.e. conserved under mutation), this property could potentially be used to develop a universal antiviral influenza drug.

One property shared by all influenza viruses is the presence of the proteins hemagglutinin (HA) and neuraminidase (NA) on the viral envelope. Different influenza viruses have slightly different versions of these proteins on their envelope. This is used to classify the viruses. For example H1N1 caused the Spanish flu in 1918, while H5N1 caused the bird flu in 2004. As of today, at least 18 different subtypes of hemagglutinin [2] and 11 different subtypes of neuraminidase are known [3].

Hemagglutinin and neuraminidase both play a critical role in the functioning of the influenza viruses. Translation of influenza RNA can only take place in the nucleus of the host cell. For the viral RNA to get inside the nucleus, the viral and endosomal membranes have to fuse. The fusion of membranes is thermodynamically favorable, but has a large energy barrier. This energy barrier is overcome by the energy supplied by hemagglutinin [4]. Neuraminidase on the other hand plays a key role in releasing viruses from the host cell [5].

In other words, without the proper functioning of hemagglutinin, an influenza virus cannot deliver its RNA to the nucleus and therefore cannot replicate. Much research is being done to better understand the structure of hemagglutinin and the mechanism of membrane fusion, in order to find ways to disrupt hemagglutinin function.

The structure and stability of H3 type influenza hemagglutinin has very recently been investigated. In this bachelor thesis, I will investigate H1 and H5-type influenza hemagglutinin. When the structure and stability of these subtypes have been investigated, I will look for conserved (common) properties between these different types, critical to hemagglutinin function.

First, influenza viruses, proteins, and HA membrane fusion will be discussed. In chapter 2, the molecular dynamics simulations will be explained. In chapter 3 and 4, the results for H1 and H5-type influenza will be given, respectively. Finally, a discussion as well as a conclusion section will be given.

Chapter 1 - HA mediated membrane fusion

1.1 - The influenza virus

Influenza viruses are sphere shaped particles (figure 1). The viral RNA of influenza viruses is surrounded by a lipid bilayer called the envelope. This outer shell or envelope is covered with around 500 spike-like proteins of which around 20% is NA and around 80% is HA [6]. In this bachelor research we look at HA of influenza virus subtypes H1 and H5.

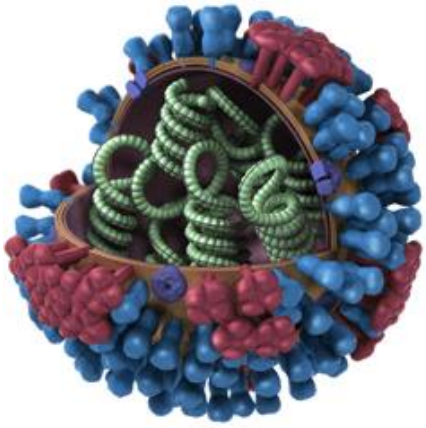


Fig. 1 - Schematic of an Influenza virus particle. The green coils inside the virus represent the viral RNA. The red proteins represent NA. The blue proteins represent HA [7].

1.2 - Proteins

Proteins are large macromolecules made of building blocks called amino acids. A total of 21 different amino acids exist (appendix A). Amino acids are often called residues in the context of peptide chains. Since we will be looking extensively at the structure of hemagglutinin it is useful to understand the different levels of structure involved in proteins.

When talking about the structure of proteins, four levels can be identified. See figure 2. The primary structure is a chain of amino acids linked together by peptide bonds. The secondary structure is the way the primary chain is folded into alpha helices or beta sheets. The tertiary structure is the way the alpha helices and beta sheets are arranged. The quaternary structure involves multiple proteins in a complex.

In the primary structure of a protein, the main-chain always has the same structure but the side-chains do not. In fact, different residues have very different side-chains. Side-chains can be positively charged, negatively charged, uncharged, polar, or hydrophobic. This makes the formation of H-bonds and salt bridges between the different side-chains possible, as well as hydrophobic interactions. These interactions make the secondary and tertiary structure of proteins possible and are the reason for their stability. Looking for *critical interactions* in HA means looking for side-chain interactions that play a *key role* in the stability of the secondary and tertiary structure of HA.

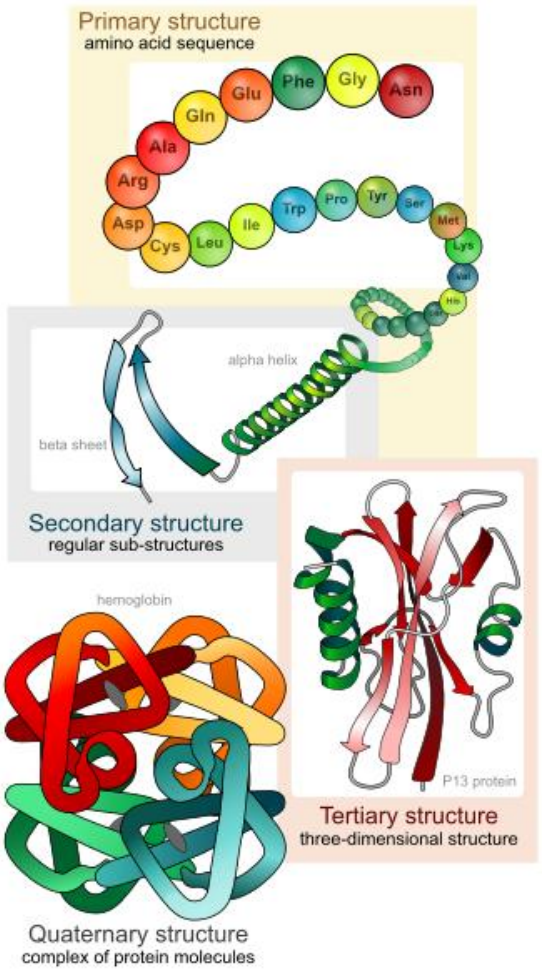


Fig. 2 - Protein structure levels [8].

1.3 - HA mediated membrane fusion

HA proteins covering the viral membrane play a key role in the fusion of the viral and endosomal membranes. The fusion of the membranes is thermodynamically favorable, but has a large kinetic barrier. This barrier is overcome by the energy supplied by conformational changes of HA [4].

HA is a trimer consisting of three identical polypeptide chains. Each chain can be divided in two parts: the HA1 subunit and the HA2 subunit. The three HA1 subunits together shield the three HA2 subunits from the environment (The HA2 subunits lie on the inside of the protein). The HA2 subunit is the part of HA that facilitates the membrane fusion [4]. The HA2 subunit consists of 175 residues and can be divided in segments running from A to K (see figure 3).

The generally proposed fusion mechanism is as follows: prior to the activation of HA, a priming and a triggering step are required. The priming step is already performed during viral replication and involves cleaving HA into the HA1 and HA2 subunits [4]. This means the HA1 and HA2 subunit are no longer connected in their primary structure. They are however kept together by side-chain interactions such as H-bonds and salt bridges until HA is triggered.

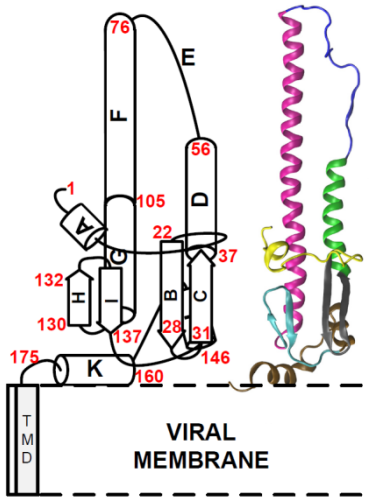


Fig. 3 - Schematic of the HA2 subunit [9].

After the viral HA binds to an endosomal membrane (figure 4A), a lowering of the pH to around 5.5 (proton ligand binding) triggers HA (4B): HA1 opens up and allows HA2 to form a triple alpha helix bundle, which extends towards the endosomal membrane. This bundle is called the extended intermediate. The end of the extended intermediate is very hydrophobic and will anchor itself in the (also hydrophobic) endosomal membrane (4C). Subsequent collapse of the extended intermediate draws the viral and nuclear membranes together (4D). The membranes are pulled towards each other until the kinetic barrier is overcome, after which membrane fusion happens spontaneously [4]. Membrane fusion eventually leads to pore formation (4E), allowing the viral RNA to enter.

In the postfusion structure of HA2, helix G lies antiparallel to helix F and is still covered with the beta sheets of segments H and I. Helix J (146 - 154) and helix K (163 - 171) however have dissociated from the top (defined as residues 112 to 140 for helix J and as 112 to 145 for helix K) and unfolded [9]. Since helix J and K are the only helices showing conformational changes between the prefusion and postfusion structure, we are mainly interested in helix J and K unfolding (loss of secondary structure) and helix J and K dissociation (loss of tertiary structure).

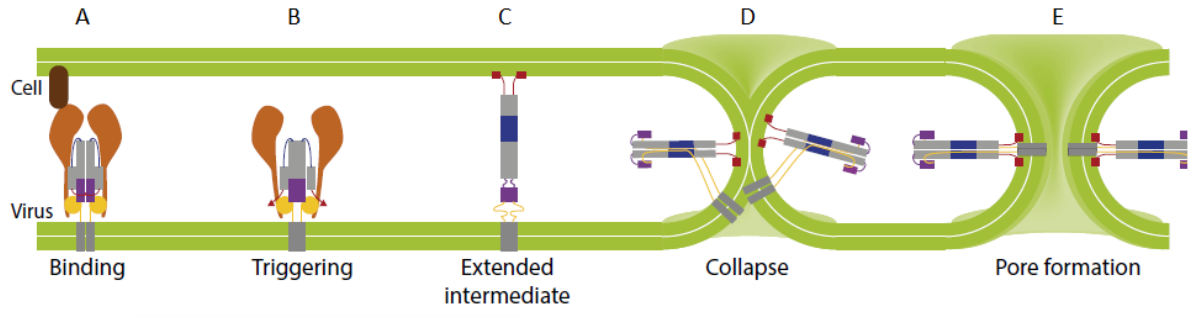


Fig. 4 - Schematic of the steps involved in membrane fusion mediated by HA [9].

1.4 - Proposed unproductive pathway of HA

By assuming that the proposed mechanism for HA mediated membrane fusion is correct, it is critical to the mechanism that the hydrophobic fusion peptide has enough time to 'find' and anchor itself to the target membrane before the globular bottom of HA2 unfolds. If the fusion peptide is not inserted into the target membrane before the globular bottom unfolds, premature collapse and subsequent inactivation of HA is the result (see figure 5). It is therefore very interesting to look for interactions in the globular bottom responsible for its stability.

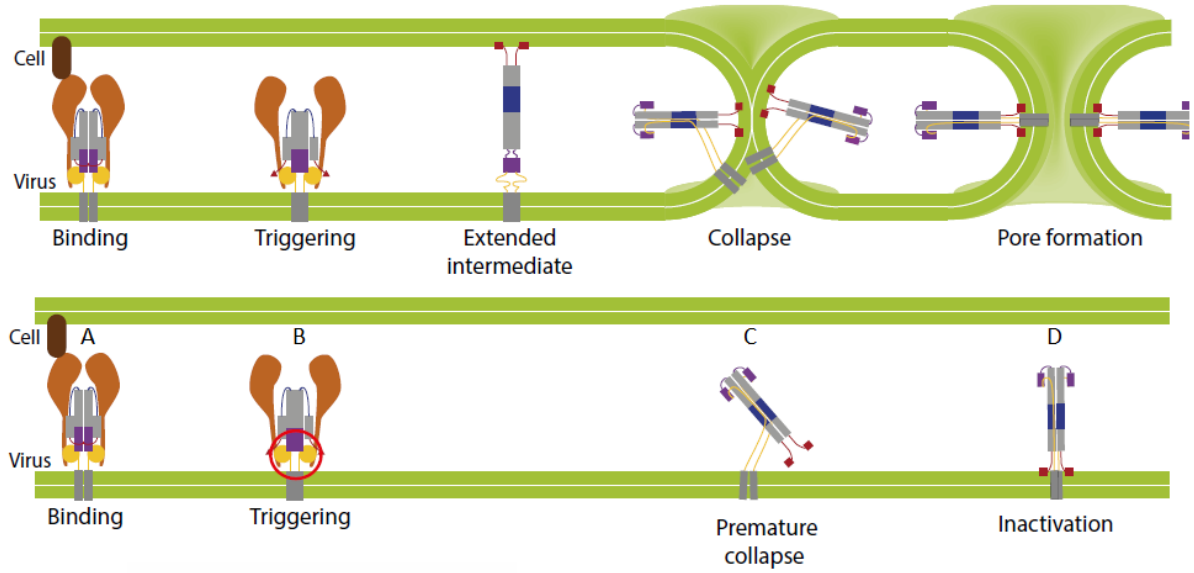


Fig 5. - Schematic of premature collapse of HA. First, HA binds to the target membrane (A). HA is triggered by a lowering in pH, and the hydrophobic fusion peptide is exposed (B). Before the hydrophobic fusion peptide has had enough time to anchor itself in the target membrane, the globular bottom unfolds causing premature collapse (C) and subsequent inactivation of HA (D) [9].

Chapter 2 - Methods

Determining which specific interactions are responsible for globular bottom stability is near impossible in vivo or in vitro. Therefore, molecular dynamics simulations are performed to investigate stabilizing interactions in the globular bottom of HA2.

As mentioned, the HA2 subunit consist of 175 residues. Since the first 111 residues do not play a role in globular bottom stability, these residues are removed as not to waste computing time. The N-terminal residues (Asp-112 for both H1 and H5) are position restrained, and a downward constant-force is applied at the C-terminal residues 175 (Gly-175 for H1, Ser-175 for H5). Pulling apart the globular bottom allows us to look at the loss of tertiary structure and the unfolding of secondary structure of helices J and K. It also allows us to identify key stabilizing interactions in the globular bottom.

Before constant-force pulling simulations can be performed however, two things need to be considered: the starting configurations for the simulations and the pulling force to be used.

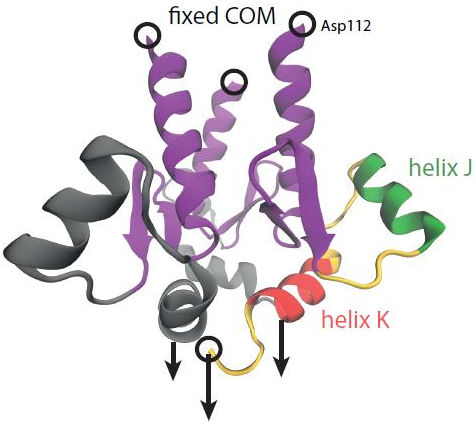


Fig. 6 - General simulation setup [9].

2.1 - Starting configurations

To acquire starting configurations for the constant-force pulling, a simulation without constraints or pulling is performed on a truncated model of HA2 comprising of residues 90 to 175. From this simulation, starting configurations for further pulling simulations are acquired.

Although this first simulation is performed using residues 90 to 175 of HA2, the constant-force pulling simulations are performed using only residues 112 to 175. This saves even more computing time, but also significantly shortens the alpha-helical top, making it unstable**.** To suppress unfolding of these parts, H-bonds responsible for the helicity of the top are now constrained as well, in addition to the N-terminal residues Asp-112.

Next to providing starting configurations, this first simulation can be used to look at which H-bonding partners are most occupied (during the simulation). This already gives us an idea of which residues are important to globular bottom stability.

2.2 - Pulling force

When determining the force to be used during constant-force pulling, several things need to be considered. First of all, if the top (helices F2 and G, see figure 3) unfolds, the force is too high. When this happens, you are effectively measuring two variables at once because now not only the unfolding of the globular bottom has influence on the extension of the protein, but also the unfolding of the top. A rule of thumb used during similar simulations on H3 simulation was: if the helical top does not unfold, the force can be considered low enough. Secondly, the protein needs to be able to unfold stochastically. If the force is too high, the protein will be pulled apart immediately regardless of any interactions, and will therefore provide no useful information.

To determine which force should be used during the constant-force pulling, a number of constant-velocity pulling simulations are performed. During these simulations the pulling happens at a constant rate instead of a constant force, thus allowing the force to vary. After the simulation has ended, the average force can be determined, as well as the helicity of the top. The time-average forces measured here already give a good indication of which forces can and cannot be used. Additionally, a comparison can be made with the time-average force measured in similar simulations on H3-type HA2, as well as the force eventually used in constant-force pulling simulations in H3.

2.3 - Further simulation details

Gromacs is used for the molecular dynamics simulations. Gromacs is a versatile molecular dynamics simulation program, allowing for fast parallel calculations. All simulations are performed using the CHARMM36 forcefield in combination with the sTIP3P (standard TIP3P) water model [10]. For all simu-lations, a temperature of 300 K at a pressure of 1 bar is used. The protein structures used for the simulations are 3LZG [11] for H1 and 2FK0 [12] for H5.

Chapter 3 - Results H1-type HA2

3.1 - Initial simulation

The initial simulation ran for a total of 31 ns. From this simulation, the average distance between the Asp-112 residues in the different chains was determined to be 1.34 nm. Starting configurations were extracted at points where the distance between the Asp-112s of the three chains lay between 1.29 and 1.39 nm. This is done because the distance between the ASP-112s fluctuates a lot during the simulation and when we position restrain them, we want the distance to be around the time-average distance. A minimum separation time of 1 ns was used between different starting configurations in order for the configurations to differ significantly.

Below, three tables are given, displaying the top 10 most occupied H-bonds for each of the chains. H-bonding partners consistently showing up have been colored. H-bonds involving residues 171 to 175 are not considered as possible candidates for being critical (because they cannot possibly provide stability to either helix K or helix J) and have therefore not been colored. Residues that only show up in one chain have not been colored either.

Table 1 - Top 10 most occupied H-bonds, chain A

H-bond partners	Occupancy	H-bond partners	Occupancy
Lys-143-Glu-165	1.28	Ser-163-Asn-129	0.26
Asn-129-Tyr-157	0.99	Lys-167-Asp-174	0.18
Asn-169-Tyr-141	0.72	Tyr-157-Gln-125	0.17
Arg-170-Asn-128	0.65	Lys-167-Glu-171	0.13
Tyr-157-His-142	0.54	Tyr-159-Asn-128	0.13

Table 2 - Top 10 most occupied H-bonds, chain B

H-bond partners	Occupancy	H-bond partners	Occupancy
Lys-143-Glu-165	1.69	Tyr-159-Asn-129	0.20
Tyr-157-Asn-129	0.96	Lys-167-Asp-174	0.18
Tyr-157-His-142	0.95	Tyr-157-Gln-125	0.16
Asn-169-Tyr-141	0.74	Arg-170-Asn-128	0.15
Gln-125-Val-152	0.39	Lys-167-Glu-171	0.15

Table 3 - Top 10 most occupied H-bonds, chain C

H-bond partners	Occupancy	H-bond partners	Occupancy
Lys-143-Glu-165	1.39	Lys-172-Glu-139	0.41
Asn-129-Tyr-157	0.96	Lys-131-Asp-174	0.27
Tyr-157-His-142	0.64	Arg-170-Asn-128	0.26
Tyr-159-Asn-128	0.60	Tyr-159-Asn-129	0.23
Asn-169-Tyr-141	0.59	Arg-170-Asn-129	0.17

We see five H-bonds consistently showing up. The associated residues of interest are Asn-169, Glu-165, Tyr-159 and Tyr-157. All four of these residues can be considered suspect for being critical. Asn-169 and Glu-165 are part of helix K (residues 163 to 171) and have therefore probably more to do with helix K stability, while Tyr-157 and Tyr-159 have probably more to do with helix J stability since they lie between helix J (residues 146 to 154) and helix K. These results should be kept in mind when looking at the results of the constant-force pulling simulations.

3.2 - Constant-velocity simulations

A total of 7 constant-velocity pulling simulations have been performed: 4 at a rate of 1 Å/ns, 3 at 0.2 Å/ns. At this point it should be mentioned that a mistake was made during 3 of the 4 simulations pulling at 1 Å/ns, and all 3 of the 0.2 Å/ns simulations. The wrong terminals were specified in the input files for Gromacs, causing the terminal ends (Gly-175) to be charged instead of neutral. This led to a force peak in the first 3 simulations at 1 Å/ns. This force peak can be seen at the start of the first 3 runs at 1 Å/ns (figure 7b): the average force rises to about 100 pN, and then drops down to zero. These simulations can still be used for determining the average force, since after the peak the terminal ends have been pulled away from the globular bottom and cannot make any interactions anymore. From this point onward, the simulations can be used to determine the average pulling force. The results of these constant-velocity simulations will now be presented.

Constant-velocity pulling simulations at 1 Å/ns

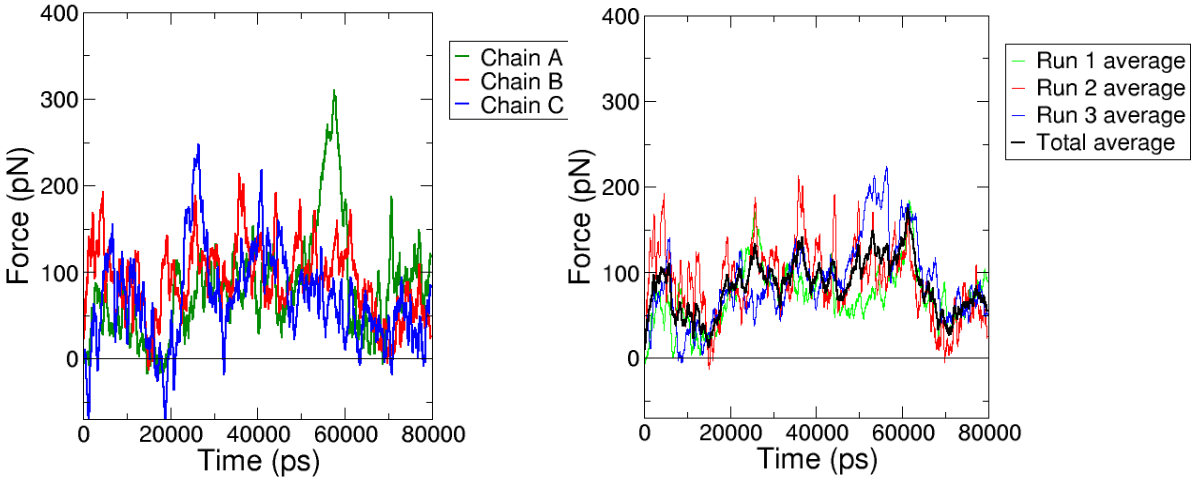


Fig. 7 - Constant-velocity pulling at 1 Å/ns. A) The pulling force on the individual chains in run 1 (2 and 3 look similar). B) The average force of runs 1, 2 and 3, as well as the total average pulling force. A peak and subsequent drop to zero are caused by incorrect specification of the charge of the end terminal. The force responsible after this peak can still be used to compute the time-average force needed for pulling at this rate.

Figure 8 shows run 4. In this run the correct charge of the end terminals was specified.

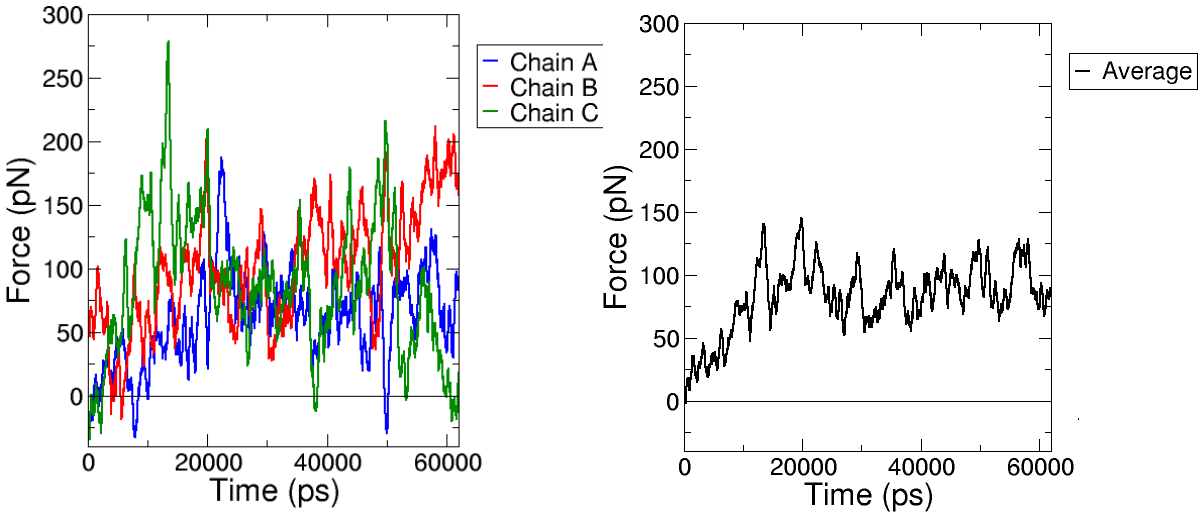


Fig. 8 - Constant-velocity pulling at 1 Å/ns, run 4. A) pulling force on the individual chains. B) average force of the run. No force peak and no subsequent drop to zero are visible in the average force.

Table 4 shows the time-average force for each chain for all 4 runs, as well as the time-average force. The 'after peak' point is defined as the point after the peak where the average force drops to zero.

Table 4 - Time-average force of constant-force pulling simulations at 1 Å/ns

Run	**After peak point (ns)**	**Time-average force per chain (pN)**		**Time-average force (pN)**
1	15	A	91	83
		B	89	
		C	70	
2	17	A	103	84
		B	67	
		C	83	
3	10	A	78	94
		B	85	
		C	119	
4	Whole run is used.	A	61	82
		B	100	
		C	85	
Time-average force (all runs)				**86 pN**

Constant-velocity pulling simulations at 0.2 Å/ns

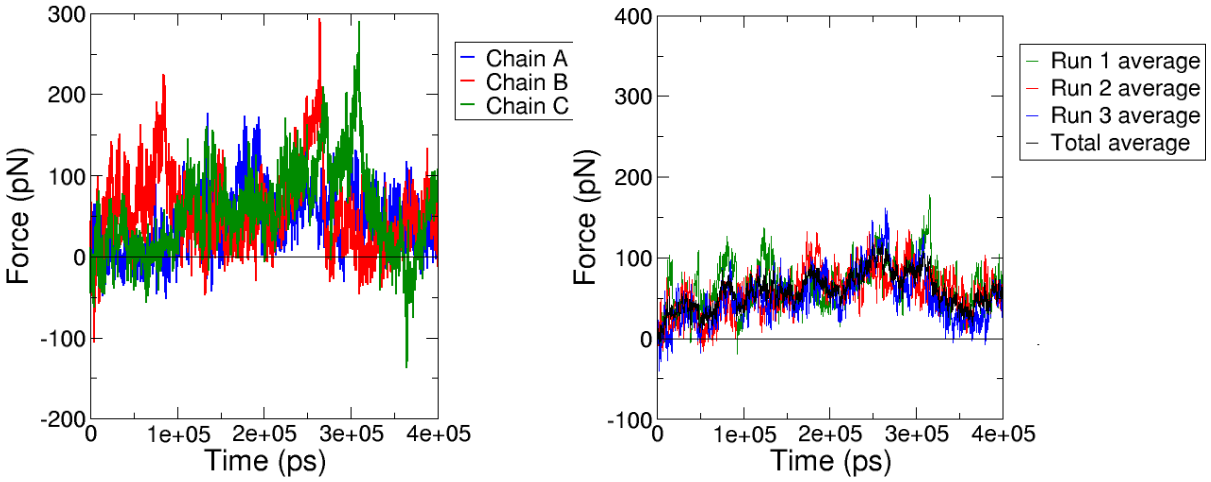


Fig. 9 - Constant-velocity pulling at 0.2 Å/ns, run 3 (1 and 2 look similar). A) Pulling force on the individual chains. B) Average force of runs 1, 2 and 3, as well as the total average pulling force. No force peak and no subsequent drop are visible in the (total) average force.

Table 5 - Time-average force of constant-force pulling simulations at 0.2 Å/ns

Run	**Chain**	**Time-average force per chain (pN)**	**Time-average force (pN)**
1	A	54	66
	B	107	
	C	38	
2	A	70	57
	B	31	
	C	71	
3	A	44	51
	B	54	
	C	56	
Time-average force (all runs)			**58**

Concluding, we found the total time-average force during constant-velocity pulling at 1 Å/ns to be 86 pN, and the time-average force at 0.2 Å/ns to be 58 pN. At no point during any of the simulations did the helicity of helix G drop below 0.8. This implies that at these forces the top does not unfold. This indicates that the average forces measured here are small enough.

To get a better idea of what force to use, a comparison can be made with H3 type influenza. During constant-force pulling simulations on H3, a force of 100 pN was used. This force worked well for H3 and yielded usable results. During constant-velocity simulations performed on H3, an average force of 106 pN was measured at a pulling rate of 1 Å/ns, and an average force of 84 pN at a pulling rate of 0.2 Å/ns.

Comparing the average forces found during constant-velocity pulling in H3 to the force used during constant-force pulling in H3, the following ratios can be established:

At 1 Å/ns 100/106 = 0.94
At 0.2 Å/ns 100/84 = 1.19

If we now apply these ratios to H1, we get the following:

At 1 Å/ns 0.94*86 = 81 pN
At 0.2 Å/ns 1.19*58 = 69 pN

Averaging these two values out, we get an average force of 75 pN. It was chosen to round this to above and use 80 pN for the constant-force simulations since the top did not unfold at 86 pN, indicating a force of 80 pN is low enough but also not too high.

Although it is possible to arrive at slightly different forces using different methods, the force of 80 pN worked quite well and gave usable results as we will see in the next section.

3.3 - Constant-force pulling simulations at 80 pN

As mentioned, we are mainly interested in helix J and K dissociation (loss of tertiary structure) and helix J and K unfolding (loss of secondary structure). The helix dissociation time is defined as the last moment at which the minimum distance between all heavy (non-hydrogen) atoms in the helix and the top is less than 3.5 Å. Furthermore, The helix unfolding time is defined as the last moment at which the helicity of the helix is larger than zero, preceding at least 10 ns of zero helicity. The amount of correlation between helix dissociation and unfolding and macroscopic unfolding gives us a good idea of how critical helix unfolding and dissociation are to the stability of HA2.

A total of 10 constant-force pulling simulations have been performed using a force of 80 pN. Within the simulated time of 1000 ns, all runs reached macroscopic unfolding except for run 10. The macroscopic unfolding time is defined as the moment at which helices J and K of the fastest chain have both dissociated, and a distance of 10.3 nm between the pulling groups (Arg-112 and Gly-175) is achieved. Ideally, a slightly larger distance of 11.5 nm is used (as was the case with H3 simulations) to ensure helix J and K always dissociate before macroscopic unfolding, but due to periodic boundary conditions this was not possible (the simulation box was defined a little too small). Run 10 got stuck in a local energy minimum (caused by inter-chain interactions at the bottom of the protein) and was stopped after running for 1108 as not to waste computing time. Macroscopic unfolding times of the other nine simulations were on average 250 ± 86 ns.

Figure 10 shows the correlation between macroscopic unfolding and both helix dissociation and helix unfolding of helices J and K.

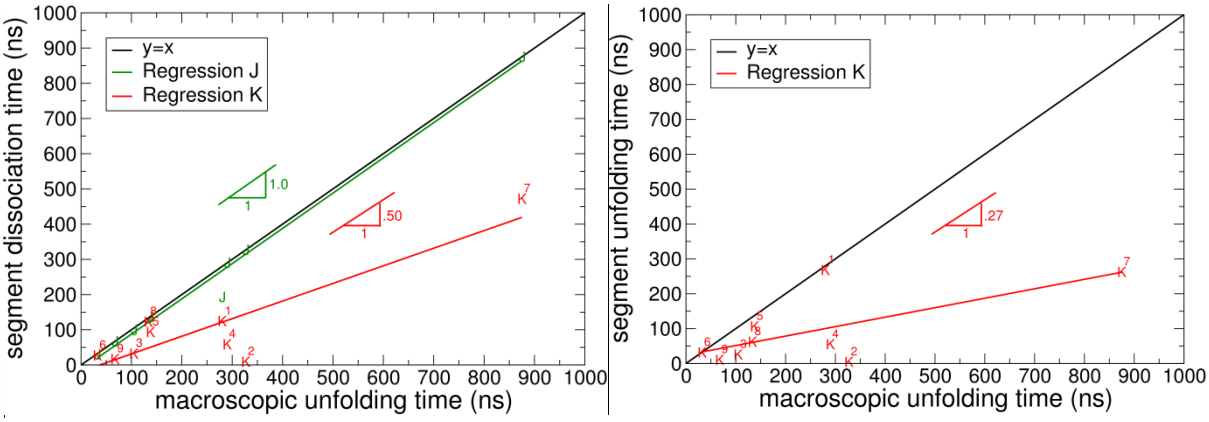


Fig. 10 - Helix J and K dissociation and unfolding. A) Helix dissociation. B) Helix unfolding.

Helix J dissociation is highly correlated with macroscopic unfolding with a CC (correlation coefficient) of 0.99 and a RC (regression coefficient) of 1.0. Helix K dissociation is less correlated to macroscopic unfolding with a CC of 0.89 and a RC of 0.50. The RCs of helix J and K differ quite a lot, indicating no fixed delay between helix J and helix K dissociation. On average, helix K dissociates in half the time of helix J dissociation. Helix K always dissociates before helix J.

Helix J never unfolds before the macroscopic unfolding event and is therefore not shown in figure 10b. Helix K unfolding has a CC of 0.68 and a RC of 0.27 and is therefore not correlated to macroscopic unfolding.

It should be mentioned that no value for helix J dissociation was acquired in run 8, even though helix J did dissociate from the top. This has to do with the definition used for helix dissociation in Gromacs. In run 8, at least on heavy atom of helix J was still less than 3.5 Å away from the top when the macroscopic unfolding distance was reached. It was chosen not to include the data point of helix J of run 8 in the graph.

It should also be mentioned that in run 1, macroscopic unfolding did not follow immediately upon helix J dissociation (see figure 10a). This was caused by inter-chain interactions at the bottom of the chain, delaying macroscopic unfolding and causing the data point of helix J not to lie on the y = x line.

When the pulling force is applied, the chains start to unfold at the C-terminal end. Residues will start to dissociate, starting at the bottom and going upwards along the chain. The time at which a residue dissociates from (makes its last contact with) the top of the protein has been determined for residues 146 to 175. The dissociation time of a residue is defined as the last moment a H-bond is present between the residue and the top of the protein. The distinction between the top and the bottom of the protein is used because vertical H-bonds deliver most of the stability against the vertical pulling.

To determine which residues are critical, a number of points are considered:

- *Correlation*. If a residue with a stabilizing interaction is reached, the unfolding stops until the interaction is broken. After the most critical contact is broken, the chain unfolds without further delay. Therefore, the most critical contact will have the highest correlation with the macroscopic unfolding event.
- *Regression*. A residue with a high regression coefficient indicates an absence of a further delay after the residue dissociates. A high regression coefficient therefore points towards the residue under consideration being critical.
- *F-value*. The F-value is defined as the mean regression sum of squares divided by the mean error sum of squares. A high F-value indicates contacts show consistency over all runs by a low spread around the regression curve. Residues with a high F-value are more likely to be critical.
- *Occupancy*. For a critical residue, not only the last contact, but also the amount of contact(s) before residue dissociation is important. Residues with a high occupancy (up until their last contact) are therefore more likely to be critical.
- *Initial Simulation*. The initial simulation already provided us with information about the occupancy of different residues and should be kept in mind.

Figure 11 shows the correlation and regression coefficients of residues 146 to 175, their average occupancy and their F-values.

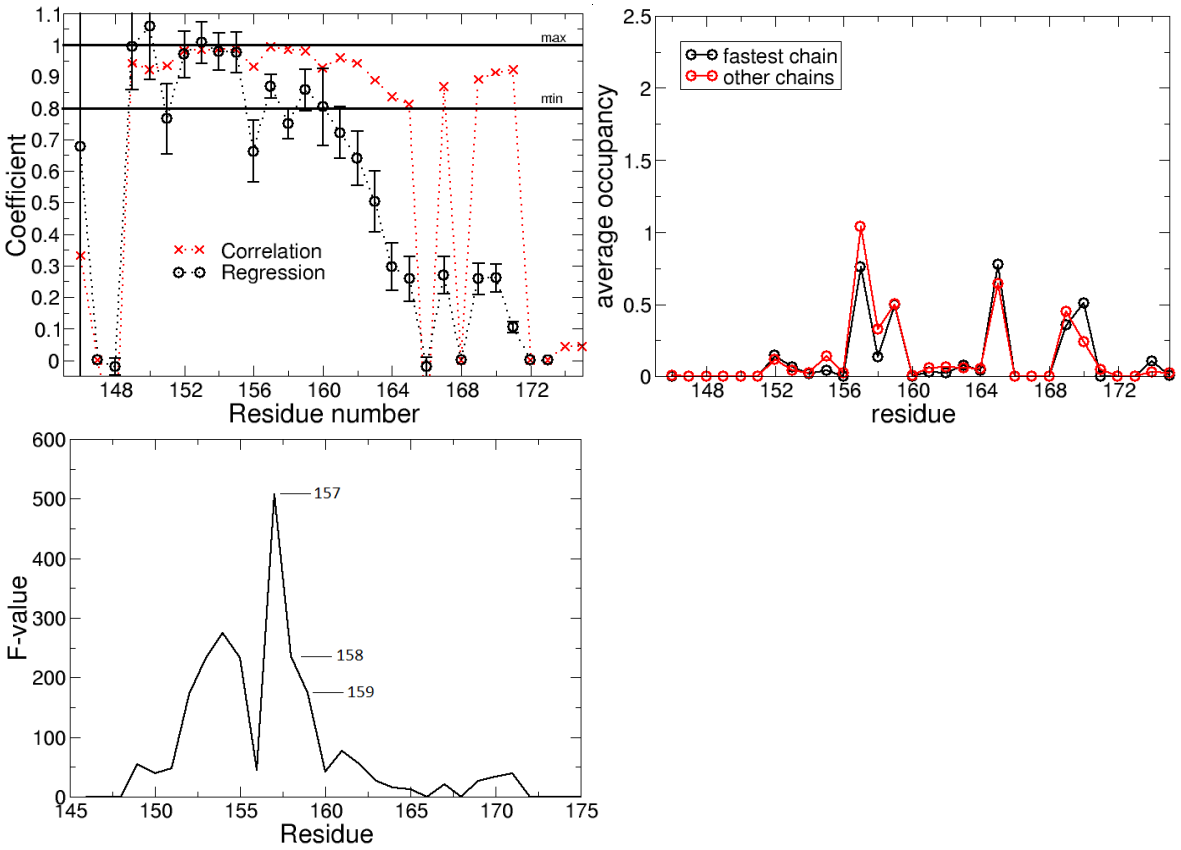


Fig. 11 - A) Correlation and regression coefficients for the LVC-regression (LVC stands for Last Vertical Contact) of residues 146 - 175. B) Average occupancy up until the LVC. Both the fastest chain and the other chains are shown. C) F-values for residues 146 to 175.

The residues with both their CC and RC convincingly higher than 0.8 are Tyr-159 (CC = 0.98, RC = 0.86) and Tyr-157 (CC = 0.99, RC = 0.87). Once the most critical contact is broken, the chain unfolds without further significant delays. This naturally leads to residues after the most critical residue having high CCs and RCs with as well, even though they are not critical at all. So although it looks like residues 152 to 155 could be critical since they all have their CCs an RCs above 0.8, this is to be expected if either Tyr-157 or Tyr-159 is the most critical residue. In fact, this reasoning could even apply to Tyr-157. If Tyr-159 is the most critical residue, it would not be strange if Tyr-157 has almost the same CC and RC values as Tyr-159. This will however not be the case if Tyr-157 is the most critical residue. Therefore, Tyr-159 seems to be the most critical residue when looking solely at the CCs and RCs.

Looking at the F-values (figure 11c), the three highest ones are Tyr-157 (508), Asp-158 (235) and Tyr-157 (174). Since The F-value of Tyr-157 is almost three times as large as the F-value of Tyr-157, the F-values point towards Tyr-157 as being the most critical residue.

When looking at the average occupancy (figure 11b), we see 4 peaks. From right to left, the first peak we see is associated with residues Arg-170 and Asn-169. The second peak is associated with residue Glu-165. Since all three of these residues are part of helix K, these residues probably play a role in helix K stability. Since helix K dissociation is not a critical event, this is not so interesting. More interesting are the highly occupied residues Tyr-157 and Tyr-159. These two residues sit in between helix J and helix K and probably play a role in helix J stability. The occupancy of Tyr-159 is the same in both the fastest and the other two chains, but the occupancy of Tyr-157 clearly shows a reduction when it belongs to the fastest chain. This indicates that a lower occupancy of the H-bonds of Tyr-157 leads to earlier chain unfolding. This is an indication that Tyr-157 might be more critical than Tyr-159.

Looking in the protein structure during the simulations before helix K dissociation (figure 12), we see that Glu-165 and Asn-169 indeed play a large role in helix K stability (12b), as we already suspected from the average occupancy. We also see a hydrophobic interaction between the aromatic rings of Tyr-162, Tyr-159, Tyr-157, His-142 and Tyr-141. This hydrophobic interaction may contribute significantly to the stability of helix K.

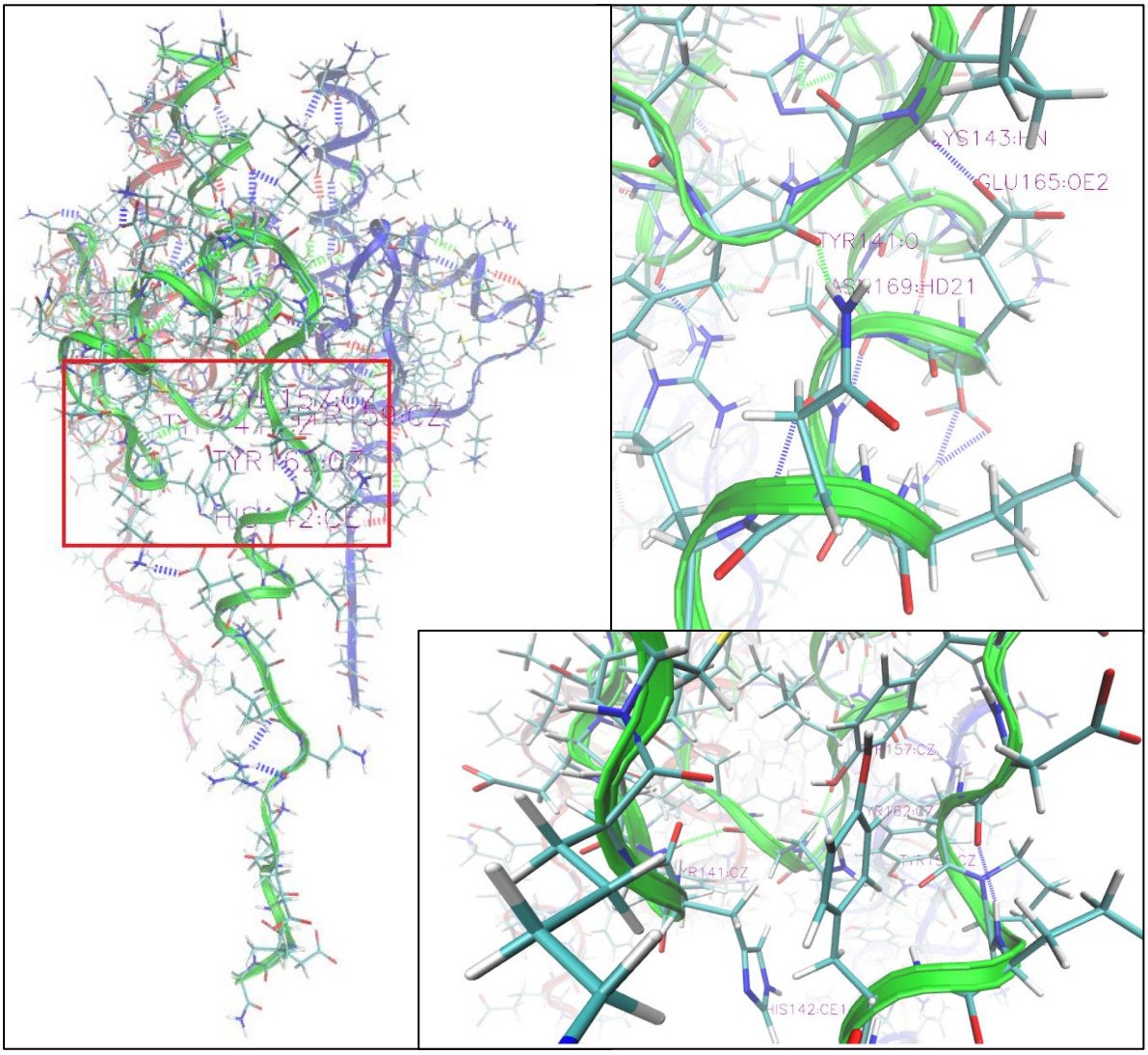


Fig. 12 - Run 3 at 21 ns. A) The red box is where we zoom in. B) Hydrogen bonds between Asn-169 and Tyr-141 and Glu-165 and Lys-143 provide stability to helix K against the vertical pulling. C) The aromatic side-chains of Tyr-157, Tyr-162 and His-142 cluster together due to their hydrophobic interaction.

Looking in the protein structure after helix K has dissociated but before macroscopic unfolding (figure 13), we see some interesting things happening. We see a hydrophobic interaction between Tyr-159, Tyr-157, His-142 and Tyr-141. No H-bonds are seen. The hydrophobic interaction between the aromatic rings seems to be the only interaction preventing helix J dissociation. The aromatic rings are so close together that pi-stacking interactions may very well contribute to the effect as well.

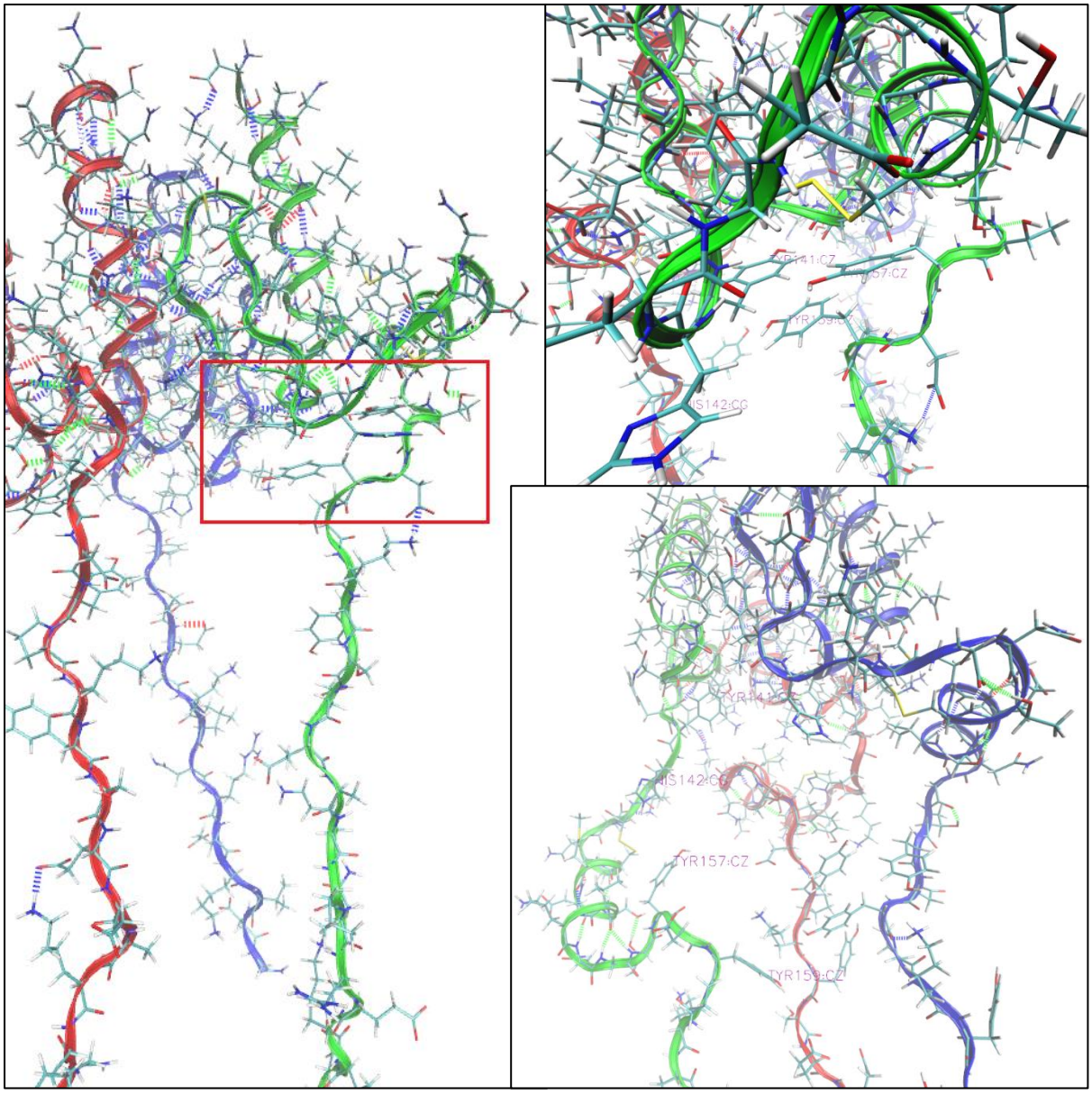


Fig. 13 - Run 3 at 80 ns. A) The red box is where we zoom in. B) A hydrophobic interaction between Tyr-157, Tyr-159 and Tyr-141 seems to be the only interaction preventing helix J dissociation. No H-bonds are present. C) Run 3 at 104 ns. Helix J dissociates as soon as the interaction between the different aromatic rings falls apart, further confirming the significant influence of the hydrophobic interaction on helix J stability.

Concluding, the correlation and regression coefficients point towards two residues: Tyr-159 and Tyr-157. Tyr-159 seems to be the more likely candidate when solely looking at the CCs and RCs. When looking at the average occupancy, Tyr-157 seems to be the more critical residue. The F-values also point towards Tyr-157 as being the more critical residue. Evidence is not conclusive, and additional simulations will have to be performed to conclude whether Tyr-159 or Tyr-157 is most critical.

3.4 - Constant-force pulling simulations at 80 pN on mutations

To conclusively determine whether Tyr-157 or Tyr-159 is the most critical residue, 20 new constant-force pulling simulations have been performed. In 10 of these simulations, Tyr-157 is mutated to Gly-157, while in the other 10 Tyr-159 is mutated to Gly-159. Glycine is the smallest amino acid and has no side-chain (see appendix A). Therefore, a mutation using glycine significantly diminishes the ability to form H-bonds or salt bridges, as well as any hydrophobic interaction. All 20 simulations were performed using identical starting conditions (except for the mutation).

The average macroscopic unfolding time of the 10 runs with a mutation from Tyr-159 to Gly-159 was 64 ± 10 ns. Compared to the average unfolding time of the wildtype (250 ± 86 ns), this is a factor of 3.9 faster. The average macroscopic unfolding time of the 10 runs with a mutation from Tyr-157 to Gly-157 was 146 ± 47 ns. Compared to the average unfolding time of the wildtype, this is a factor of 1.7 faster.

Since the Tyr-159 mutations are a factor 3.9 faster whereas the Tyr-157 mutations are only a factor of 1.7 faster, Tyr-159 clearly is the most critical residue in H1. This does not mean that Tyr-157 is unimportant. With the Tyr-157 mutation, unfolding times were still a factor of 1.7 faster.

It is interesting to look at the dissociation times of helices J and K during the different mutation simulations. These are shown in figure 14.

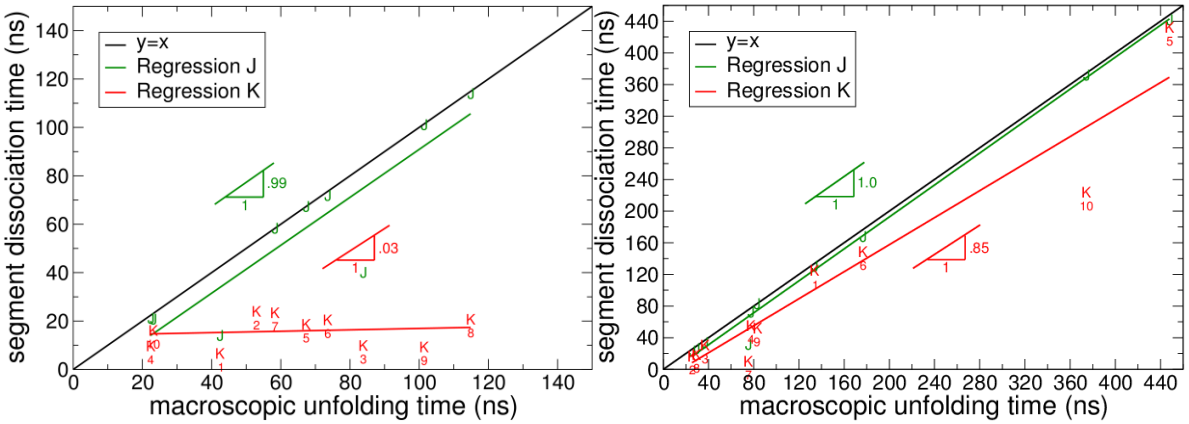


Fig. 14 - Helix J and K dissociation versus macroscopic unfolding. A) Tyr-159 mutation. B) Tyr-157 mutation.

Surprisingly, the regression of helix K is lower in the Tyr-159 mutation (0.03) than in the wildtype (0.50), while the regression of helix K in the Tyr-157 mutation (0.85) is higher than in the wildtype.

The most likely explanation for this behavior is as follows: due to its unique position, Tyr-159 is able to stabilize both helix K and helix J (figure 15). This is partially done by making H-bonds with residues such as Asn-128, but mostly by the hydrophobic interaction Tyr-159 has with His-142 and Tyr-141. Tyr-157 on the other hand resides further upwards in the chain and is therefore only able to provide stability to helix J. This is done partially by making H-bond with residues such as Asn-129 and His-142, but again mostly by the hydrophobic interaction Tyr-157 has with His-142 and Tyr-141.

When Tyr-159 is mutated to Gly-159, helix K is no longer stabilized by Tyr-159 and will therefore dissociate much faster, while Tyr-157 is still there to provide stability to helix J, explaining the small regression of helix K in figure 14a. When Tyr-157 is mutated to Gly-157, helix K remains just as stable as in the wildtype because helix K could not profit from Tyr-157 anyway, while helix J becomes more instable because it can no longer profit from the stability provided by Tyr-157 earlier, explaining the high helix K regression in figure 14b.

If the mechanism assumed above is correct, this would not only explain why Tyr-159 is much more critical than Tyr-157, but also why a mutation of Tyr-157 still has very significant effect on the macroscopic unfolding time.

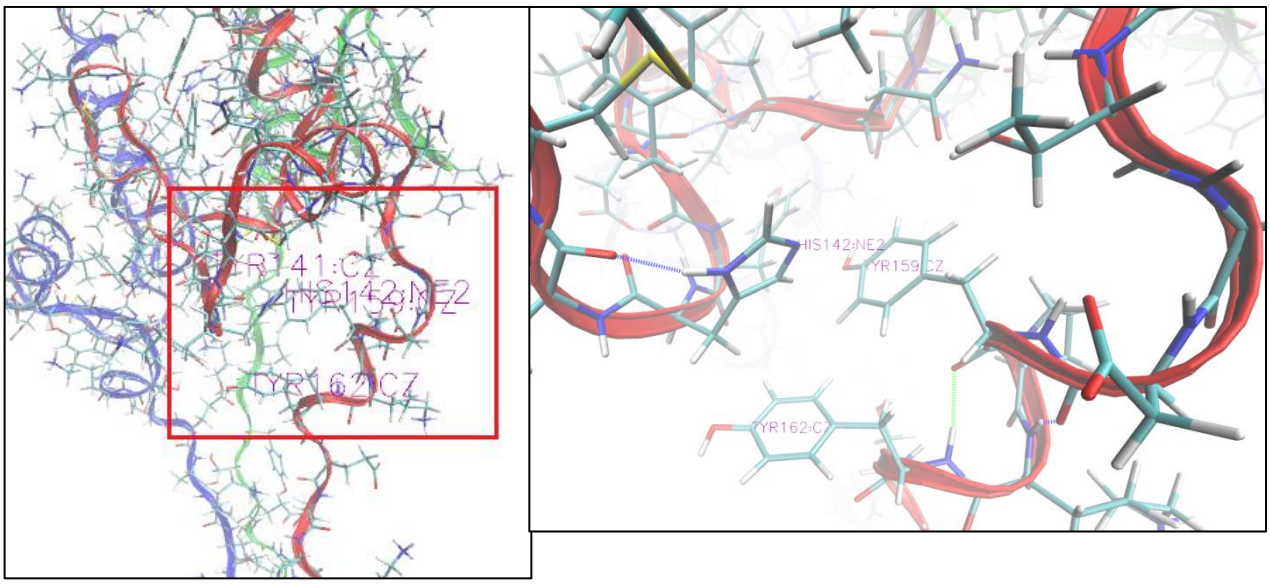


Fig. 15 - Tyr-157 mutation, run 4 at 65 ns. A) The red box is where we zoom in. B) Thanks to the unique position of Tyr-159, it is able to stabilize both helix J and K via the hydrophobic interaction between Tyr-159 and His-142 (and/or Tyr-141).

Chapter 4 - Results H5-type HA2

4.1 - Initial simulation

The initial simulation ran for a total of 60 ns. The average distance between the Asp-112 residues in the different chains during the first 13 ns was determined to be 1.44 nm. After 13 ns, the protein started to deform due to the destabilizing effect the removal of the first 90 residues had on the top and the fact that nothing is constrained. Starting configurations were extracted during the first 13 ns at points where the distance between the Asp-112's of the three chains lay between 1.39 and 1.49 nm. A minimum separation time of 0.5 ns was used between the different starting configurations in order for the configurations to differ significantly.

Again, three tables are given, displaying the top 10 most occupied H-bonds for each of the chains. H-bonding partners consistently showing up have been colored. H-bonds involving residues 171 to 175 are not considered, and neither are residues that only show up in one chain.

Table 6 - Top 10 most occupied H-bonds, chain A

H-bond partners	Occupancy	H-bond partners	Occupancy
A-Tyr-159-A-Asp-128	1.00	A-Arg-170-C-Glu-171	0.36
A-Lys-143-A-Glu-165	0.99	A-Arg-170-A-Asp-128	0.35
A-Asn-129-A-Tyr-157	0.95	A-Tyr-159-A-Asn-129	0.27
A-Arg-167-A-Asp-128	0.71	A-Arg-170-C-Asp-128	0.13
A-Tyr-157-A-His-142	0.60	B-Ser-175-A-Glu-171	0.10

Table 7 - Top 10 most occupied H-bonds, chain B

H-bond partners	Occupancy	H-bond partners	Occupancy
B-Tyr-159-B-Asp-128	0.98	B-Ser-175-C-Ser-174	0.28
B-Asn-129-B-Tyr-157	0.94	B-Tyr-157-B-Gln-125	0.26
B-Arg-170-B-Asp-128	0.66	C-Ser-175-B-Glu-171	0.24
B-Tyr-157-B-His-142	0.61	B-Ser-174-C-Ser-175	0.22
B-Lys-143-B-Glu-165	0.52	B-Arg-153-B-Gly-136	0.18

Table 8 - Top 10 most occupied H-bonds, chain C

H-bond partners	Occupancy	H-bond partners	Occupancy
C-Lys-143-C-Glu-165	1.00	C-Tyr-157-C-His-142	0.40
C-Tyr-159-C-Asp-128	0.98	C-Asn-129-C-Tyr-159	0.37
C-Asn-129-C-Tyr-157	0.93	A-Arg-170-C-Glu-171	0.36
C-Arg-170-C-Asp-128	0.83	C-Ser-174-B-Ser-175	0.28
C-Gln-125-C-Gly-155	0.51	C-Ser-175-B-Glu-171	0.24

We see five H-bonds consistently showing up. The associated residues of interest are Arg-170, Glu-165, Tyr-159 and Tyr-157. We can consider all four of these residues suspect for being critical. Arg-170 and Glu-165 are part of helix K (residues 163 to 171) and have therefore probably more to do with helix K stability, while Tyr-157 and Tyr-159 have probably more to do with helix J stability since they lie between helix J (residues 146 to 154) and helix K.

4.2 - Constant-velocity simulations

Three constant-velocity pulling simulations have been performed. All three simulations were performed at a constant pulling rate of 1 Å/ns. Again, the wrong terminals were used. In this case however, no force peak and no subsequent drop are present in the first part of the simulation. In my interpretation, it is therefore justified to determine the average pulling force over the entire simulation.

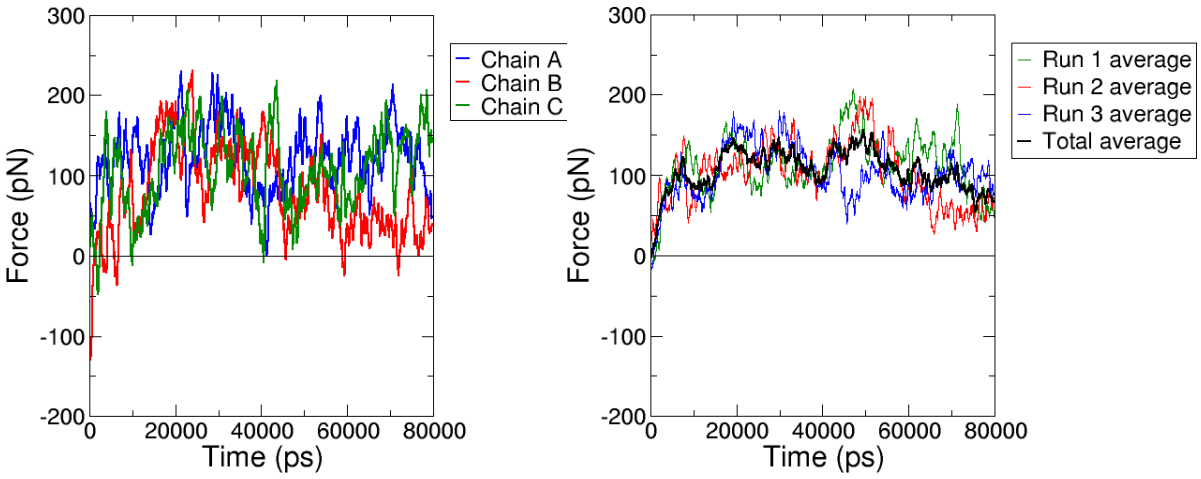


Fig. 16 - Constant-velocity pulling at 1 Å/ns, run 3 (1 and 2 look similar). A) Pulling force on the individual chains. B) Average force of runs 1, 2 and 3, as well as the total -average pulling force. No force peak and no subsequent drop are visible in the (total) average force.

Table 9 - Time-average force of constant-velocity pulling simulations at 1 Å/ns

Run	Chain	Time-average force per chain (pN)	Time-average force (pN)
1	A	91	112
	B	114	
	C	131	
2	A	83	102
	B	127	
	C	96	
3	A	122	103
	B	79	
	C	107	
Average force (all runs)			**106**

Concluding, during constant-velocity simulations at a rate of 1 Å/ns, a time-average pulling force of 106 pN is observed. During run 1 helix G got pulled apart, but this did not happen during run 2 or run 3. This means that 106 pN is around the maximum force that can be used for constant-force pulling.

To get a better idea of what force to use, once more a comparison can be made with H3. Since an average force of 106 pN was also observed in H3 during 1 Å/ns pulling, we can also use 100 pN for the constant-force pulling simulations on H5.

4.3 - Constant-force pulling simulations at 100 pN

A total of 10 constant-force pulling simulations have been performed using a force of 100 pN. All runs reached macroscopic unfolding. The macroscopic unfolding time is again defined as the moment at which helices J and K of the fastest chain have both dissociated, but this time a distance of 11 nm between the pulling groups (Arg-112 and Ser-175) is used. A distance of 11 nm is used because this was the maximum the periodic boundary of the simulation box allowed for in H5. Macroscopic unfolding times were on average 100 ± 16 ns.

Figure 17 shows the correlation between macroscopic unfolding and both helix dissociation and helix unfolding of helices J and K.

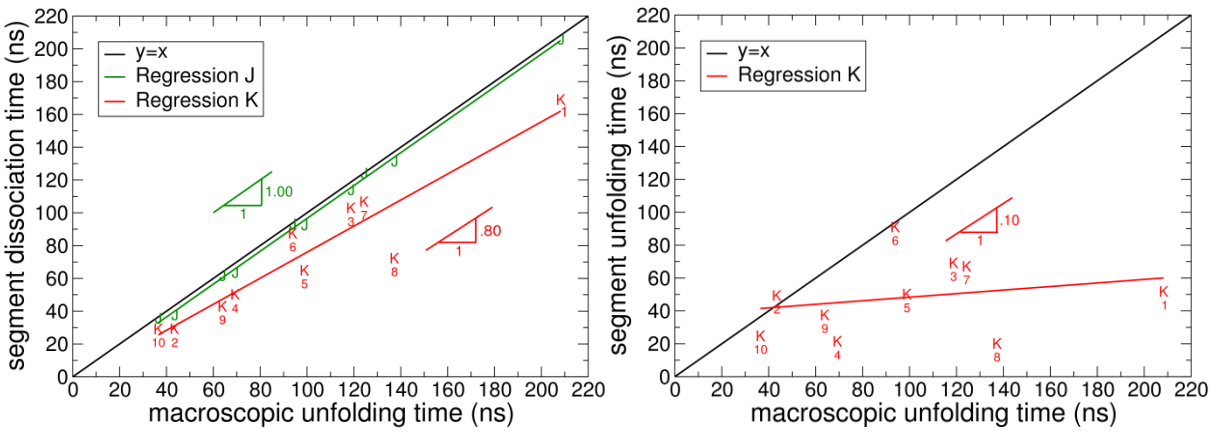


Fig. 17 - Helix J and K dissociation and unfolding. A) Helix dissociation. B) Helix unfolding.

Helix J dissociation is highly correlated with macroscopic unfolding with a CC of 1.0 and a RC of 1.0. The dissociation of helix K is somewhat correlated to macroscopic unfolding with a CC of 0.94 and a RC of 0.80. The RCs of helices J and K differ considerably, indicating no real fixed delay between helix J and helix K dissociation exists. On average, helix K dissociates at about 80% of the time of helix J dissociation. Helix K always dissociates before helix J.

Helix J never unfolds before the macroscopic unfolding event and is therefore not shown in figure 17b. Helix K unfolding has a CC of 0.24 and a RC of 0.11 and is therefore not correlated to macroscopic unfolding.

Moving on to residue dissociation, figure 18 shows the CCs and RCs of residues 146 to 175, as well as their average occupancy and F-values.

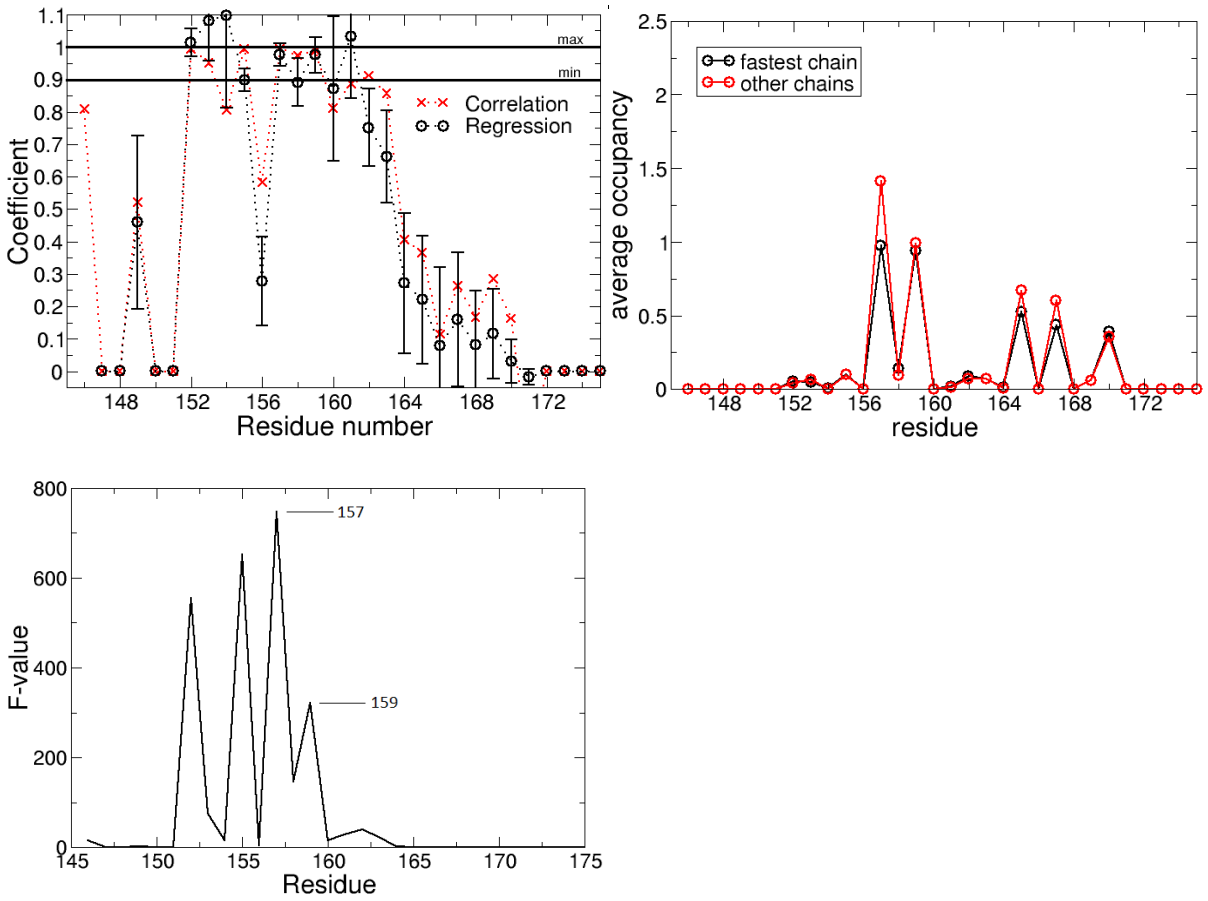


Fig. 18 - A) Correlation and regression coefficients for the LVC-regression of residues 146 to 175. B) Average occupancy up until the LVC. Both the fastest chain and the other chains are shown. C) F-values for residues 146 to 175.

The residues with both their CC and RC convincingly higher than 0.9 are Tyr-159 (CC = 0.99, RC = 0.97) and Tyr-157 (CC = 0.99, RC = 0.98). We set the lower limit at 0.9 (we used 0.8 in H1) because the regression coefficients in H5 are slightly larger overall as compared to H1. In addition to Tyr-159 and Tyr-159, residues 152 and 153 also have RCs and CCs above 0.9*, but the reasoning used in H1 also applies here: this is to be expected in either Tyr-157 or Tyr-159 is the most critical residue. And again, if Tyr-159 is the most critical residue, it would not be strange if Tyr-157 has almost the same CC and RC values as Tyr-159. Therefore, Tyr-159 seems to be more likely to be the critical residue when looking solely at the CCs and RCs.

Looking at the F-values (figure 18c), the four highest ones are Tyr-157 (748), Gly-155 (652), Val-152 (556) and Tyr-159 (321). The F-value of Tyr-157 is 2.3 times as large as the F-value of Tyr-159, therefore the F-values point towards Tyr-157 as being the most critical residue.

*The RCs with a value larger than 1 can be easily explained. They are caused by a particular placement of the data points through which the regression is drawn. See appendix B for an example.

When looking at the average occupancy (figure 18b) we see five peaks. From right to left, the first peak we seek is associated with residue Arg-170, the second peak is associated with Arg-167 and the third peak is associated with Glu-165. Since all three of these residues are part of helix K, these residues probably play a role in helix K stability. Since helix K dissociation is not the most critical event in H5, these residues are not too interesting. More interesting are the highly occupied residues Tyr-157 and Tyr-159. These two residues probably play a role in helix J stability. The occupancy of Tyr-159 is the same in both the fastest and the other two chains, but the occupancy of Tyr-157 clearly shows a reduction when it belongs to the fastest chain. This indicates that a lower occupancy of the H-bonds of Tyr-157 leads to earlier chain unfolding. Just as in H1, this gives an indication that Tyr-157 might be more critical than Tyr-159.

Looking in the protein structure during the simulations before helix K dissociation (figure 19), we see that Glu-165, Arg-167 and Arg-170 indeed plays a role in helix K stability (Arg-167 and Arg-170 are not shown in figure 19b), just as we expected from the average occupancy results. We also see a hydrophobic interaction between the aromatic rings of Tyr-157, Tyr-159, Tyr-162, His-142 and Tyr-141. This hydrophobic interaction may contribute significantly to the stability of helix K.

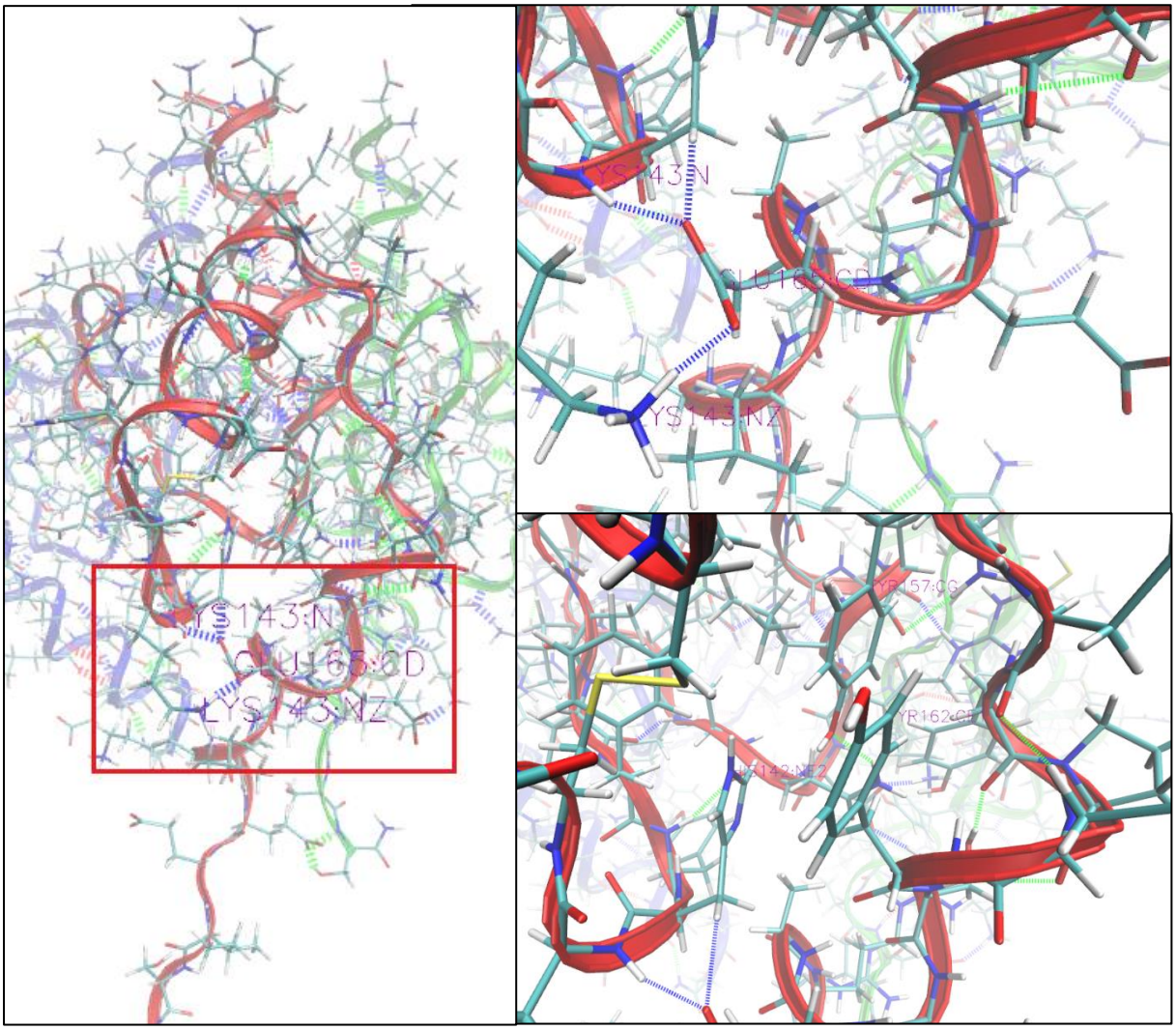


Fig. 19 - Run 4 at 8 ns. A) The red box is where we zoom in. B) Hydrogen bonds between Glu-165 and Lysr-143 provide stability to helix K against the vertical pulling. C) The aromatic side-chains of Tyr-157, Tyr-162 and His-142 cluster together due to their hydrophobic interaction.

Looking in the protein structure after helix K has dissociated but before macroscopic unfolding (figure 20), we see a lot of hydrophobic interaction between Tyr-159, Tyr-157, His-142 and Tyr-141. Again, no H-bonds seem to significantly contribute to helix J stability.

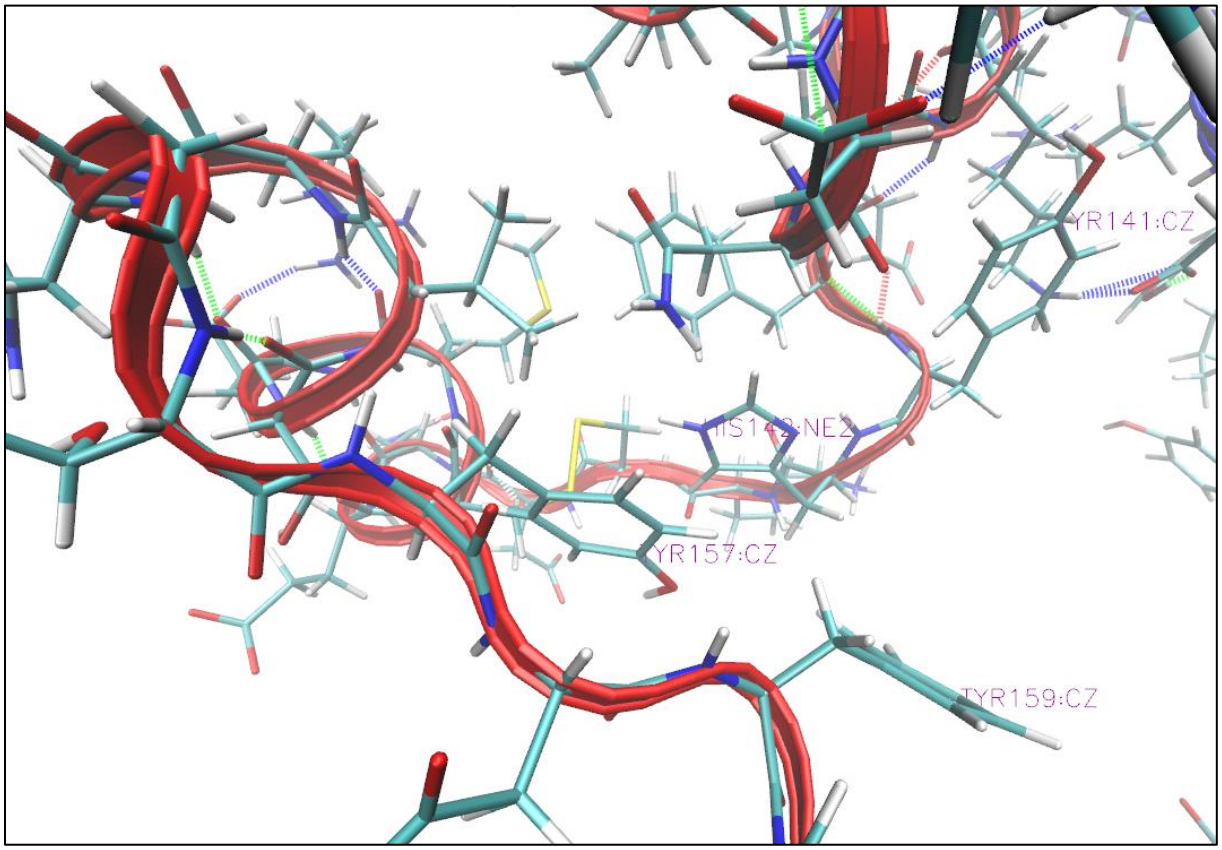


Fig. 20 - Run 4 at 60 ns. A) The hydrophobic interaction between the aromatic rings of Tyr-159, Tyr-157, His-142 and Tyr-141 seems to be the only interaction preventing helix J dissociation. Although Tyr-141 and Tyr-159 are a little further apart in this particular frame, they definitely contribute to helix J stability.

Concluding, the correlation and regression again point toward Tyr-159 and Tyr-157. Tyr-159 seems to be the more likely candidate when solely looking at the CCs and RCs. When looking at the average occupancy, Tyr-157 seems to be the most critical residue. The F-values also point towards Tyr-157 as being the most critical residue. Just as in H1, evidence is not conclusive and additional simulations will have to be performed to conclude whether Tyr-159 or Tyr-157 is the most critical residue.

4.4 - Constant-force pulling simulations at 100 pN on mutations

To conclusively determine whether Tyr-157 or Tyr-159 is the most critical residue, 20 new constant-force pulling simulations have been performed. In 10 of these simulations, Tyr-157 is mutated to Gly-157, while in the other 10 Tyr-159 is mutated to Gly-159. All 20 simulations were performed using identical starting conditions (except for the mutation).

The average macroscopic unfolding time of the 10 runs with a mutation from Tyr-159 to Gly-159 was 32 ± 12 ns. Compared to the average unfolding time of the wildtype (100 ± 16 ns), this is a factor of 3.1 faster. The average macroscopic unfolding time of the 10 runs with a mutation from Tyr-157 to Gly-157 was 67 ± 19 ns. Compared to the average unfolding time of the wildtype, this is a factor of 1.5 faster.

Since the Tyr-159 mutations are a factor 3.1 faster whereas the Tyr-157 mutations are only a factor of 1.5 faster, Tyr-159 clearly is the most critical residue in H5. This does not mean that Tyr-157 is unimportant. With the Tyr-157 mutation, unfolding times were still a factor of 1.5 faster.

Just as in H1, also here it can be interesting to look at the dissociation times of helices J and K during the mutation simulations. These are shown in figure 21.

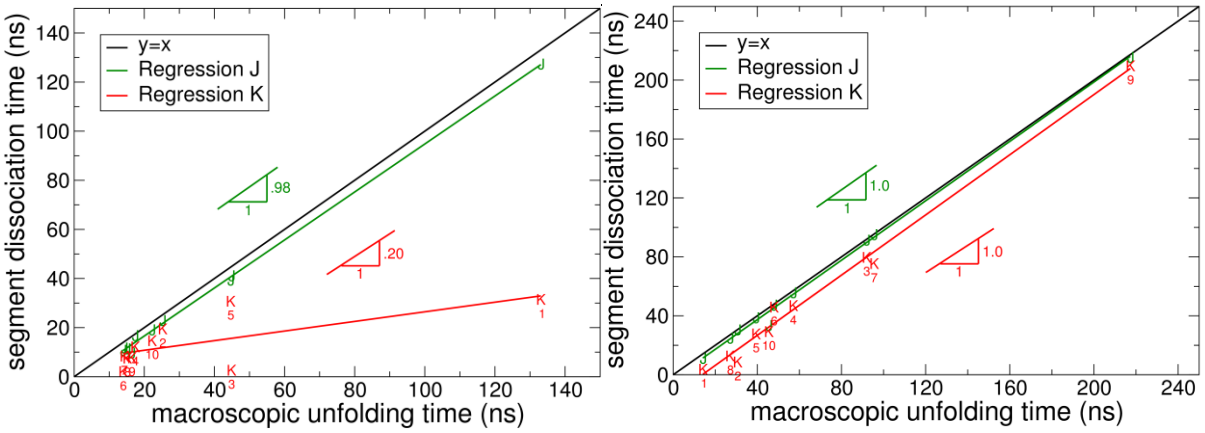


Fig. 21 - Helix J and K dissociation versus macroscopic unfolding. A) Tyr-159 mutation. B) Tyr-157 mutation.

Also here we see that the regression of helix K is lower in the Tyr-159 mutation (0.20) than in the wildtype (0.80), while the regression of helix K in the Tyr-157 mutation (1.0) is higher than in the wildtype.

The behavior of helix K in figure 21 looks a lot like the behavior of helix K during H1 mutations (figure 14), and probably has a similar explanation. Also in H5 Tyr-159 resides on a unique position, being able to stabilize both helix J and helix K. Tyr-157 on the other hand is only able to stabilize helix J. This explains the small RC of helix K in the Tyr-159 mutation simulations and the large RC of helix K in the Tyr-157 mutation simulations.

This would again explain why Tyr-159 is much more critical than Tyr-157, but also why a mutation of Tyr-157 still has a significant effect on the macroscopic unfolding time.

Chapter 5 - Discussion

In chapters 3 and 4, the results of all simulation on H1 and H5-type HA2 are shown, respectively. A lot of small results and observations have already been interpreted or discussed in these chapters. A more general discussion of the most important observation and results will now be presented.

Molecular dynamics simulations haven been performed on H1 and H5-type HA2 to find interactions critical to the stability of HA2. Constant-force simulations on the wildtype of H1 and H5-type HA2 showed that for both types helix J dissociation is a critical event in globular bottom unfolding. In both types, helix K dissociation does play a role, but is not critical to globular bottom stability.

Furthermore, the results of the wildtype mutations suggested that in both H1 and H5 type HA2, Tyr-159 and Tyr-157 are the most critical residues. Regression and correlation coefficients were in favor of Tyr-159, while the F-values and average occupancy were in favor of Tyr-157. Additional simulations performed on mutations of Tyr-159 to Gly-159 showed a reduction of factor of 3.9 in macroscopic unfolding times in H1 and a reduction of a factor of 3.1 in H5. Simulations performed on mutations of Tyr-157 to Gly-157 only showed a reduction of a factor of 1.7 in macroscopic unfolding times in H1 and reduction of a factor of 1.5 in H5.

The large difference in reduction of macroscopic unfolding times between Tyr-159 and Tyr-157 mutations makes Tyr-159 the most critical residue in both H1 and H5 type HA2, irrespective of the ambiguity of the wildtype simulation results or the underlying interactions of Tyr-159. When looking at the proposed unproductive pathway of HA (chapter 1.4), the only thing that really determines if a residue is critical is its influence on the macroscopic unfolding time.

In the previous chapters, we already briefly discussed why Tyr-159 might be the most critical residue. In both H1 and H5, this is most likely due to its unique location in the peptide chain. Due to its unique location, Tyr-159 is able to provide stability to both helix K and helix J. This is partially done by making H-bonds with residues such as Asn-128 and His-142, but probably mostly by the hydrophobic interaction the aromatic side chain of Tyr-159 has with Tyr-162, Tyr-157, His-142 and Tyr-141 (see for example figure 13b).

The fact that the role of the hydrophobic interaction is not quantified in the simulation results makes it difficult to tell how large its influence really is, and probably explains the ambiguity in the results of the wildtype simulation. In reality, Tyr-159 and Tyr-157 actually work together in preventing helix J dissociation via their hydrophobic interaction with Tyr-141 and His-142. The fact that Tyr-159 is also able to provide stability to helix K (also probably mostly via hydrophobic interactions) is what explains the difference between the unfolding times during the mutations, and why Tyr-159 is more critical than Tyr-157.

All residues involved in the hydrophobic interaction preventing helix J and K dissociation (Tyr-162, Tyr-159, Tyr-157, His-142 and Tyr-141) are conserved in both H1, H5 and H3 type HA2 (see appendix C) [13]. The only non-conserved residue is 159. Tyr-159 becomes His-159 in H3, but this has no influence on the hydrophobicity (they both have an aromatic ring). The fact that all of these residues are conserved explains why H1 and H5 type HA2 look similar during the simulations, and why in both types Tyr-159 is the most critical residue. This does however not explain why H5-type HA2 seems to be more stable than H1-type HA2.

The reason H5-type HA2 is more stable than H1-type HA2 is most likely due to the mutation of the neutral residue Asn-128 (H1) to the negatively charged residue Asp-128 (H5) [13]. The positively charged and conserved residue Arg-170 is now able to make salt bridges with Arg-128, which was not possible in H1 with Asn-128 (see appendix D). This makes helix K more stable and therefore makes H5 as a whole more stable. This would also explain why the regression coefficient of helix K dissociation in H5 is larger (0.80) than the regression coefficient of helix K dissociation in H1 (0.50).

Comparing H1 and H5 to H3-type HA2, a large difference in average macroscopic unfolding time is observed (figure 22). This difference can be explained as follows: the most critical interaction in H3 is Arg-163 [9]. Arg-163 is positively charged and forms a salt bridge with the negatively charged Glu-128. The interaction between these residues is very strong and prevents helix K dissociation. Once the interaction is broken, helix K quickly dissociates. Since all hydrophobic residues playing a role in helix J and K stability in H1 and H5 are conserved, they are also present in H3 and they probably also help stabilize helix J and K in H3. This is supported by looking at the macroscopic unfolding time of H3-type HA2 simulations where Arg-163 is mutated to Ala-163 (figure 22). Macroscopic unfolding times are now on average only 76 ns instead of 269 [9]. 76 ns lies close to the macroscopic unfolding time of the H5 simulation on the wildtype (in both simulations a force of 100 pN was used). This all suggest that the hydrophobic interaction that is providing stability to helix J and helix K in H1 and H5 is also present in H3, but in H3 an extra 'layer' of stability is added on top of this by the salt bridge between Arg-163 and Glu-128.

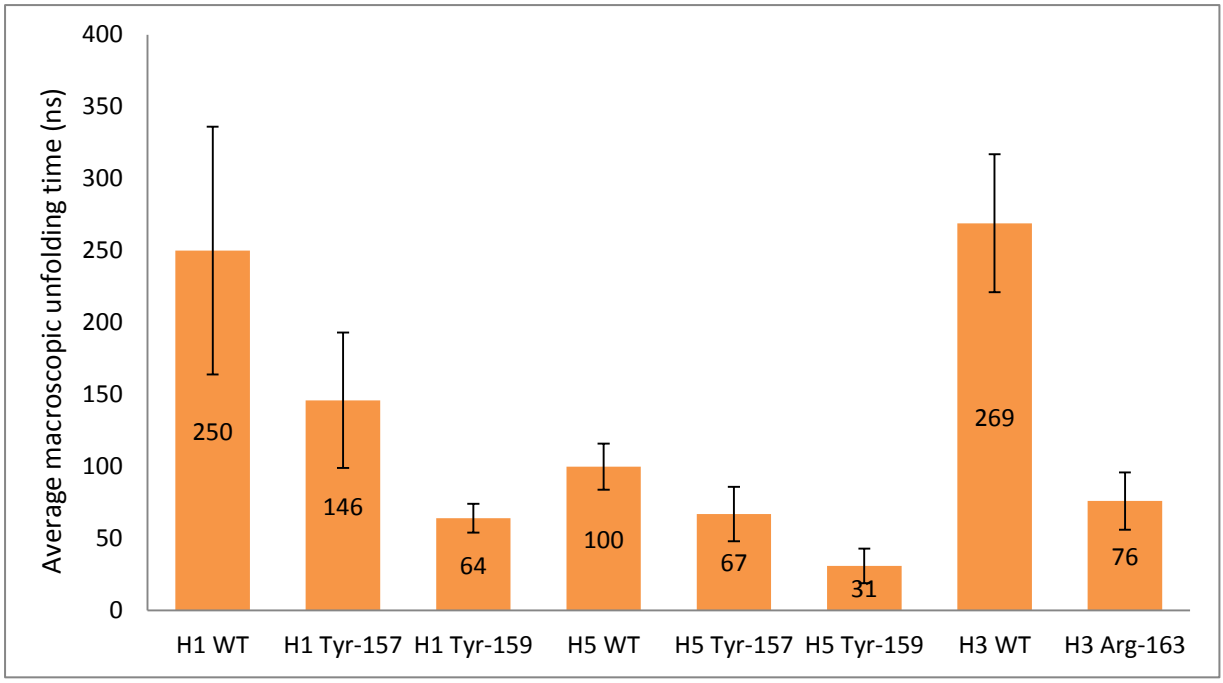


Fig. 22 - Average macroscopic unfolding times of various HA2 subtypes and mutations.

Chapter 6 - Conclusions

Interactions and mechanisms critical to HA2 stability have been identified in H1 and H5 type HA2. Helix J dissociation is a critical event in both H1 and H5 type HA2, while helix K dissociation is a critical event in H3 type HA2. The most critical residue in both H1 and H5 type HA2 is Tyr-159. The most critical residue in H3 type HA2 is Arg-163 [9]. Helix J dissociation as a critical event is conserved in H1 and H5 type HA2. The most critical residue, Tyr-159, is also conserved in H1 and H5 type HA2. Although the most critical residue of H3, Arg-163, is not conserved in H1 and H5, the hydrophobic interaction largely responsible for helix J an helix K stability in H1 and H5 is also present in H3 and therefore conserved in all three subtypes.

Chapter 7 - Recommendations

The most important recommendation for further research is quantifying the role of hydrophobic interactions in globular bottom stability. Other recommendations are investigating the pH dependence of the stability of H1 and H5 type HA2 by protonating histidine or glutamic acid residues, and reperforming the constant-force pulling simulations on the wildtype of H1 at a force of 100 pN for a better comparability with H3 an H5.

Chapter 8 - References

[1] World Health Organization 2016, *Influenza (seasonal)*, World Health Organization, viewed 23 April 2016, <http://www.who.int/mediacentre/factsheets/fs211/en/>

[2] Tong, SX, Zhu, XY, Li, Y, Shi, M, Zhang, J, Bourgeois, M, Yang, H, Chen, XF, Recuenco, S, Gomez, J, Chen, LM, Johnson, A, Tao, Y, Dreyfus, C, Yu, WL, McBride, R, Carney, PJ, Gilbert, AT, Chang, J, Guo, Z, Davis, CT, Paulson, JC, Stevens, J, Rupprecht, CE, Holmes, EC, Wilson, IA & Donis, RO 2013, 'New World Bats Harbor Diverse Influenza A Viruses', *PLoS Pathogens*, vol. 9, no. 10, pp. 1-12.

[3] Center for Disease Control and Prevention 2014, *Types of Influenza Viruses*, Center for Disease Control and Prevention, viewed 1 July 2016, <http://www.cdc.gov/flu/about/viruses/types.htm>

[4] Harrison, SC 2015, 'Viral Membrane Fusion', *Virology*, vol. 479, pp. 498-507

[5] Huang, IC, Li, W, Sui, J, Maracso, W, Choe, H & Farzan, M 2008, 'Influenza A virus neuraminidase limits viral superinfection', *Journal of Virology*, vol. 82, no. 10, pp. 4834-4843.

[6] Shors, T 2009, *Understanding Viruses*, first edition, Jones and Bartlett Publishers.

[7] Center for Disease Control and Prevention 2009, *3D Influenza Transparent Key Pieslice*, viewed April 2016, <http://www.cdc.gov/h1n1flu/images/3D_Influenza_transparent_key_pieslice_sml.gif>

[8] Wikipedia 2008, *Protein Structure*, viewed April 2016, <https://en.wikipedia.org/wiki/Protein_structure>

[9] Boonstra S, 2016, private information.

[10] Boonstra S, Onck, PR & van der Giessen, E 2016, 'CHARMM TIP3P Water Model Suppresses Peptide Folding by Solvating the Unfolded State', *Physical Chemistry*, vol. 120, pp. 3692-3698.

[11] Protein Data Bank 2016, *3LZG*, downloaded April 2016, <http://www.rcsb.org/pdb/explore/explore.do?structureId=3LZG>

[12] Protein Data Bank 2016, *2FK0*, downloaded April 2016, <http://www.rcsb.org/pdb/explore/explore.do?structureId=2FK0>

[13] Nobusawa, E, Aoymama, T, Kato, H, Suzuki, Y, Tateno Y & Nakajima, K 1991, 'Comparison of Complete Amino Acid Sequence and Receptor-Binding Properties among 13 Serotypes of Hemagglutinins of Influenza A viruses', *Virology*, vol. 182, pp. 478-479.

[14] Wikipedia 2010, *Amino acids*, viewed April 2016, <https://en.wikipedia.org/wiki/Amino_acids>

Chapter 9 - Appendix

9.1 - Appendix A - The Amino acids

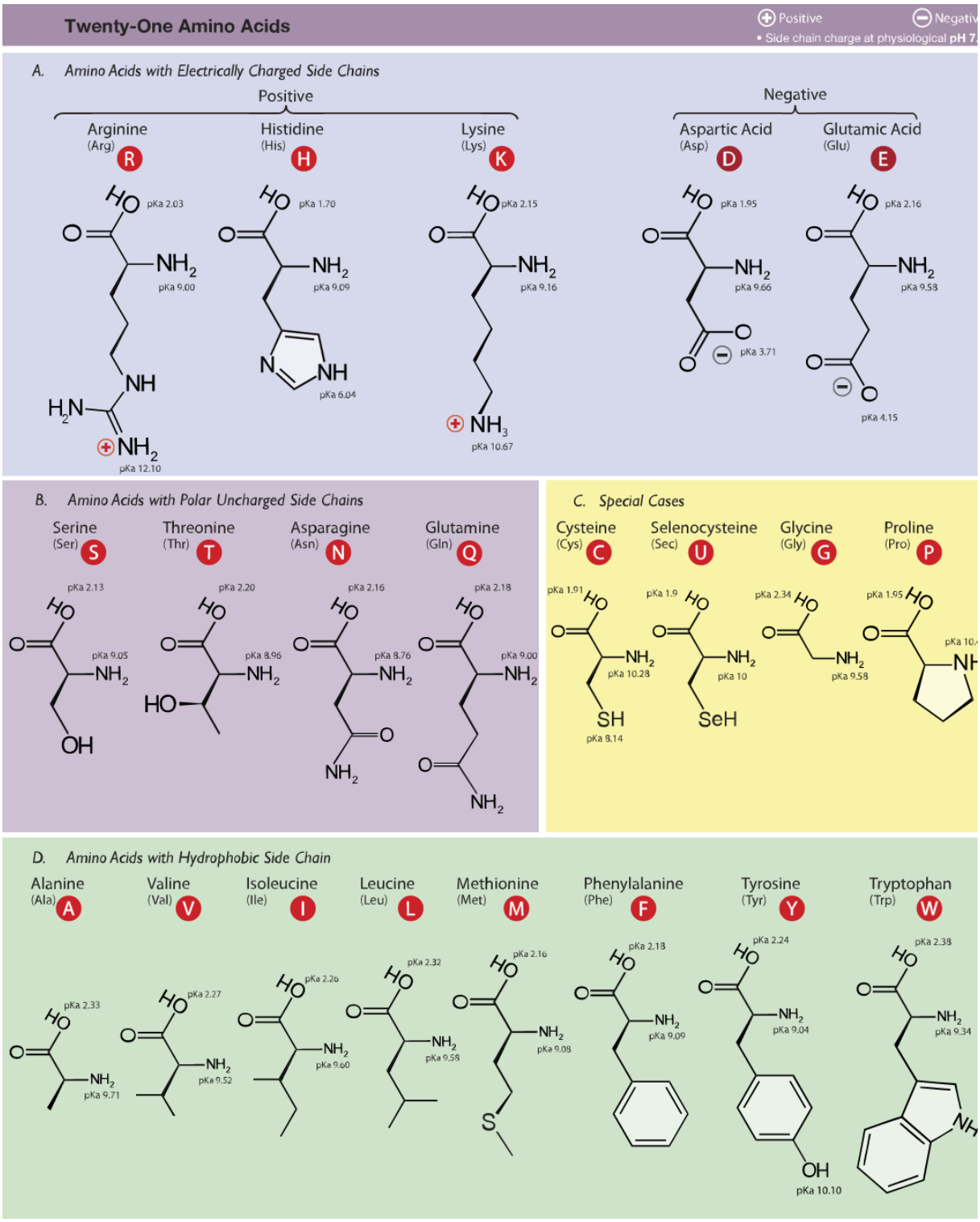


Fig. 23 -The 21 amino acids [14].

9.2 - Appendix B - LVC regression with RC > 1

Constant-force pulling at 100 pN on H5 wildtype. Explanation for an LVC regression higher than 1 (LVC regression of residue 153 is shown).

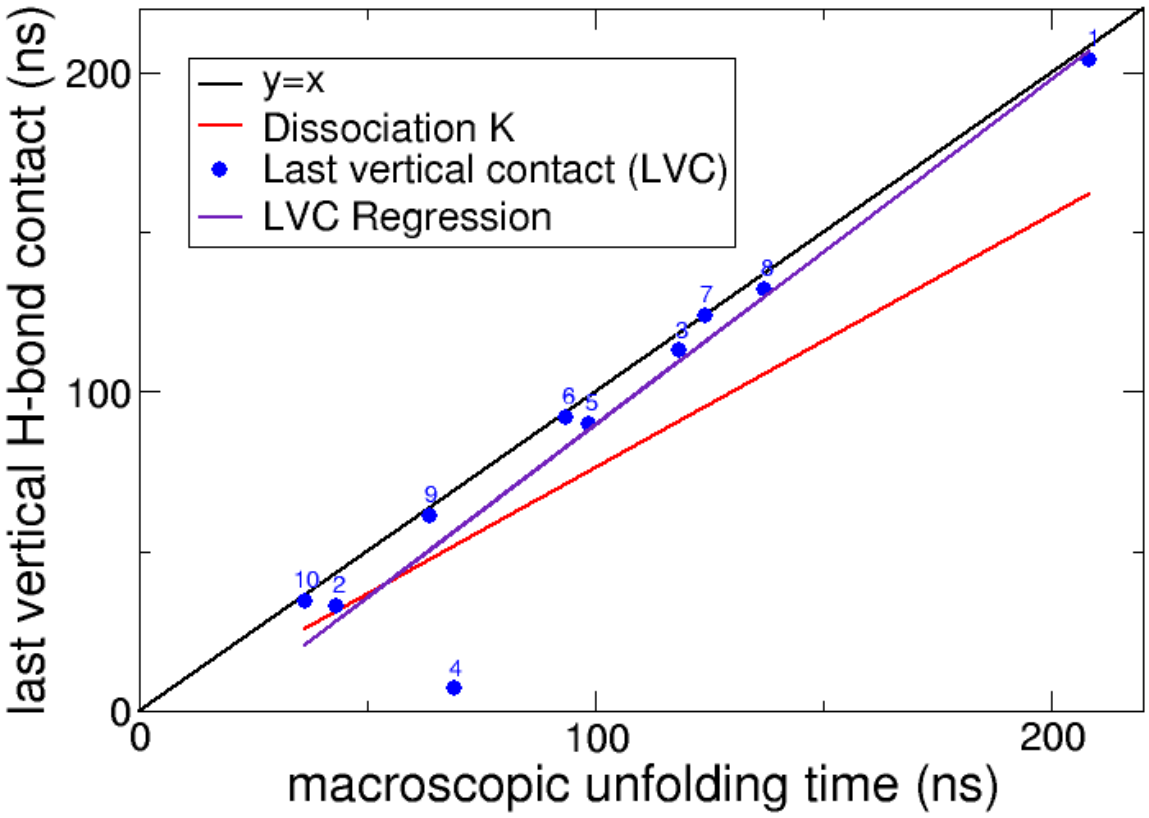


Fig. 24 - LVC regression of residue 153, constant-force pulling on H5 wildtype.

9.3 - Appendix C - Residue comparison, hydrophobic and polar

	H1	H5		H3			H1	H5		H3	
112	Asp	Asp	c	Asp	c	144	Cys	Cys	c	Cys	c
113	Ser	Ser	c	Ser	c	145	Asp	Asp	c	Asp	c
114	Asn	Asn	c	Glu		146	Asn	Asn	c	Asn	c
115	Val	Val	c	Met		147	Thr	Glu		Ala	
116	Lys	Lys	c	Asn		148	Cys	Cys	c	Cys	c
117	Asn	Asn	c	Lys		149	Met	Met	c	Ile	
118	Leu	Leu	c	Leu	c	150	Glu	Glu	c	Glu	c
119	Tyr	Tyr	c	Phe		151	Ser	Ser	c	Ser	c
120	Glu	Asp		Glu		152	Val	Val	c	Ile	
121	Lys	Lys	c	Lys	c	153	Lys	Arg		Arg	
122	Val	Val	c	Thr		154	Asn	Asn	c	Asn	c
123	Arg	Arg	c	Arg	c	155	Gly	Gly	c	Gly	c
124	Ser	Leu		Arg		156	Thr	Thr	c	Thr	c
125	Gln	Gln	c	Gln	c	157	Tyr	Tyr	c	Tyr	c
126	Leu	Leu	c	Leu	c	158	Asp	Asp	c	Asp	c
127	Lys	Arg		Arg		159	Tyr	Tyr	c	His	
128	Asn	Asp		Glu		160	Pro	Pro	c	Asp	
129	Asn	Asn	c	Asn	c	161	Lys	Gln		Val	
130	Ala	Ala	c	Ala	c	162	Tyr	Tyr	c	Tyr	c
131	Lys	Lys	c	Glu		163	Ser	Ser	c	Arg	
132	Glu	Glu	c	Glu	c	164	Glu	Glu	c	Asp	
133	Ile	Leu		Met		165	Glu	Glu	c	Glu	c
134	Gly	Gly	c	Gly	c	166	Ala	Ala	c	Ala	c
135	Asn	Asn	c	Asn	c	167	Lys	Arg		Leu	
136	Gly	Gly	c	Gly	c	168	Leu	Leu	c	Asn	
137	Cys	Cys	c	Cys	c	169	Asn	Lys		Asn	
138	Phe	Phe	c	Phe	c	170	Arg	Arg	c	Arg	
139	Glu	Glu	c	Lys		171	Glu	Glu	c	Phe	
140	Phe	Phe	c	Ile		172	Lys	Glu		Gln	
141	Tyr	Tyr	c	Tyr	c	173	Val	Ile		Ile	
142	His	His	c	His	c	174	Asp	Ser		Lys	
143	Lys	Lys	c	Lys	c	175	Gly	Ser		Gly	

Fig. 25 - Residue comparison of H1, H3 and H5 [13]. Hydrophobic residues are colored yellow, polar residues are colored blue.

9.4 - Appendix D - Residue comparison, positive and negative

	H1	H5		H3			H1	H5		H3	
112	Asp	Asp	c	Asp	c	144	Cys	Cys	c	Cys	c
113	Ser	Ser	c	Ser	c	145	Asp	Asp	c	Asp	c
114	Asn	Asn	c	Glu		146	Asn	Asn	c	Asn	c
115	Val	Val	c	Met		147	Thr	Glu		Ala	
116	Lys	Lys	c	Asn		148	Cys	Cys	c	Cys	c
117	Asn	Asn	c	Lys		149	Met	Met	c	Ile	
118	Leu	Leu	c	Leu	c	150	Glu	Glu	c	Glu	c
119	Tyr	Tyr	c	Phe		151	Ser	Ser	c	Ser	c
120	Glu	Asp		Glu		152	Val	Val	c	Ile	
121	Lys	Lys	c	Lys	c	153	Lys	Arg		Arg	
122	Val	Val	c	Thr		154	Asn	Asn	c	Asn	c
123	Arg	Arg	c	Arg	c	155	Gly	Gly	c	Gly	c
124	Ser	Leu		Arg		156	Thr	Thr	c	Thr	c
125	Gln	Gln	c	Gln	c	157	Tyr	Tyr	c	Tyr	c
126	Leu	Leu	c	Leu	c	158	Asp	Asp	c	Asp	c
127	Lys	Arg		Arg		159	Tyr	Tyr	c	His	
128	Asn	Asp		Glu		160	Pro	Pro	c	Asp	
129	Asn	Asn	c	Asn	c	161	Lys	Gln		Val	
130	Ala	Ala	c	Ala	c	162	Tyr	Tyr	c	Tyr	c
131	Lys	Lys	c	Glu		163	Ser	Ser	c	Arg	
132	Glu	Glu	c	Glu	c	164	Glu	Glu	c	Asp	
133	Ile	Leu		Met		165	Glu	Glu	c	Glu	c
134	Gly	Gly	c	Gly	c	166	Ala	Ala	c	Ala	c
135	Asn	Asn	c	Asn	c	167	Lys	Arg		Leu	
136	Gly	Gly	c	Gly	c	168	Leu	Leu	c	Asn	
137	Cys	Cys	c	Cys	c	169	Asn	Lys		Asn	
138	Phe	Phe	c	Phe	c	170	Arg	Arg	c	Arg	
139	Glu	Glu	c	Lys		171	Glu	Glu	c	Phe	
140	Phe	Phe	c	Ile		172	Lys	Glu		Gln	
141	Tyr	Tyr	c	Tyr	c	173	Val	Ile		Ile	
142	His	His	c	His	c	174	Asp	Ser		Lys	
143	Lys	Lys	c	Lys	c	175	Gly	Ser		Gly	

Fig. 26 - Residue comparison of H1, H3 and H5 [13]. Positively charged residues are colored green, negatively charged residues are colored red.