Computational studies of influenza hemagglutinin
Boonstra, Sander
Introduction
The influenza virus infects about 5 to 15% of the world population every year, causing a disease that is well-known as ‘the flu’. These infections result in an estimated three to five million cases of severe illness, causing on average 250 to 500 thousand mortalities annually.\textsuperscript{24} Influenza has a significant economic impact in terms of work and school absence, loss of productivity, the cost of research on new treatments and the development and application of seasonal vaccinations.\textsuperscript{19} Yearly updates of broadly neutralizing influenza vaccines can help protect the most vulnerable people from the virus, but these drugs do not offer complete protection against infection.\textsuperscript{16} Moreover, spontaneous genetic mutations can produce new virus strains that are immune to current vaccines, which can potentially cause a world-wide pandemic.\textsuperscript{10} The 1918 ‘Spanish flu’ has been the largest documented influenza pandemic in history, infecting an estimated 500 million people and killing at least 50 million of those across the world.\textsuperscript{20} These observations illustrate the significance of ongoing fundamental research into the mechanism of influenza virus infection. The severe threat of a new pandemic remains, unless a common mechanism in the replication cycle of influenza is found that can be suppressed for all influenza virus strains.\textsuperscript{11}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Schematic view of an influenza virus particle, with its main constituents: Hemagglutinin (blue); Neuraminidase (red); the M2 ion channel (purple); the viral RNP, which contains the RNA (green); the lipid bilayer (light brown) and the M1 protein coat (maroon).\textsuperscript{2}}
\end{figure}
1.1 Influenza viral entry and replication

Virus particles can be regarded as tiny devices that take over a host cell and use it to generate their offspring, thereby ensuring the survival of their own species.\(^\text{17}\) The basic design of such a device is illustrated for the influenza virus in Figure 1.1. Influenza virus particles can be either spherical or filamentous, and have an average size of 80 to 100 nm. At the core of the particle lies the blueprint of the design – the viral genome (ribonucleoprotein; RNP) that encodes the sequences of the viral proteins. The purpose of the particle is to deliver the viral genome to the nucleus of a target cell. The genome is protected from the environment by a matrix protein M1 layer, which, in turn, is surrounded by an enveloping lipid bilayer. The three proteins that are embedded within this membrane are the M2 ion channel and the glycoproteins neuraminidase (NA) and hemagglutinin (HA).\(^\text{13,14}\) Influenza virus strains are named according to the antigenic character of their HA and NA proteins, currently ranging from H1 to H18 and N1 to N11, so a strain with HA subtype 3 and NA subtype 2 is called H3N2. All virus proteins have a role in the replication cycle of the virus, as discussed next.

The invasion strategy of the influenza virus is illustrated in Figure 1.2. Upon enter-
ing the respiratory tract of the host, receptor binding domains on the globular head of HA adhere the particle to sialic-acid moieties on the surface of epithelial cells. The particle is then internalized into an endosome by either clathrin-mediated endocytosis or macropinocytosis. The host cell traffics the endosome towards its nucleus, thereby gradually increasing the acidity within this compartment. Between the maturing and late endosome, a pH below 6 triggers HA to carry out its next task: membrane fusion. HA merges the viral and endosomal membrane, thereby exposing the viral genome to the cell cytoplasm. Concurrent acidification of the particle interior by the M2 ion channel causes uncoating of the genome through dissociation of the M1 matrix protein. The viral RNP enters the host cell nucleus through the nuclear pore complex and starts the synthesis of viral proteins and ribonucleic acids (RNA). This ultimately leads to new virus particles budding from the cell membrane and their NA-mediated release.

### 1.2 Entry inhibition

A number of steps in the replication cycle of the influenza virus can be used as a target for antiviral drugs. Entry inhibition strategies interfere with, for example, binding of HA to cellular receptors, cellular processes that mediate endocytosis, M2-mediated uncoating, HA-mediated membrane fusion or the import of the viral genome into the cell nucleus. The pivotal role of HA in both binding and fusion, combined with its exposure to extracellular compounds, makes it a particularly popular target for neutralizing antibodies or small-molecule inhibitors. However, mutations in the amino acid sequence of the protein can happen during replication, in a process called antigenic drift. These mutations cause small changes in the appearance of an HA subtype that make it less recognizable by antibodies or inhibitors and therefore more successful in infection. Antigenic drift happens within the subtype and is not be confused with antigenic shift, a process in which two or more different strains of a virus infect the same host and combine into a new one, e.g., H3N2 and H5N1 could form H5N2. Because the human immune system would have difficulty recognizing such a new subtype, it may result in a highly dangerous, pandemic virus strain. An ideal antiviral drug should therefore universally apply to all virus strains, inhibiting a mechanism that is crucial for their replication, such as membrane fusion.

### 1.3 Hemagglutinin-mediated membrane fusion

During cell entry of influenza viruses, fusion of the viral and endosomal membranes is mediated by HA. The fusion of two lipid bilayers progresses over a number of intermediate states that are separated by appreciable energy barriers. Membrane fusion would therefore not occur on a biologically relevant timescale without the input of energy from a fusion catalyst, such as HA. HA is a trimeric glycoprotein consisting of 1647 residues that can
be divided into two parts: HA1 and HA2. HA1 is mostly globular and is mainly responsible for binding. HA2 is the fusion-active subunit and its central triple-stranded coiled coil forms the core of the protein. HA1 covers HA2, maintaining the protein in a metastable conformation.

A proposed pathway for HA-mediated membrane fusion is depicted in Figure 1.3. This sequence of events has been deduced from comparison of the known structures of HA at neutral pH (the prefusion state, Figure 1.3a) and at low pH (postfusion, Figure 1.3d), showing a large conformational change. HA1 dissociates upon acidification (Figure 1.3a-b), enabling HA2 to get from its metastable prefusion state into an extended intermediate via a coil-to-helix transition, inserting the amphipathic fusion peptide (red star in the figure) into the target membrane (Figure 1.3b-c). A helix-to-coil transition causes the intermediate to collapse, bringing the two membranes together for fusion (Figure 1.3c-e).

This working hypothesis for HA-mediated fusion leaves a number of open questions: What are the exact intermediates and what is the kinetics of the transformations between them? Which of the steps are pH-dependent? And which of the residues in HA are essential

![Figure 1.3: Proposed sequence of conformational changes in influenza HA that drive fusion, based on its (a) prefusion and (b) postfusion structure. HA is shown in cartoon representation, with each monomer in a different colour (red, green and blue). The position of one fusion peptide is marked with a red star. The gray bar represents the (top) target and (bottom) viral membrane, in which HA is anchored by a transmembrane domain (coloured bars). From (a) to (b), the pH drops and HA1 dissociates. From (b) to (c), the fusion peptides are released, a subsequent coil-to-helix transition extends HA2 and the fusion peptides insert into the target membrane. The extended intermediate collapses from (c) to (d), with a helix-to-coil transition and unfolding of the globular domain at the bottom of HA2. In (d) this globular domain has zippered up along the central coiled coil, carrying the transmembrane domain towards the fusion peptide. A tight interaction between the regions near the fusion peptide and the transmembrane domain (e) subsequently drives formation of the fusion pore. Adapted by permission from Macmillan Publishers Ltd: Nature Structural & Molecular Biology, copyright (2008).](image)
for the process? Are these conserved through different influenza subtypes? Also, how much energy can HA deliver in order to fuse the membranes? Consequently, how many HAs are needed to successfully fuse the membranes?

### 1.4 Molecular dynamics simulations

We set out to answer the research questions on HA-mediated membrane fusion using molecular dynamics simulations. These simulations give a trajectory of the atomic positions of the system (solute and solvent) over time, from which the desired properties can be extracted by further analysis. In order to generate such a trajectory, the displacement of a particle is calculated over a finite time step, based on the instantaneous forces between the particles at the beginning of that step. These forces are given by a predefined force field, parametrized against observables from experiments or quantum-mechanical simulations.

There are three major challenges in modeling HA and simulating its mechanistic behaviour using this method:

- **System size**
  The HA trimer consists of more than 15,000 atoms, between which the forces need to be calculated at every time step.

- **Large conformational change**
  Dynamical simulation of the dramatic conformational change in HA requires long simulation times, so a large number of time steps, and/or enhanced sampling methods.

- **pH-dependence**
  Methods for dynamic protonation of ionizable residues during molecular dynamics simulations increase the computational demands substantially.

As not all of these challenges can be tackled at the same time, our approach is to reduce system size and complexity by making informed assumptions and approximations. For example, only the parts of the protein that are relevant for the research question at hand are simulated.

### 1.5 Thesis outline

This thesis is devoted to improving our understanding of hemagglutinin-mediated membrane fusion, with a potential view on the development of a universal anti-influenza drug. After the brief introduction on HA-mediated membrane fusion given already here, Chapter 2 discusses the current literature on this subject more thoroughly. This includes a description of the intermediate membrane configurations involved in the membrane fusion
process and a quantification of the energy barriers between them. The pathway of the HA rearrangements and their role in mediating membrane fusion are discussed next. The description of the stochastic model that follows, explaining how multiple HAs can mediate fusion together, completes the current biophysical perspective on HA-mediated membrane fusion.

In Chapter 3, two molecular dynamics explicit solvent models are being compared, in the search for an accurate representation of peptide and protein conformations. The comparison focuses on the correct balance between folded and unfolded conformations, which is important in, for example, the coil-to-helix and helix-to-coil transition in HA. Three different peptides are simulated using an enhanced sampling technique, and the results are compared to their folding characteristics in experiments. The combination of force field and water model thus found is applied in all explicit solvent simulations reported on later in this thesis.

Chapter 4 presents the simulation results of only a small part of HA: the globular bottom of HA2. The stability of this domain is hypothesized to determine HA productivity, by regulating the amount of time that is available for the fusion peptides to insert into the target membrane. Steered molecular dynamics simulations in explicit solvent are used to study the unfolding behaviour of this domain and to determine which residues are critical for its stability. Preliminary results from single-particle fusion kinetics experiments are consistent with our expectation that the found residues influence HA-mediated fusion efficiency. Additionally, the results indicate a possible pH-dependence in globular bottom stability that might trigger this intermediate step, and even include a number of conserved residues that seem to be crucial for productive membrane fusion.

In the final chapter, the amount of energy that one HA can deliver to the fusion process is calculated using the confinement free energy method. Because only HA2 is actively involved in the fusion process, HA1 does not need to be simulated. Still, the calculation of this single quantity requires a huge computational effort and the use of enhanced sampling techniques. Furthermore, the used method currently does not seem feasible in explicit water, so an implicit water model is used. The amount of energy that is found for one HA is similar to what has been found experimentally for other fusion proteins, and supports a model in which a local cluster of three HAs is needed to mediate fusion.

References


