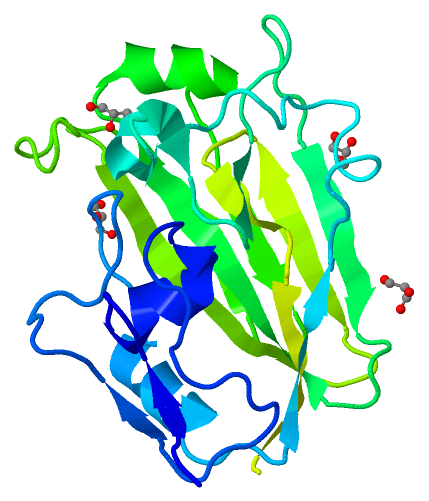
**Crystallization of Hemagglutinin Receptor Binding Domain to Determine the Binding Mode of a Novel Inhibitor**



**Atsarina Larasati Anindya**

**S2655063**

Supervisor:

Matthew R. Groves

This report is made for the completion of the first research project in Medical Pharmaceutical Sciences master program, University of Groningen

**Table of Contents**

[Abbreviations](#h.gjdgxs)

[List of Figures](#h.30j0zll)

[Abstract](#h.1fob9te)

[Introduction](#h.3znysh7)

[Overview of Influenza Virus and Possible Drug Targets](#h.2et92p0)

[Hemagglutinin as an Antiviral Target](#h.tyjcwt)

[Protein Structure Determination](#h.3dy6vkm)

[Crystallization of Protein - Small Molecules Complexes](#h.1t3h5sf)

[Summary of Methods Used to Obtain HA-RBD Protein Crystal](#h.4d34og8)

[Methods](#h.2jxsxqh)

[Protein Expression](#h.z337ya)

[Cell harvesting and Washing of Inclusion Bodies](#h.3j2qqm3)

[Denaturation of Protein in Inclusion Bodies](#h.1y810tw)

[Screening of Protein Refolding Methods (Small Scale Experiments)](#h.4i7ojhp)

[1. Fast Dialysis](#h.2xcytpi)

[2. Slow Dialysis](#h.1ci93xb)

[3. Immobilized Metal Affinity Chromatography (IMAC)](#h.3whwml4)

[4. Shock Dilution](#h.2bn6wsx)

[Scaling Up Protein Refolding (Large Scale Experiment)](#h.qsh70q)

[1. Scaling up with IMAC](#h.3as4poj)

[2. Scaling up with Shock Dilution](#h.1pxezwc)

[Thermofluor Assay for Crystallization Buffer\_](#h.49x2ik5)

[Scaling Up Protein Refolding (Large Scale Experiment) with Selected Thermofluor Buffers](#h.2p2csry)

[Crystallization with Sitting Drop Method](#h.147n2zr)

[Crystallization Screening with Dimethyl Sulfoxide (DMSO)](#h.3o7alnk)

[Results and Discussion](#h.23ckvvd)

[HA-RBD Protein Expression in Inclusion Bodies](#h.ihv636)

[Denaturation of Protein in Inclusion Bodies](#h.32hioqz)

[Screening of Protein Refolding Methods (Small Scale Experiments)](#h.1hmsyys)

[Scaling Up Protein Refolding (Large Scale Experiment)](#h.41mghml)

[1. Scaling up with IMAC](#h.2grqrue)

[2. Scaling up with shock dilution](#h.vx1227)

[Thermofluor Assay for Crystallization Buffer](#h.3fwokq0)

[Scaling Up Protein Refolding (Large Scale Experiment) with Selected Thermofluor (DSF) Buffers](#h.1v1yuxt)

[Crystallization Screening with Sitting Drop Method](#h.4f1mdlm)

[Crystal Diffraction](#h.2u6wntf)

[Crystallization Screening with Dimethyl Sulfoxide (DMSO)](#h.19c6y18)

[References](#h.3tbugp1)

[APPENDIX A: HA-RBD Gene for Protein Expression](#h.34g0dwd)

[APPENDIX B: Composition of Refolding Buffers in Small Scale Refolding](#h.1jlao46)

[APPENDIX C: Thermofluor Buffer Conditions](#h.43ky6rz)

# Abbreviations

EDTA : Ethylenediaminetetraacetic acid

PBS : Phospate-based Buffer  
TBS : Tris-based Buffer

BME : β-mercaptoethanol

LB : Lysogeny Broth

IPTG : Isopropyl β-D-1-thiogalactopyranoside

IMAC : Immobilized Metal Affinity Chromatography

DSF : Differential Scanning Fluorimetry

PDB : Protein Data Bank

SDS-PAGE : Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

RCF : Relative Centrifugal Force

# List of Figures

**FIG. 1.** (A) Crystal structure of hemagglutinin from A/H1N1/2009 (PDB entry 3LZG(42)), each monomer is differentiated with colors, with receptor binding domain (HA-RBD) located in amino acid sequence 63-286. (B) Details of HA63-286-RBD (PDB entry 3MLH) with sialic acid binding pocket. Blue color signifies N terminus and green color signifies C terminus of the domain.(17)………………………………………………………………………………...11

**FIG. 2.** Summary of small scale refolding methods with sample names according to method, buffer, and additives used. In samples containing arginine, a letter A was assigned and in samples containing βME, a letter β was assigned. Samples which do not contain the additive(s) were assigned the number 0 in the place of the absent additive(s)………………17

**FIG. 3.** First crystallization screening ……………………………………………………….21

**FIG. 4.** Optimization for crystallization buffer scheme ……………………………………..21

**FIG. 5.** Crystallization with DMSO scheme ………………………………………………...22

**FIG. 6.** Compound Soaking Experiment Conditions………………………………………...22

**FIG. 7.** Expression of HA-RBD from E. coli culture ……………………………………….24

**FIG. 8.** Concentration of denatured HA-RBD protein ……………………………………...25

**FIG. 9.** SDS-PAGE gel of fast dialysis refolding samples ……………………………….....25

**FIG. 10.** SDS-PAGE gel of slow dialysis refolding samples ……………………………….26

**FIG. 10.** SDS-PAGE gel of IMAC refolding samples ……………………………………...26

**FIG. 11.** SDS-PAGE gel of shock dilution refolding samples ……………………………...26

**FIG. 12.** Chromatogram of protein refolding results showing protein peak around 0.5 column volume (fraction 17) with dimer aggregation peak around 0.35 column volume (fraction 7).

………………………………………………………………………………………………..26

**FIG. 13.** SDS-PAGE gel of size exclusion chromatography fractions ……………………...27

**FIG. 14.** SDS-PAGE of scale up with IMAC method ………………………………………27

**FIG. 15.** SDS-PAGE of scale up with shock dilution method ……………………………...28

**FIG. 16.** Thermofluor results in different buffer conditions. (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in F1 buffer (100 mM Na-citrate). (C) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7). (D) Thermofluor curve for HA-RBD in D4 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8). (E) Thermofluor curve for HA-RBD in B5 buffer (60 mM Tris-HCl at pH 8.5). (F) Thermofluor curve for HA-RBD in C9 buffer (60 mM Na-phosphate and 250 mM NaCl at pH 7). (G) Thermofluor curve for HA-RBD in F2 buffer (100 mM Na-sulphate). (H) Thermofluor curve for HA-RBD in B4 buffer (60 mM Tris-HCl at pH 8). (I) Thermofluor curve for HA-RBD in D5 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8.5)………….29

**FIG. 17.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in C9 buffer (60 mM Na-phosphate and 250 mM NaCl at pH 7). (C) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7)……………30

**FIG. 18.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in B5 buffer (60 mM Tris-HCl at pH 8.5). (C) Thermofluor curve for HA-RBD in B4 buffer (60 mM Tris-HCl at pH 8)……………………………………………………...30

**FIG. 19.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in D4 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8). (C) Thermofluor curve for HA-RBD in D5 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8.5)………….31

**FIG. 20.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in B8 buffer (10% DMSO, 10 mM Tris-HCl, and 100 mM NaCl at pH 8) ……………………………………………………………………………………………..31

**FIG. 21.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in F2 buffer (100 mM Na-sulphate). (C) Thermofluor curve for HA-RBD in F1 buffer (100 mM Na-citrate). (D) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7) …………………………………………………………………………...32

**FIG. 22.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in D1 buffer (60 mM HEPES and 250 mM NaCl at pH 7.5) …………………32

**FIG. 23.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. Other curves represents stability of HA-RBD in several metal-containing buffers. See **APPENDIX C** for a complete list of buffer solutions used .....................................................................................................33

**FIG. 24.** Measured concentration of protein at different volume points ……………………33

**FIG. 25.** Measured concentration of protein at different volume points (in graph) ………...34

**FIG. 26.** SDS-PAGE of refolded HA-RBD, concentration 5.6 mg/Ml ……………………..35

**FIG. 27.** HA-RBD crystals from sitting drop method ………………………………………35

**FIG. 28.** Peak profile of refolded HA-RBD sample used in the crystallization screening in **FIG. 27.**………………………………………………………………………………………36

**FIG. 29.** HA-RBD protein crystal after purification by size exclusion chromatography …...36

**FIG. 30.** Electron density map of protein crystal diffraction. Left image shows the protein as multiple crystals and the right image shows the protein as a single crystal.……………………………………………………………………………………..…37

**FIG. 31.** Electron density map of protein crystal diffraction. Left image shows the protein as multiple crystals and the right image shows the protein as a single crystal………………….37

**FIG. 32.** Diffraction statistics for HA-RBD protein…………………………………………38

**FIG. 33.** Structure solution of HA-RBD crystal soaked in small compound at low concentration…………………………………………………………………………………39

**FIG. 34.** Structure solution of HA-RBD crystal soaked in small compound in high concentration…………………………………………………………………………………40

# Abstract

Hemagglutinin (HA) is one of the major surface protein in influenza virus, along with neuraminidase and M2 proteins. It plays a major role at the beginning of the viral cycle, allowing the influenza virus to enter host cells by binding to sialic acid on the host cell’s surface. While drugs targeting neuraminidase and M2 proteins have been marketed widely, such as amantadines, rimantadines, zantamivir, and oseltamivir,(15) hemagglutinin-directed drugs still remain to be developed. In addition to its major role in the initial stage of viral attachment to host cells, hemagglutinin is not prone to mutation, unlike neuraminidase and M2. The conformation of its receptor-binding domain is relatively conserved due to the binding specificity to sialic acid.(24) Mutation in this region may risk the ability of the virus to bind to the host cell in the first place. The sialic acid binding pocket in hemagglutinin protein is therefore an interesting starting place to develop antiviral drugs. Using *in silico* modelling ANCHOR.QUERY, HA-directed small molecules have been synthesized and screened for activity. Crystallization is needed to determine the binding mode of several small molecules to the sialic acid binding pocket.

HA-receptor binding domain (HA-RBD) was known to be successfully expressed from *E. coli* bacteria as insoluble inclusion bodies, which need to be denatured by a strong chaotropic agent, for example urea.(5)(12) Denatured protein can be refolded into its native state in optimal conditions, otherwise it can be misfolded into aggregates. To estimate this optimal condition, 4 refolding methods were screened: fast dialysis, slow dialysis, Immobilized Metal Affinity Chromatography (IMAC), and shock dilution. Arginine and ßME were also used in the experiment to help promote correct refolding. HA-RBD protein was found to be refolded in all refolding methods to various degrees, with IMAC and shock dilution methods having the best results for producing monomeric protein in solution.

HA-RBD protein crystals were obtained from scaling up expression and refolding protocols (using the shock dilution method), at a protein concentration of 5.6 mg/mL. HA-RBD protein was shown to have a better thermal stability in our dialysis buffer than the buffer used in a previous study by DuBois *et al.* (2011). Diffraction of HA-RBD crystals shows the protein has crystallised and we have performed crystal-compound soaking experiments.

# Introduction

## Overview of Influenza Virus and Possible Drug Targets

Influenza virus is a major cause of morbidity and mortality due to rapidly changing mutations and there is a shortage of available vaccines and antivirals. The type A influenza virus can spread widely, causing both epidemics and pandemics, and thus is of particular interest for antiviral development. The type A influenza virus is further subtyped according to the number of its surface proteins, hemagglutinin (HA) and neuraminidase (NA). Subtypes A that commonly infect people are H1N1, H2N2, H3N2. Another group of subtype A, which caused the most recent influenza outbreak is the avian influenza group, including subtypes H5 and H7. Each A subtype has several strains which can appear over time through a mutation process known as antigenic drifting. Mutations can also occur through antigenic shifting, which creates a new subtype with different surface protein (indicated by surface protein number).(13)(15)(24)(43)

Influenza virus genetic material is contained in single-stranded RNAs with negative polarity. The viral genome is divided into several segments and each segment is encapsulated by viral nucleoproteins to form a RNP (ribonucleoprotein). Each RNP also contains RNA polymerase components (PA, PB1, and PB2) which are necessary for transcription and replication. The RNP unit is therefore capable of independent transcription and replication of the viral genome segment.(4)

The life cycle of influenza virus starts with viral attachment to sialic acid on the host cell receptor and subsequent entry into by receptor-mediated endocytosis. The acidic environment of endosomes promotes uncoating and entry of viral genetic materials into the cytoplasm. This triggers a change in the hemagglutinin conformation which leads to membrane fusion.(24) Before fusion, H+ ion flows through M2 ion channel to cause dissociation of M1 protein from RNP so that later it can be mobilized from the cytoplasm to the host cell nucleus. Membrane fusion between viral membrane and endosome causes the internal core of the virus to be released into the host cell. When the virus is ready for replication, viral RNPs will move inside the host cell's nucleus through the nuclear pores. Once inside, the viral RNA polymerase attached to the RNP catalyzes the transcription of negative-sense viral RNA (vRNA) into positive-sense mRNA (messenger RNA) and cRNA (complementary RNA). Subsequent priming in an mRNA sequence resulting in a structure with fragments capping 5’ end and additional poly(A) tail at the 3' end. mRNAs are transported outside the host cell's nucleus and translated into proteins using the host cell's ribosomes. Proteins synthesized from mRNAs make up the components used in replication, transcription, and the new viral structure (including surface proteins). Unlike mRNAs, cRNAs are not primed and thus serve as a template for new viral genome synthesis. The vRNAs synthesized from cRNAs are direct copies of viral genome, which will be assembled together with proteins synthesized from mRNAs into new viruses. The new vRNAs transcribed from cRNAs and proteins translated from mRNAs are exported to the plasma membrane and assembled to form new viruses. To enable the budding virus to be released from the host cell’s plasma membrane, neuraminidase catalyzes sialic acid degradation from the host cell and the new viruses are released from the host cell’s plasma membrane.(4)(14)(22)(30)

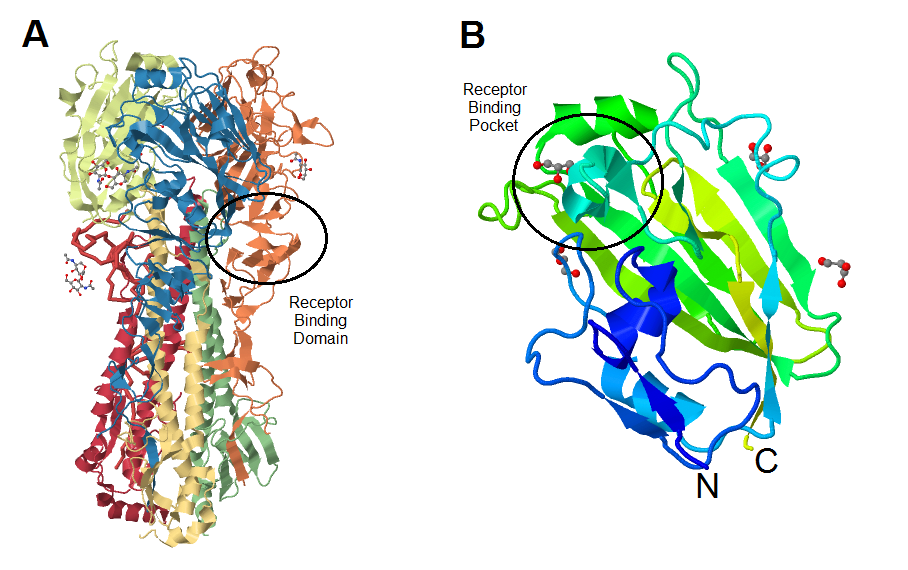
Several surface proteins of the influenza virus play significant roles in the viral life cycle, particularly at the early stage of interacting with their host cells. The major surface proteins are hemagglutinin, neuraminidase, and matrix protein 2. Hemagglutinin (HA) allows the influenza virus to enter host cells by attaching its sialic acid receptor to sialic acid from the host cell and is the major antigen. Neuraminidase (NA) allows the release of new viral particles from host cells by catalyzing the cleavage of linkages to sialic acid. Matrix protein 2 (M2) maintains the pH balance across viral membranes during host cell entry, ensuring the stability of the whole process.(24)(30) These three proteins are the main targets of influenza antiviral research.

There are two types of antiviral drugs that are currently available, M2 ion channel blockers and neuraminidase inhibitors. Antiviral drugs that target M2 ion channel belong to the adamantanes class, such as amantadine(27) (sold under the trade name Symmetrel by Endo Pharmaceuticals) and rimantadine (sold under the trade name Flumadine(21)). Amantadine works by increasing endosomal pH or blocking M2 specifically. In the neuramindase inhibitor class, oseltamivir (sold under the trade name Tamiflu by Roche(8)), and zanamivir (sold under the trade name Relenza by GlaxoSmithKline**)** are currently on the market. The problem with these current treatments is the genes encoding both proteins are highly susceptible to mutation. There have been many reported cases of rapid cross-resistance due to a single point mutation, particularly for M2 ion channel blockers.(15) As of 2013, a study found that 45.2% of the currently circulating influenza A virus is resistant to adamantanes.(16) Another study found that 27.3% test subjects infected with influenza A H1N1 virus is resistant to oseltamivir, although the number of test subjects is considerably smaller.(36)

While drugs targeting the other two major surface proteins have been on the market for some time, hemagglutinin-directed drugs still remain to be developed. As hemagglutinin is responsible for viral attachment to the host cell, drugs targeting hemagglutinin could potentially inhibit viral infection at a very early stage.

## Hemagglutinin as a Possible Antiviral Target

Hemagglutinin is a trimeric protein (size: 220 kDa) that, in the infectious stage, consists of 2 subunits for each monomer, the HA1 subunit which makes the globular head and the HA2 subunit which makes the elongated tail of the protein (see **FIG. 1.**). The HA1 subunit contains the sialic acid binding domain that can attach to the host cell. Upon attachment, hemagglutinin undergoes a dramatic conformational change to allow the HA2 subunit to initiate membrane fusion. Even though hemagglutinin conformation can be changed during sialic acid binding, this binding domain in subunit HA1 remains relatively unchanged.(20) (37)



**FIG. 1.** (A) Crystal structure of hemagglutinin from A/H1N1/2009 (PDB entry 3LZG(42)), each monomer is differentiated with colors, with receptor binding domain (HA-RBD) located in amino acid sequence 63-286. (B) Details of HA63-286-RBD (PDB entry 3MLH(17)) with sialic acid binding pocket. Blue color signifies N terminus and green color signifies C terminus of the domain.

As a notable example, the amino acids constituting the receptor binding pocket in H3 subtype virus, which emerges in 1968, are no different than the amino acids in H3 present in 1981.(24)(40) This is primarily due to the necessary binding specificity of hemagglutinin to sialic acid. The binding is also further differentiated according to the species infected, which have specific galactose linkages to the sialic acid.(20)(24) This binding specificity is crucial in ensuring the viral entry to the host cell and continuation of the viral life cycle. As a consequence, a mutation in the structure of receptor binding pocket may risk the virus inability to bind to host cell in the first place.

The hemagglutinin sialic acid receptor binding domain (HA-RBD) is located within residues 63-286 from influenza virus H1N1 and its size is approximately 25 kDa, containing 4 cysteine residues.(2) In each influenza virus type, HA-RBD residues can have different conformations that bind only to a specific sialic acid. For example, the HA-RBD residue in human influenza virus recognizes only sialic acid attached to galactose by a α2,6-linkage, that is present in human epithelial cells but absent in other species. Avian species, on the other hand, recognise an α2,3-linkage between galactose and sialic acid.(20)

To develop an effective HA-RBD inhibitor, several small molecules that act as sialic acid antagonists have been synthesized and screened for activity by Tryfon-Zarganis Tzitzitkas (Drug Design research group, University of Groningen), using an *in silico* model (ANCHOR.QUERY) and an *in vitro* model. This study aims to investigate whether the small molecules get their activity from binding to HA-RBD or by other means by studying the HA-RBD that is bound to the designated sialic acid antagonists by protein crystallography. Co-crystallized complex of HA-RBD and small molecules can be analysed by X-ray diffraction experiments to determine their structure and mode of binding of the compounds.

## Protein Structure Determination

In protein crystallography, an X-ray beam is used to determine a protein structure from a crystallized sample. This beam has the same range of wavelength as the interatomic distance inside the sample, which will enable structure determination to the atomic level. Once the X-ray beam hits the sample, the electrons in sample molecules will scatter the beam which can be read by the detector as a diffraction pattern of reflections. Because each molecule configuration is unique, the reflection pattern will also be specific to that molecule. Each reflection from a crystal is characterized by its amplitude and phase. This electron-density distribution, or electron-density map, is the primary result of a diffraction experiment.(41)

Electron-density maps usually combine the basic map and the difference map. Basic electron-density map uses Fourier transformation of observed reflection amplitudes Fobs and the phases are calculated from a model. This map shows an approximation model of the structure. Difference map calculates the differences between the observed and calculated amplitudes and calculated phases.This map shows which parts exclusively exists in the model and the part which is only observed in the experiment, so that the user can make further corrections. The quality of electron-density maps depend on the number of reflections or the resolution of the diffraction data, which is represented by the unit Å.(41)

## Crystallization of Protein - Small Molecules Complex

To obtain protein-small molecules complex crystals, one requires a relatively high concentration of pure protein. The experimental design to get concentrated pure HA-RBD is as follows: expression of HA-RBD protein from glycerol stock (HA-RBD genes inserted into *E. coli* bacteria), washing and denaturation of inclusion bodies, protein refolding and further purifications, concentration of refolded protein, and co-crystallization by sitting-drop method. After co-crystallization or soaking expereiments were performed, the crystal was subjected to high intensity X-ray light from a synchrotron source to obtain structural data.

## Summary of Methods Used to Obtain HA-RBD Protein Crystal

HA-RBD was known to be successfully expressed from *E. coli* bacteria as insoluble inclusion bodies.(2)(17) The inclusion bodies need to be dissolved before it can be processed further. In order to dissolve the inclusion bodies, the protein has to be denatured until it loses its quarternary (if said protein consists of more than one subunit), tertiary, and secondary structure.(3)(12)(29) As the protein still retains structural information from its primary structure, under the optimal condition, the protein should refold back to its native, functional state.

Proteins in the unfolded state can go through two different pathways of refolding, the one leads to a native state and the other leads to misfolded state, which will commonly result in aggregation due to highly exposed hydrophobic surface (protein-protein hydrophobic interaction are more likely to occur). Because in an optimal folding condition, native folding pathway is the most stable state, the energy required to go through this pathway is less than energy required in the misfolding pathway, thus it is expected to be the default process. Unfortunately, there is no information about optimal folding conditions for an individual protein that can be derived from the protein sequence. However, one can estimate the required condition for a protein to behave as in its natural environment, such as optimization of pH, salt contents in buffer, and temperature.

In order to obtain the optimal condition, several refolding methods were screened in this report. There are 2 basic requirements for protein refolding, one is the removal of denaturant by reducing its concentration and the optimization of native refolding condition once the denaturant is removed. The optimal condition introduced to the denatured sample depends on the pH, temperature, and addition of several chemicals that can help promote the correct refolding process. There are 4 different methods that allows reduction of denaturant concentration while introducing the optimal condition (by means of refolding buffers): fast dialysis,(31) slow dialysis,(38) Immobilized Metal Affinity Chromatography (IMAC) refolding,(2)(17) and shock dilution.(25)

Both fast dialysis and slow dialysis methods use the advantage of changing the buffer composition from the sample with the refolding buffer (a buffer that promotes optimal refolding condition), thus reducing the concentration of the denaturant. In fast dialysis, the amount of denaturant is dropped sharply by introducing the refolding buffer which contains no denaturant, at a ratio of 1:100 (denatured sample:refolding buffer). As not all proteins have natural one-state folding kinetics, it is presumed that a certain amount of denaturant also helps the refolding process. Protein molecules may rapidly equilibrate in several unfolded states before completely refolded. The presence of denaturant may help preventing these unfolded states to enter a misfolding pathway. Thus, rather than drastically dropping the amount of denaturant, an experiment with gradual decrease of denaturant was also needed for screening, as represented by slow dialysis.

Immobilized Metal Affinity Chromatography (IMAC)-assisted refolding makes use of selective binding of the protein to a metal resin, thus denaturant and other possible contaminations can be washed away. This protocol uses Ni-NTA resin, which can bind selectively to the His-tag sequence in the expressed HA-RBD. The bound protein is washed several times with refolding buffer to promote correct refolding and stabilization of refolded protein before it was removed from the resin by competetion with high concentrations of imidazole.

Shock dilution method simply uses rapid decreasing of the denaturant by diluting the sample into large amount of refolding buffer. However, this method has a disadvantage of a very high volume of refolded sample, thus further concentration is needed.

Although buffer properties (pH, salt contents, and temperature) play the most significant role in protein refolding, some additives can be used to make the refolded protein more stable and reducing the potential to form a partially-folded protein sample. The additives used in this experiment are arginine and -mercaptoethanol (BME).(25)(10) As HA-RBD contains several hydrophillic regions on its surface, arginine helps reducing aggregation from protein-protein interaction by binding to these regions.(10) BME is a reducing agent, which can help the folding of proteins containing cysteine residues. HA-RBD contains 5 cysteine residues, in which the thiol group can bind to each other by forming a disulfide bond (S-S bond). In HA-RBD, the disulfide bonds are located between C97-C139 and C64-C76, both are outside the actual binding pocket but in the immediate vicinity. We can hypothesize that disulfide bonds are created in the late stages of folding and have no effect on the initial folding process, though it can improve the solubility of folded protein. In order to avoid incorrect cross-linking between pairs of cysteines, the reducing agent BME can be added in the refolding process.

# Methods

## Protein Expression

HA-RBD protein was expressed from *E. coli* glycerol stock (HA-RBD DNA(17) inserted in vector pETM11, transformed in *E. coli* BL21\*(DE3), see **APPENDIX C**) strain by Tryfon-Zarganis Tzitzikas at Drug Design research group, University of Groningen).

*E. coli* from bacteria stock was pre-cultured in 25 mL LB media(7) (with addition of 10 mg/mL kanamycin and 10 mg/mL chloramphenicol) by incubating the mixture in a shaking incubator at 37oC overnight.

Pre-cultured bacteria was transferred into 1 L LB media (with addition of 1 mL kanamycin and 1 mL chloramphenicol, both in 10 mg/mL concentration) and incubated in a shaking incubator at 37oC. When the optical density of the culture reached 0.6, 1 mM isopropyl-β-D-thiogalactoside (IPTG) was added to the culture. Incubation was continued in a shaking incubator at 18oC overnight.

To confirm HA-RBD protein expression, cultured bacteria (1, 2, 5, 10, 15, and 20 μL concentration) was mixed with 5 μL of Coomassie Blue dye for SDS-PAGE(18) with 8% acrylamide gel.

## Cell harvesting and Washing of Inclusion Bodies

Bacteria cells were harvested in the form of pellets after centrifugation (Beckman SLA-3000 rotor) at 5000 rpm and 4oC for 30 minutes.(25)(26) After harvesting, the expressed HA-RBD protein was contained within the inclusion bodies. To obtain a relatively pure sample of HA-RBD, the inclusion bodies needed to be separated from other cell materials. Pure HA-RBD sample was needed to have a standardized material for comparison should the experiment be repeated in the future.

To break down the outer cell walls, the harvested cells were suspended in 70 mL lysis buffer containing detergent (50 mM Tris-HCl, 300 mM NaCl, 5% (v/v) glycerol, 3 mM βME, 5 mM EDTA, and 0.5% (v/v) Triton X-100 at pH 8) and was sonicated for 2 minutes. After sonication, 0.01 mg/mL DNAse and 0.1 mg/mL lysozyme were added into the suspension to digest the DNA and break down inner cell walls, respectively. 10 mM MgSO4 was also added to chelate the EDTA in the lysis buffer. The mixture was incubated in room temperature for 15 minutes then centrifuged at 5000 rpm for 15 minutes. These washing procedures were repeated 5 times, with the last 2 repetitions without the addition of MgSO4, DNAse, and lysozyme after sonication.(26)

Inclusion bodies collected from the previous washing steps were suspended in the same lysis buffer, in the absence of Triton X-100 detergent (50 mM Tris-HCl, 300 mM NaCl, 5% (v/v) glycerol, 3 mM βME, and 5 mM EDTA). The same washing steps were performed (without using MgSO4, DNAse, and lysozyme) 3 times.(26)

## Denaturation of Protein in Inclusion Bodies

HA-RBD protein in the form of inclusion bodies could not be processed further, therefore HA-RBD inclusion bodies were denatured in a solution of Urea at high concentration. Inclusion bodies collected from all washing steps were divided into 4 Falcon tubes. Each sample was dissolved in 35 mL TBS-buffered urea solution (8 M urea, 50 mM Tris-HCl, and 150 mM NaCl at pH 8) then incubated at 37oC overnight. Protein concentration from each sample was measured by absorption spectroscopy at 280 nm (extinction coefficient 2.046).

## Screening of Protein Refolding Methods (Small Scale Experiments)

The dissolved HA-RBD protein was in a denatured state and needed to be refolded back into its native state. Several protein refolding methods were screened with small amounts of unfolded HA-RBD protein solution: fast dialysis, slow dialysis, IMAC-assisted, and shock dilution. Two additives, arginine and βME, were also added to promote correct refolding.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | **Additives in Buffer** | | | |
| **Method** | **Buffer** | Aβ | A0 | 0β | 00 |
| 1  (fast dialysis) | TBS | 1AβT | 1A0T | 10βT | 100T |
| PBS | 1AβP | 1A0P | 10βP | 100P |
| 2  (slow dialysis) | TBS | 2AβT | 2A0T | 20βT | 200T |
| PBS | 2AβP | 2A0P | 20βP | 200P |
| 3  (IMAC) | TBS | 3AβT | 3A0T | 30βT | 300T |
| PBS | 3AβP | 3A0P | 30βP | 300P |
| 4  (shock dilution) | TBS | 4AβT1 | 4A0T1 |  | |
| 4AβT2 | 4A0T2 |
| PBS | 4AβP1 | 4A0P1 |
| 4AβP2 | A0P2 |

**FIG. 2.** Summary of small scale refolding methods with sample names according to method, buffer, and additives used. In samples containing arginine, a letter A was assigned and in samples containing βME, a letter β was assigned. Samples which do not contain the additive(s) were assigned the number 0 in the place of the absent additive(s).

### 1. Fast Dialysis

1 mL of unfolded HA-RBD protein was dialysed in 100 mL of each refolding buffer (see **APPENDIX A** for the composition of buffers used) for 4 hours then dialysed further with 1 L of the respective buffer for 16 hours.

Each sample from the fast dialysis method was loaded into a Ni-NTA spin column (equilibrated with respective buffer) in a small filtered tube (used QIAGEN miniprep columns) and eluted with 700 μL standard elution buffer (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, 5% (v/v) glycerol, and 3 mM βME). All samples were analysed on an SDS-PAGE gel.

### 2. Slow Dialysis

1 mL of unfolded HA-RBD protein was diluted in 2 mL of each refolding buffer (see **APPENDIX A** for the composition of buffers used), with addition of 2.7 M urea to each buffer, then dialysed in 100 mL of the same urea-containing buffer overnight. Each dialysed sample was further dialysed with 100 mL of the respective buffer (this time, the urea concentration was reduced to 1.3 M) overnight. Each dialysed sample was further dialysed with 100 mL of the respective buffer (this time without urea and the NaCl concentration was increased to 250 mM) overnight.

Each sample from the slow dialysis method was loaded into Ni-NTA spin column (equilibrated with respective buffer) in a filtered small tube and eluted with 700 μL standard elution buffer. All samples were analysed on an SDS-PAGE gel.

### 3. Immobilized Metal Affinity Chromatography (IMAC)

700 μL of unfolded HA-RBD protein was loaded into 200 μL Ni-NTA spin column (each column was equilibrated with each refolding buffer in **FIG. 2**, see **APPENDIX A** for the composition of buffers used) in a filtered small tube and the flow through was collected. Each column was washed 2 times with respective refolding buffer and eluted with 700 μL standard elution buffer. All samples were analysed on an into SDS-PAGE gel.

### 4. Shock Dilution

40 μL of unfolded HA-RBD protein was diluted in 40 mL of each refolding buffer (see Appendix A for the composition of buffers used) .

1 mL of each sample from the shock dilution method was centrifuged at 4500 rpm for 20 minutes to remove precipitation. All samples were run analysed on an run into SDS-PAGE gel.

Selected samples from all small scale refolding methods were characterised on a size exclusion column. From the peak profile results, IMAC and shock dilution were deemed suitable for the large scale experiment as the two methods show the highest refolded protein concentration compared to other methods used.

## Scaling Up Protein Refolding (Large Scale Experiment)

For scaling up process, a fresh bacterial culture was used. Procedures used for protein expression, cell harvesting and washing of inclusion bodies, and denaturation of protein in inclusion bodies were the same as described above for small scale experiments.

### 1. Scaling up with IMAC

34 mL denatured HA-RBD sample (3.7 mg/mL) was mixed with 2 mL Ni-NTA beads (equilibrated with refolding buffer AβT) in a 50 mL Falcon tube and incubated on a moving platform at 4oC for 60 minutes.

The mixture was loaded into a filtered column and the flowthrough was kept for further use. The column was washed with refolding buffer AβT for 2 times then the HA-RBD protein was eluted with 8 mL standard elution buffer.

It was found out that a significant amount of HA-RBD protein was still present in the flowthrough, therefore the flowthrough needed to be reloaded back into the column.

Before reloading the flowthrough, the column was washed with 0.5 M NaOH for 30 minutes, followed by a cleaning solution (1 mM EDTA and 1% w/v SDS) for 15 minutes, followed by 8 M urea overnight, then recharged with 100 mM NiCl2 solution for 15 minutes. The flowthrough was reloaded into the column for 2 times, each time with a preceding cleaning procedure as described above. Total eluted protein was collected and dialysed with 1 L dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8) for 4 hours to remove imidazole from the elution buffer.

### 2. Scaling up with Shock Dilution

3 mL denatured HA-RBD sample (3.7 mg/mL) was put drop by drop into stirring 100 mL refolding buffer AβT. The mixture was left to stir for 1 hour.

The mixture was dialysed against 2 L dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8) overnight at 4oC, stirred.

The dialysed sample was concentrated to 20 mL using a stirred cell Amicon-6050 concentrator (with 10 kDa membrane pore size) at 4oC and then concentrated in a centrifuge at 2000 RCF and 4oC.

## Thermofluor Assay for Crystallization Buffer (24)(32)

Information about the suitable buffer condition for concentrated protein sample was needed to prepare crystallization. Thermofluor assay is used to measure protein stability in different conditions by measuring its unfolding profile over increasing temperature. The assay can differentiate folded and unfolded states of a protein by means of a fluorescence marker, for example SYPRO Orange. The fluorescence marker can bind into hydrophobic surface of the protein which is normally not exposed in a folded protein. As the protein unfolds, the hydrophobic surface inside is exposed and an increased fluorescence signal can be detected.

For this thermofluor assay, 500 μL concentrated HA-RBD sample (2 mg/mL) was mixed with 2.5 μL SYPRO Orange dye (5000X concentrated solution in DMSO)(33) and spun at 1000 rcf for 1 minute. The mixture was put into 96-well plate with different buffer conditions (see **APPENDIX C**). Each well contains 5 μL mixture of concentrated protein and SYPRO Orange dye and 45 μL buffer. The samples in 96-well plate was put in a RT-PCR (BioRad CFX96) machine and run through several thermal cycles, with temperature increasing by 0.5oC per cycle. The higher the unfolding temperature, the more stable the buffer condition is for the HA-RBD protein. The most stable buffers condition was selected for crystallization.

## Scaling Up Protein Refolding (Large Scale Experiment) with Selected Thermofluor Buffers

3 mL of denatured protein sample (3.7 mg/mL) was put drop by drop into stirring 100 mL refolding buffer AβT. The mixture was left to stir for 1 hour.

The mixture was dialysed in 2 L dialysis buffer selected from the thermofluor assay (10 mM Tris-HCl, 100 mM NaCl, and 10 mM Na-citrate at pH 8) overnight at 4oC, stirred.

The dialysed sample was concentrated to 20 mL using the Amicon-6050 concentrator (membrane pore size 10 kDa) at 4oC and then concentrated in a centrifuge at 2000 RCF and 4oC to 4 mg/mL.

## Crystallization with Sitting Drop Method

Crystallization process was done with the assistance of Yuanze Wang from Protein Structural unit, Drug Design research group, University of Groningen. 500 μL buffer condition (see **FIG. 3.**) was put in each well from the sitting drop 96-well plate.

1 μL of protein sample (5.6 mg/mL) was dropped onto an elevated post above each well and 1 μL of buffer condition (from the same well) was dropped on top of the concentrated protein droplet. The 96-well plate used was sealed and incubated at 18oC.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Buffer | | | | | |
| MES  (pH 6.5) | HEPES (pH 7) | Tris-HCl (pH 7.5) | Tris-HCl (pH 8) | Tris-HCl (pH 8.5) | Tris-HCl (pH 9) |
| PEG-2000 | 5% | A1 | A2 | A3 | A4 | A5 | A6 |
| 10% | B1 | B2 | B3 | B4 | B5 | B6 |
| 20% | C1 | C2 | C3 | C4 | C5 | C6 |
| 30% | D1 | D2 | D3 | D4 | D5 | D6 |
| 40% | E1 | E2 | E3 | E4 | E5 | E6 |

**FIG. 3.** First crystallization screening scheme

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Buffer (Tris-HCl) | | | |
| pH 8.3 | pH 8.5 | pH 8.8 | pH 9 |
| PEG-2000 | 22% | 1A | 1B | 1C | 1D |
| 24% | 2A | 2B | 2C | 2D |
| 26% | 3A | 3B | 3C | 3D |
| 28% | 4A | 4B | 4C | 4D |
| 30% | 5A | 5B | 5C | 5D |
| 32% | 6A | 6B | 6C | 6D |
| 34% | 7A | 7B | 7C | 7D |
| 36% | 8A | 8B | 8C | 8D |
| 38% | 9A | 9B | 9C | 9D |
| 40% | 10A | 10B | 10C | 10D |
| 42% | 11A | 11B | 11C | 11D |
| 44% | 12A | 12B | 12C | 12D |

**FIG. 4.** Optimization for crystallization buffer scheme

The crystal formed from condition C7 (see **FIG. 4**) was analysed by diffraction at beamline P11, PETRA III, Deutsches Elektronen-Synchrotron (DESY) in Hamburg, Germany. The resulting diffraction was analyzed by XDS.(19)

## Crystallization Screening with Dimethyl Sulfoxide (DMSO)

Because the small compounds which would be co-crystallized with protein sample were dissolved in DMSO, the effect of DMSO presence in protein sample crystallization had to be determined. The screening was done with hanging drop method.

|  |  |  |  |
| --- | --- | --- | --- |
| DMSO 5%  10 μl glycerol | DMSO 10%  8 μl glycerol | DMSO 15%  3 μl glycerol | DMSO 18% |

**FIG. 5.** Crystallization with DMSO scheme

All conditions above were put in 0.1 M Tris buffer pH 8.8 with 36% (w/v) PEG-2000.

## Compound Soaking Experiment

Based on the optimization result for HA-RBD crystallization, the crystal was seeded and soaked in HA-directed small compounds with various concentrations.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PEG-2000 26% w/v** | | | | |
| pH | 8.2 | 8.5 | 8.8 | 9 |
| Compound |  |  |  |  |
| 5% v/v | 26A | 26E | 26I | 26M |
| 10% v/v | 26B | 26F | 26J | 26N |
| 15% v/v | 26C | 26G | 26K | 26O |
| 20% v/v | 26D | 26H | 26L | 26P |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PEG-2000 30% w/v** | | | | |
| pH | 8.2 | 8.5 | 8.8 | 9 |
| Compound |  |  |  |  |
| 5% v/v | 30A | 30E | 30I | 30M |
| 10% v/v | 30B | 30F | 30J | 30N |
| 15% v/v | 30C | 30G | 30K | 30O |
| 20% v/v | 30D | 30H | 30L | 30P |

**FIG. 6.** Compound Soaking Experiment Conditions

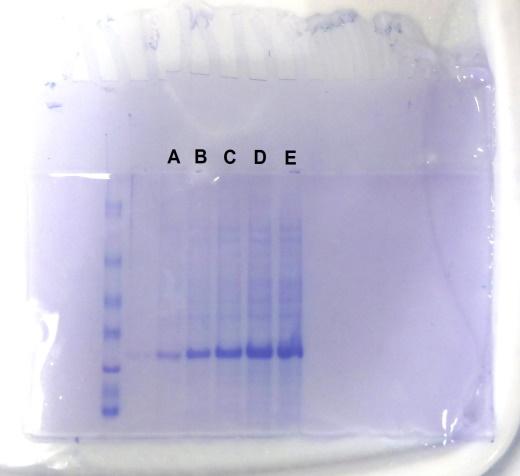
In all **FIG. 6** conditions, 0.1 M Tris-HCl buffer and 10 μM glycerol were also added to promote crystal stability. Crystals produced from this experiment were analysed by diffraction at beamline P11, PETRA III, Deutsches Elektronen-Synchrotron (DESY) in Hamburg, Germany. The resulting diffraction was analyzed by XDS(19) and the structure of protein-small compound complex was solved by molecular replacement, using the 3MLH structure from the Protein Data Bank RCSB ([www.rcsb.org](http://www.rcsb.oprg)) as the starting model.(17)(6)

# Results and Discussion

## HA-RBD Protein Expression in Inclusion Bodies

HA-RBD protein was expressed from *E. coli* glycerol stock (HA-RBD DNA inserted in vector pETM11, transformed in *E. coli* BL21\* (DE3) strain) as insoluble inclusion bodies. High expression of non-native protein inside a bacterial cell tend to lead to the formation of aggregates, as these proteins are usually partly folded and misfolded after expression(34), due to the lack of conditions which can promote correct folding. This non-conducive environment for correct refolding may arise from lack of chaperones and enzymes catalyzing folding, high concentration of folding intermediates which have limited solubility, or the reducing environment inside *E.* *coli* which can degrade native disulfide bonds in protein and make it insoluble.(14) Lack of glycosylation was also thought to contribute to the protein misfolding(14) however DuBois *et al.* (2011) stated that glycosylation is not required for proper HA-RBD folding. Glycosylation on amino acid asparagine (Asn94) in HA-RBD is possibly only required for immune evasion or receptor-binding affinity.(17)

Because inclusion bodies can be easily separated from other cell materials, the formation of inclusion bodies can be advantageous in obtaining a pure HA-RBD protein sample. After breaking the bacterial outer cell walls with a detergent-containing solution (50 mM Tris-HCl, 300 mM NaCl, 5% glycerol, 3 mM βME, 5 mM EDTA, and 0.5% Triton X-100 at pH 8) and sonication, followed by the dissolution of inner cell walls by lysozyme and DNA degradation by DNAse, dissolved cell materials can be separated from insoluble inclusion bodies by centrifugation.



**FIG. 7.** Expression of HA-RBD from E. *coli* culture.

A = 1 μL sample loading volume

B = 2 μL sample loading volume

C = 5 μL sample loading volume

D = 10 μL sample loading volume

E = 15 μL sample loading volume

F = 20 μL sample loading volume

## Denaturation of Protein in Inclusion Bodies

HA-RBD protein can not be used for srtuctural biology in the form of inclusion bodies because of its insolubility and misfolding state. To obtain correctly folded HA-RBD, aggregated protein must be unfolded first by denaturation. Under optimal conditions, unfolded protein will refold into its native state.

HA-RBD inclusion bodies were divided into 4 Falcon tubes, each labeled as sample Aβ, A0, 0β, and 00 and each was denatured in 25 mL TBS-buffered urea solution at pH 8. Samples are divided based on the refolding buffers which would be used after denaturation process (See **FIG. 2.**).

**00**

5.86 mg/mL

**Aβ**

**5.55 mg/mLAβ**

5.55 mg/mL

**0β**

4.63 mg/mL

**A0**

5.07 mg/mL

**Aβ**

5.55 mg/mL

**0β**

4.63 mg/mL

**00**

5.86 mg/mL

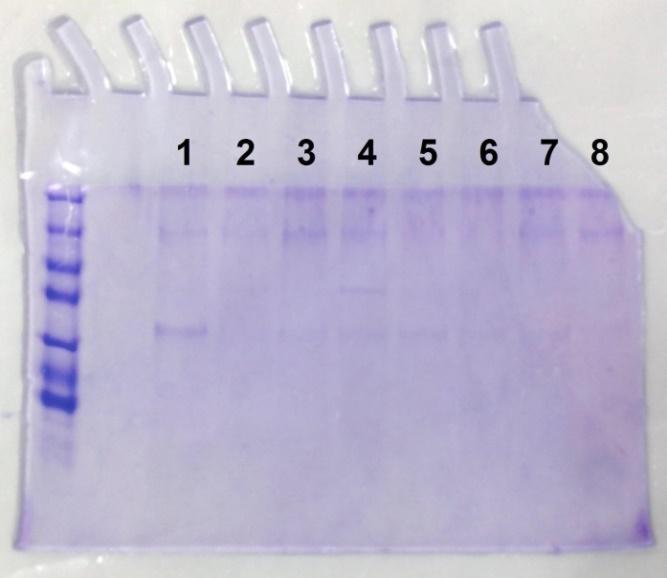
**A0**

5.07 mg/mL

**FIG. 8.** Concentration of denatured HA-RBD protein.

Protein denaturation makes use of a strong chaotropic agent, for example the urea used in this experiment. Urea disrupts inter-amide hydrogen bonds which shape alpha helices and beta sheets of secondary structure. In tertiary structure, urea disturbs hydrogen bonds in the side-chains. Primary structure remains undisturbed as it contains covalent bonds in disulfide bridges and between amino acids, which requires relatively high energy to cleave at room temperature.

## Screening of Protein Refolding Methods (Small Scale Experiments)



**FIG. 9.** SDS-PAGE gel of fast dialysis refolding samples. Based on **FIG. 2**.:

1 = 1AβT

2 = 1A0T

3 = 10βT

4 = 100T

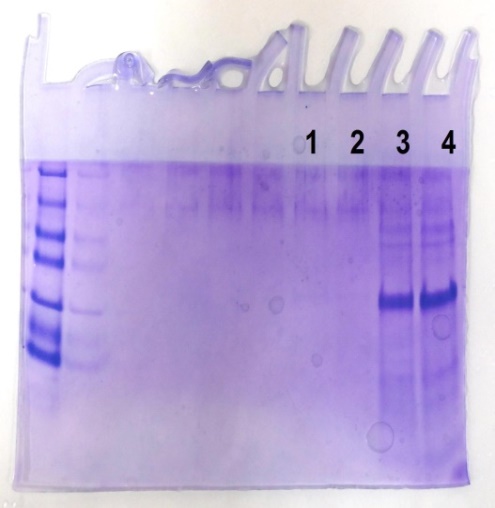
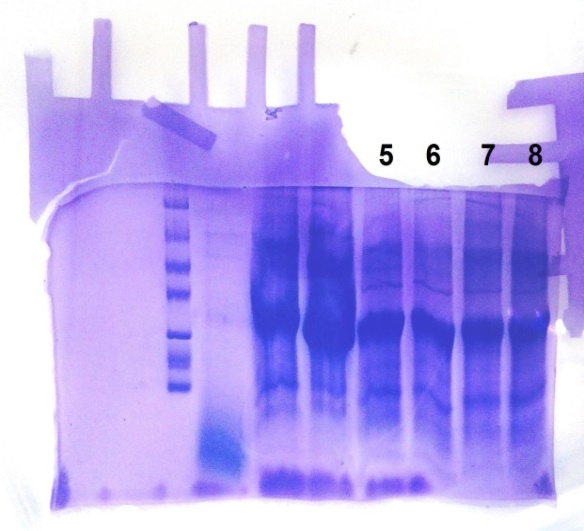
5 = 1AβP

6 = 1A0P

7 = 10βP

8 = 100P

**FIG. 10.** SDS-PAGE gel of slow dialysis refolding samples. Based on **FIG. 2.**:



1 = 2AβT

2 = 2A0T

3 = 20βT

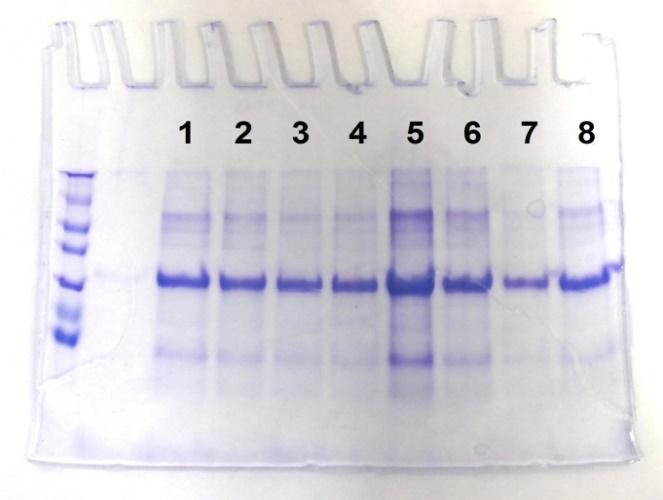
4 = 200T

5 = 2AβP

6 = 2A0P

7 = 20βP

8 = 200P



**FIG. 11.** SDS-PAGE gel of IMAC refolding samples. Based on **FIG. 2.**:

1 = 3AβT

2 = 3A0T

3 = 30βT

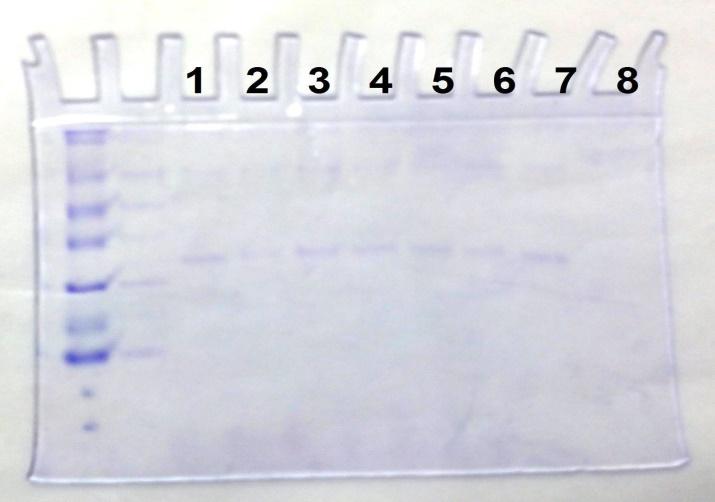
4 = 300T

5 = 3AβP

6 = 3A0P

7 = 30βP

8 = 300P



**FIG. 12.** SDS-PAGE gel of shock dilution refolding samples. Based on **FIG. 2.**:

1 = 4AβT

2 = 4A0T

3 = 40βT

4 = 400T

5 = 4AβP

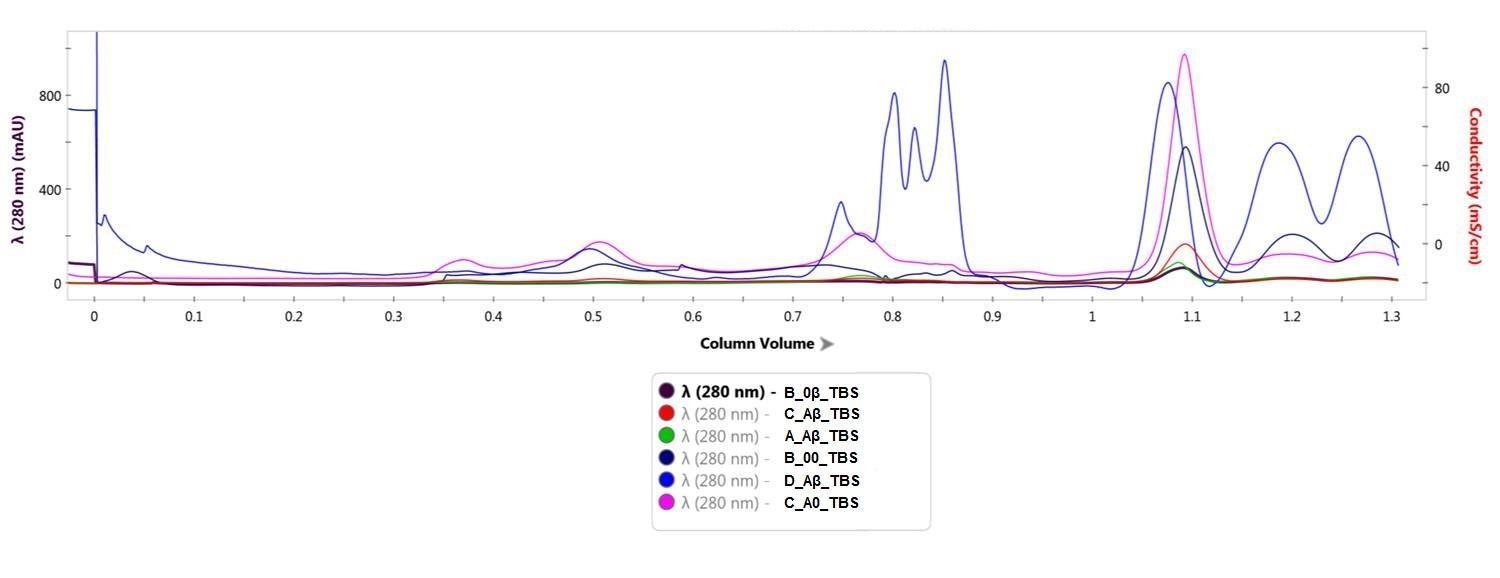
6 = 4A0P

7 = 40βP

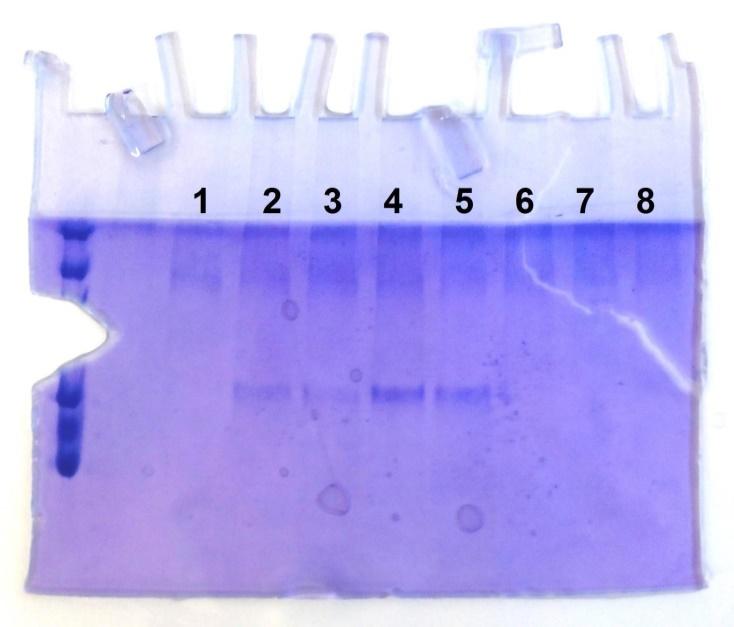
8 = 400P

All samples were not precipitated and based on the results of SDS-PAGE gel, most samples was successfully refolded. As there was no significant amount of urea in all refolded samples, the HA-RBD protein could be present as correctly folded soluble protein which could be detected from SDS-PAGE gel results or misfolded as aggregates which lead to precipitation. Samples 5, 6, 7, and 8 from SDS-PAGE gel of slow dialysis refolding method (**FIG. 9.**) appear to have significantly higher concentration of refolded protein. These samples were obtained with longer dialysis period (1 week) and continuous stirring. The SDS-PAGE gel showing the shock dilution refolding method (**FIG. 12**) is noticeably less intense than the other refolding methods as the refolded protein samples were not concentrated through Ni-NTA column.

As most samples were apparently successfully refolded, based on their solubility, (hence their appearances in SDS-PAGE gel), it should be determined whether the folded samples were also monomeric in solution. Samples from the following methods were run into size exclusion chromatography: fast dialysis 1AβT, slow dialysis 20βT and 200T, IMAC-assisted 3AβT and 3A0T, and shock dilution 4AβT. Each sample was run at flow rate of 0.5 mL/minute for 60 minutes.



**FIG. 13.** Chromatogram of protein refolding results showing protein peak around 0.5 column volume (fraction 17) with dimer aggregation peak around 0.35 column volume (fraction 7).



**FIG. 14.** SDS-PAGE gel of size exclusion chromatography fractions. Based on **FIG. 2.**:

1 = fraction 18 from shock dilution 4AβT

2 = fraction 11 from IMAC 3AβT

3 = fraction 12 from IMAC 3AβT

4 = fraction 11 from IMAC 3A0T

5 = fraction 12 from IMAC 3A0T

6 = fraction 17 from IMAC 3A0T

7 = fraction 18 from IMAC 3A0T

8 = fraction 25 from IMAC 3A0T

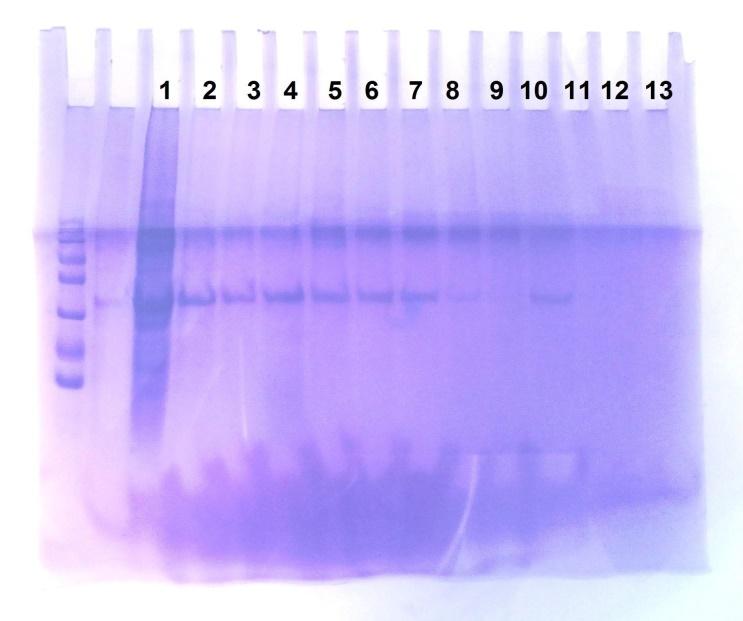
Based on the chromatogram result, the highest protein peak (around 0.5 column volume) was obtained from IMAC method using 3AβT buffer, followed by shock dilution method using 3AβT buffer. It should be noted that while the protein peak from IMAC method is significantly higher than shock dilution method, chromatogram result from IMAC method also shows significant peak at around 0.35 column volume, which was presumed to be a dimer or higher oligomeric aggregation peak. The chromatogram from shock dilution method does not have a noticeable high oligomeric peak, therefore it was expected that the refolded protein using shock dilution method would be most likely to be monomeric in solution.

## Scaling Up Protein Refolding (Large Scale Experiment)

HA-RBD protein used for the large scale experiment was obtained from a fresh bacteria culture. After denaturation, the unfolded protein has a concentration of 1.37 mg/mL.

1. **Scaling up with IMAC**





**FIG. 15.** SDS-PAGE of scale up with IMAC method.

1 = flow through

2, 3 = unfolding buffer washes

4, 5, 6, 7 = refolding buffer washes

8, 9 = elution buffer pH 8 washes

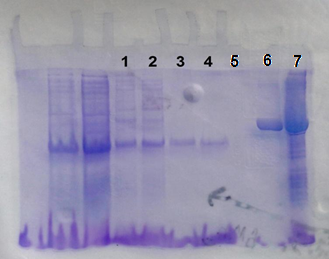
10 = elution buffer pH 4.5 wash

It was found that a large amount of HA-RBD was not bound in Ni2+ beads from the Ni-NTA column used due to initial equilibration using AβT buffer with 8 M urea (Tris-HCl 50 mM, 150 mM, 0.4 M arginine, and 5 mM ßME at pH 8). Arginine in the refolding buffer is not suitable for IMAC-assisted refolding as it has good affinity to the Ni2+ beads, thus can displace the bound protein.(1)(10)(28) According to Abe, R. *et al* (2009), this interference could be greatly reduced at low arginine concentration (below 0.2 M). Bound HA-RBD was also washed away using the same arginine-containing refolding buffer, leaving only a small amount of HA-RBD in the elutions.

In this experiment, total elution volume obtained was 40 mL then concentrated to 15 mL. Aggregation was seen and protein concentration was decreased to 0.117 mg/mL. After dialysis with dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8), sample was concentrated again to 6 mL and concentration was measured at 0.1 mg/mL.

1. **Scaling up with shock dilution**

For scaling up using shock dilution method, 103 mL of refolded protein sample was dialysed in 2 L dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8) and concentrated to 500 μL. The end concentration was measured at 1.5 mg/mL. This concentrated sample was further used for the thermofluor assay and crystallization.



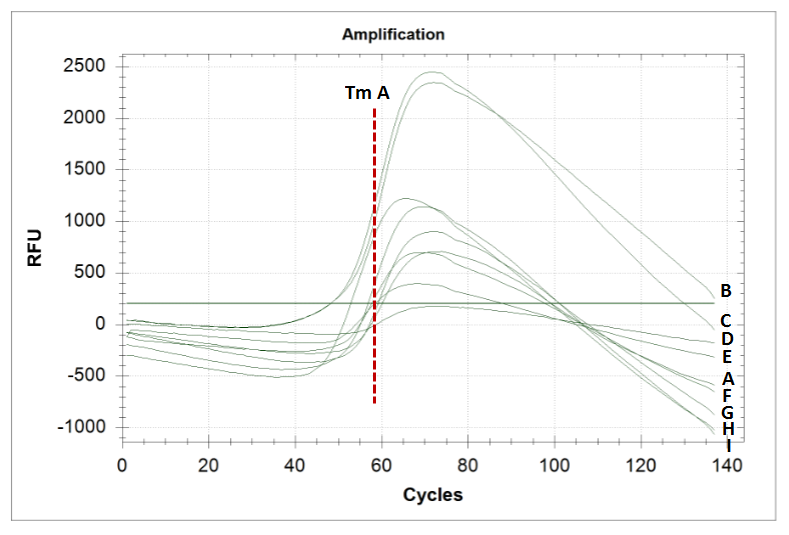
**FIG. 16.** SDS-PAGE of scale up with shock dilution method.

1, 2 = reference protein, size 25 kD, concentration around 2 mg/mL

3, 4 = refolded HA-RBD, concentrated to 16 mL from 103 mL

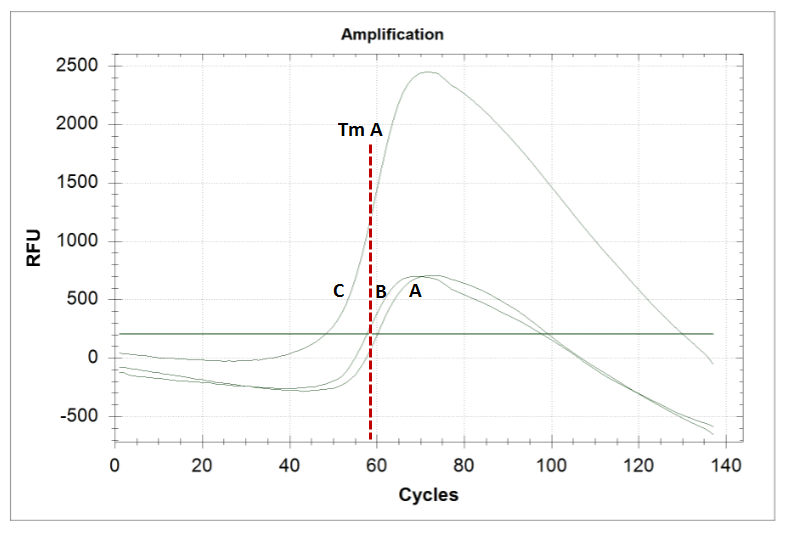
## Thermofluor Assay for Crystallization Buffer

Comparison of thermofluor stability curve in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8) and several buffer conditions (see **APPENDIX C**):



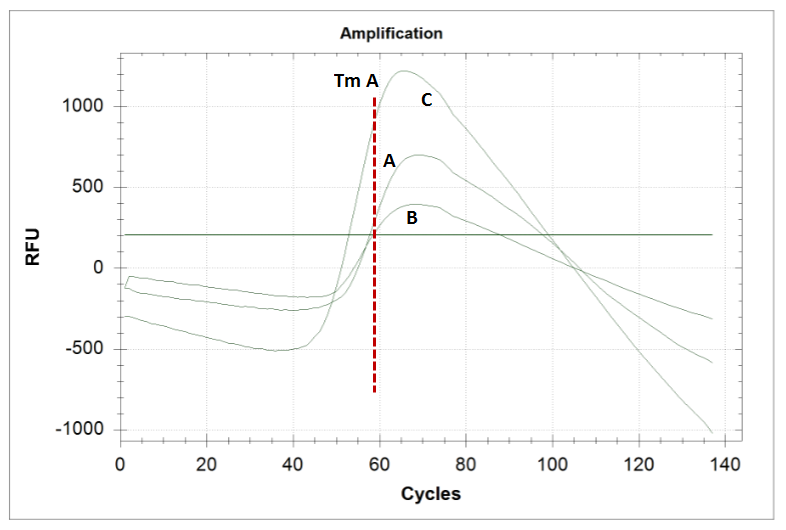
**FIG. 17.** Thermofluor results in different buffer conditions. (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in F1 buffer (100 mM Na-citrate). (C) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7). (D) Thermofluor curve for HA-RBD in D4 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8). (E) Thermofluor curve for HA-RBD in B5 buffer (60 mM Tris-HCl at pH 8.5). (F) Thermofluor curve for HA-RBD in C9 buffer (60 mM Na-phosphate and 250 mM NaCl at pH 7). (G) Thermofluor curve for HA-RBD in F2 buffer (100 mM Na-sulphate). (H) Thermofluor curve for HA-RBD in B4 buffer (60 mM Tris-HCl at pH 8). (I) Thermofluor curve for HA-RBD in D5 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8.5).

There are several conditions which provide better thermal stability for HA-RBD protein than the dialysis buffer, which was used in a previous study on HA-RBD crystallization by DuBois *et al.* (2011). To estimate the optimal condition for crystallization buffer, several parameters were used to compare HA-RBD thermal stability: pH, salt, metal presence, and DMSO presence.



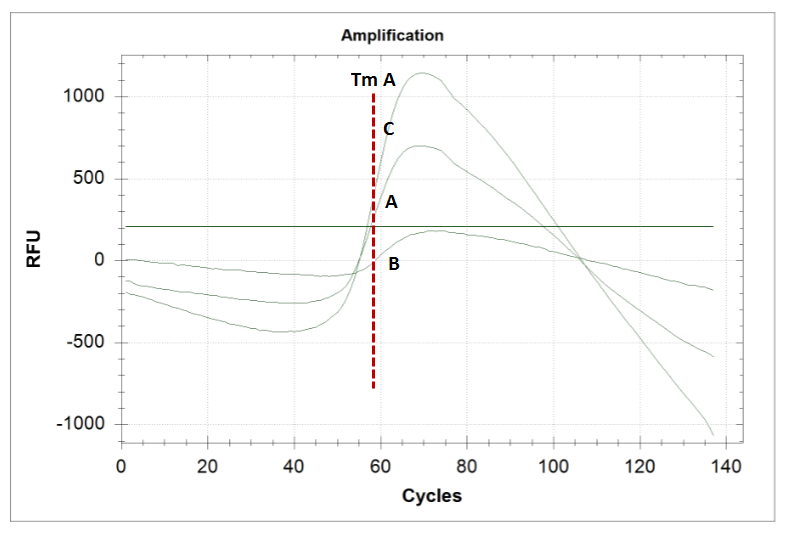
**FIG. 18.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in C9 buffer (60 mM Na-phosphate and 250 mM NaCl at pH 7). (C) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7).

**FIG. 18** above shows HA-RBD thermal stability is improved at pH 8 with respect to pH 7 and may be improved by the addition of increased concentrations of NaCl.



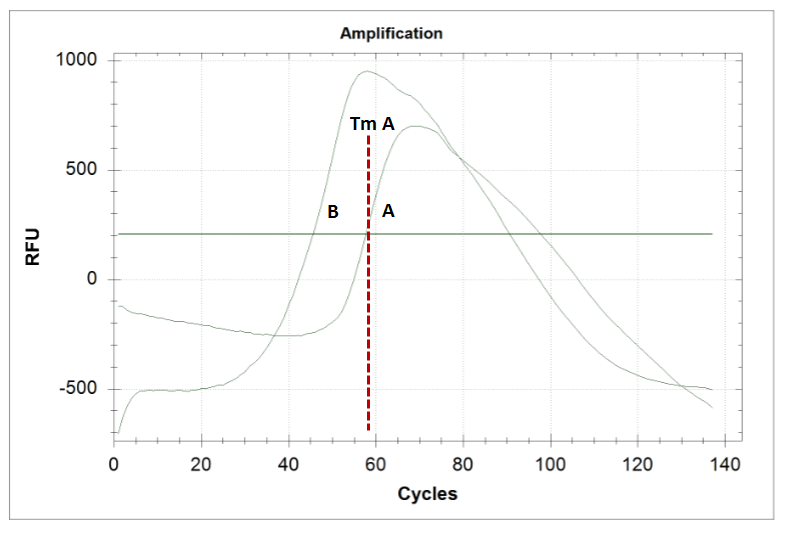
**FIG. 19.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in B5 buffer (60 mM Tris-HCl at pH 8.5). (C) Thermofluor curve for HA-RBD in B4 buffer (60 mM Tris-HCl at pH 8).

**FIG. 19** above shows HA-RBD thermal stability is improved by increased concentrations of NaCl.



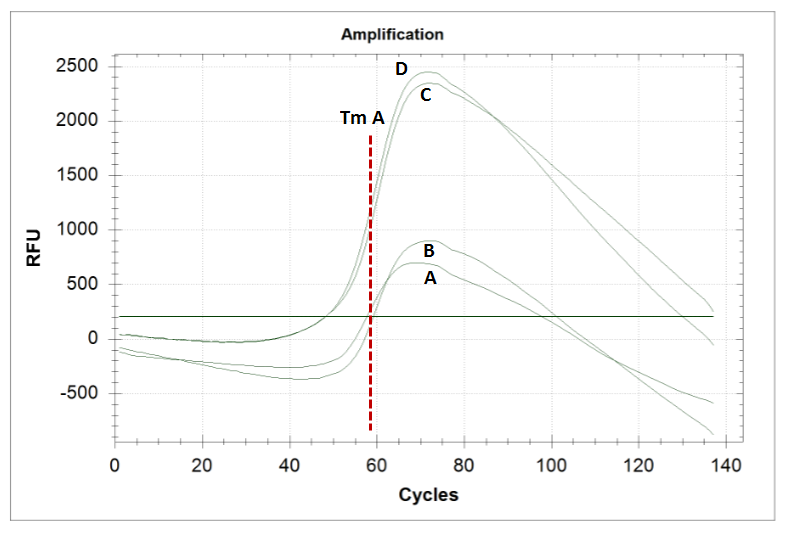
**FIG. 20.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in D4 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8). (C) Thermofluor curve for HA-RBD in D5 buffer (60 mM Tris-HCl and 250 mM NaCl at pH 8.5).

**FIG. 20** above shows HA-RBD thermal stability is improved by a slightly higher concentration of NaCl, 250 mM to 100 mM.



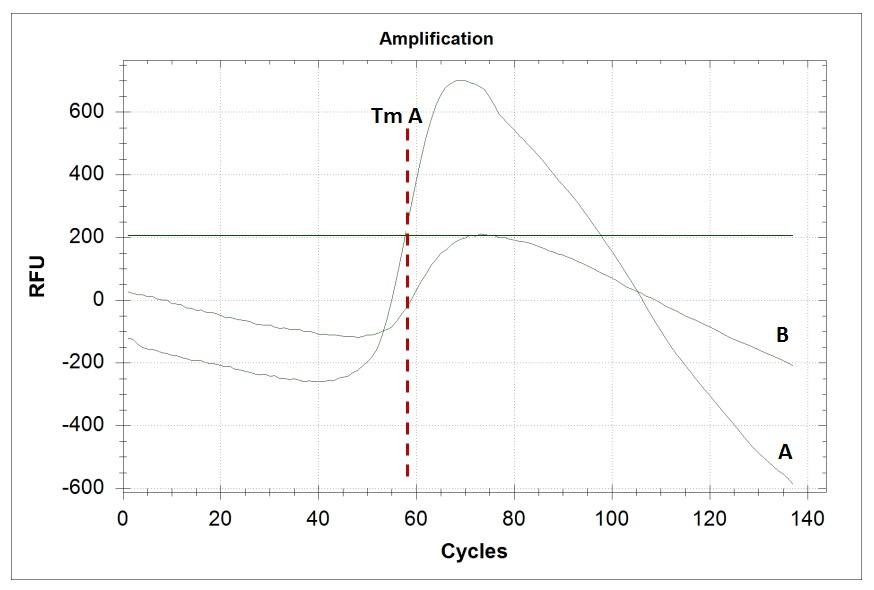
**FIG. 21.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in B8 buffer (10% DMSO, 10 mM Tris-HCl, and 100 mM NaCl at pH 8).

**FIG. 21.** above shows HA-RBD thermal stability is significantly decreased by the addition of 10% DMSO.



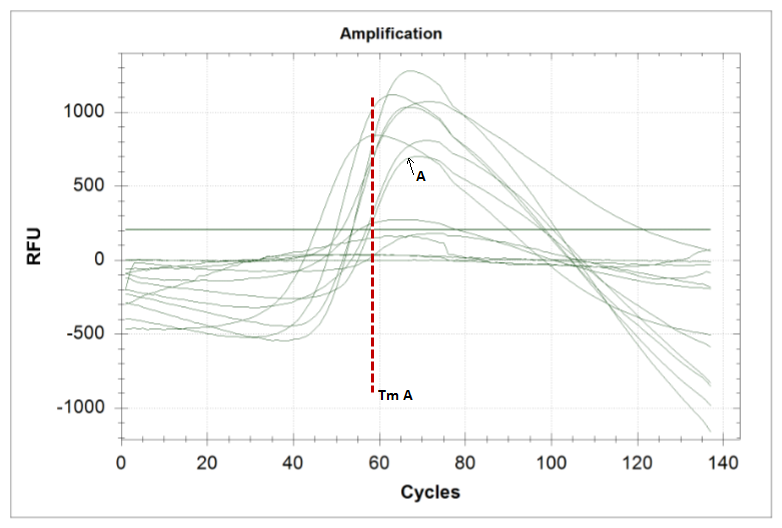
**FIG. 22.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in F2 buffer (100 mM Na-sulphate). (C) Thermofluor curve for HA-RBD in F1 buffer (100 mM Na-citrate). (D) Thermofluor curve for HA-RBD in A9 buffer (60 mM Na-phosphate at pH 7).

**FIG. 22** above shows HA-RBD thermal stability is improved best in Na-citrate and followed by Na-sulphate and Na-phosphate buffers.



**FIG. 23.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. (B) Thermofluor curve for HA-RBD in D1 buffer (60 mM HEPES and 250 mM NaCl at pH 7.5).

**FIG. 23** above shows HA-RBD thermal stability is improved in HEPES buffer.



**FIG. 24.** (A) Thermofluor curve for HA-RBD in dialysis buffer (10 mM Tris-HCl and 100 mM NaCl at pH 8). (Tm A) Approximate inflection point of curve A. Other curves represents stability of HA-RBD in several metal-containing buffers. See **APPENDIX C** for a complete list of buffer solutions used.

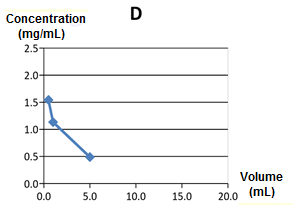
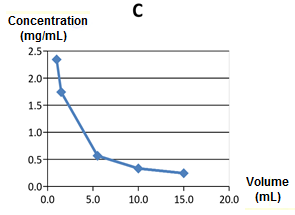
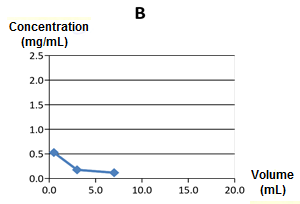
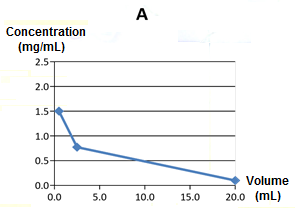
**FIG. 24** above shows HA-RBD thermal stability is decreased by the presence of all metals used in the assay. In summary, the thermofluor suggests that the following conditions can improve HA-RBD thermal stability: pH 8, slightly higher concentration of NaCl than in the current dialysis buffer (100 mM), addition of Na-citrate and use of HEPES as a buffer. Based on these results, two thermofluor (DSF) buffers was selected, namely DSF 1 (10 mM Tris-HCl, 200 mM NaCl, 10 mM Na-citrate at pH 8) and DSF2 (10 mM HEPES, 200 mM NaCl, 10 mM Na-citrate at pH 7.5).

## Scaling Up Protein Refolding (Large Scale Experiment) with Selected Thermofluor (DSF) Buffers

For scaling up using new thermofluor (DSF) buffers, a fresh batch of unfolded protein sample was used. The protein concentration from this new batch is 2.73 mg/mL. 3 mL of unfolded protein sample was diluted in each thermofluor buffer (100 mL) and concentrated at 4oC. The concentrated protein solution was measured at different volume points.

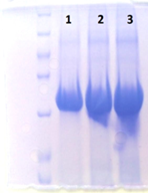
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **A** | | | **B** | | |
| Volume (mL) | Abs. | Conc (mg/ml). | Volume (mL) | Abs. | Conc (mg/ml). |
| 20.00 | 0.01 | 0.10 | 7.00 | 0.01 | 0.12 |
| 2.50 | 0.08 | 0.77 | 3.00 | 0.02 | 0.18 |
| 0.50 | 0.15 | 1.50 | 0.50 | 0.05 | 0.53 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **C** | | | **D** | | |
| Volume (mL) | Abs. | Conc (mg/ml). | Volume (mL) | Abs. | Conc (mg/ml). |
| 15.00 | 0.03 | 0.24 | 5.00 | 0.05 | 0.49 |
| 10.00 | 0.03 | 0.33 | 1.00 | 0.12 | 1.13 |
| 5.50 | 0.06 | 0.57 | 0.50 | 0.16 | 1.54 |
| 1.50 | 0.18 | 1.74 |  |  |  |
| 1.00 | 0.24 | 2.34 |  |  |  |

**FIG. 25.** Measured concentration of protein at different volume points. (A) 3 mL of unfolded protein diluted in 100 mL AßT refolding buffer (see Appendix A) then dialysed into standard dialysis buffer (10 mM Tris-HCl, 100 mM NaCl at pH 8) then concentrated. (B) 3 mL of unfolded protein diluted in 100 mL AßT refolding buffer then dialysed into DSF1 dialysis buffer then concentrated. (C) 3 mL of unfolded protein diluted in 100 mL standard refolding buffer then dialysed into DSF2 dialysis buffer then concentrated. (D) 3 mL of unfolded protein diluted in 100 mL DSF2 buffer then dialysed into DSF2 buffer then concentrated.



**FIG. 26.** Measured concentration of protein at different volume points (in graph). Sample A, B, C, and D correspond to the samples in **FIG. 17.**

Sample C was further concentrated until it reached 5.6 mg/mL and analysed on an SDS-PAGE gel. This analysis was necessary to follow the behaviour of HA-RBD during concentrations – as refolded protein is prone to precipitation and careful monitoring of the protein concentration is requred in order not to over concentrate the material.



**FIG. 27.** Refolded HA-RBD, concentration 5.6 mg/mL.

1 = sample volume 5 μL

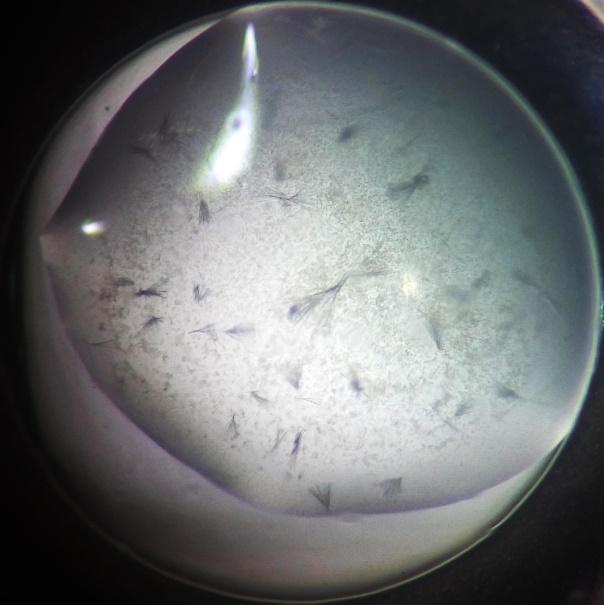
2 = sample volume 10 μL

3 = sample volume 15 μL

This result demonstrates that we have successfully obtained HA-RBD at a concentration high enough for crystallisation and that the sample is highly pure (as determined by SDS-PAGE).

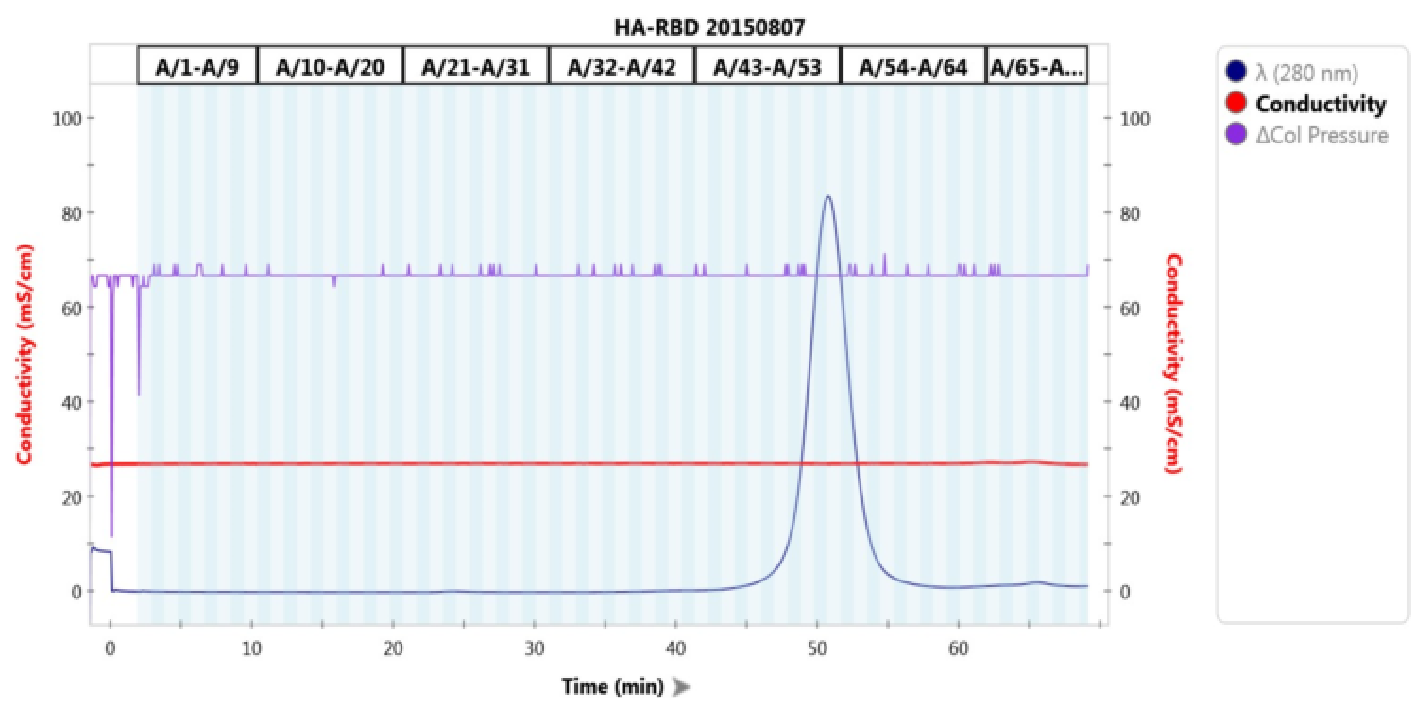
## Crystallization Screening with Sitting Drop Method

After 48 hours incubation, crystal was observed in condition D5 (see **FIG. 3.**) which contains 30% PEG at pH 8.5.



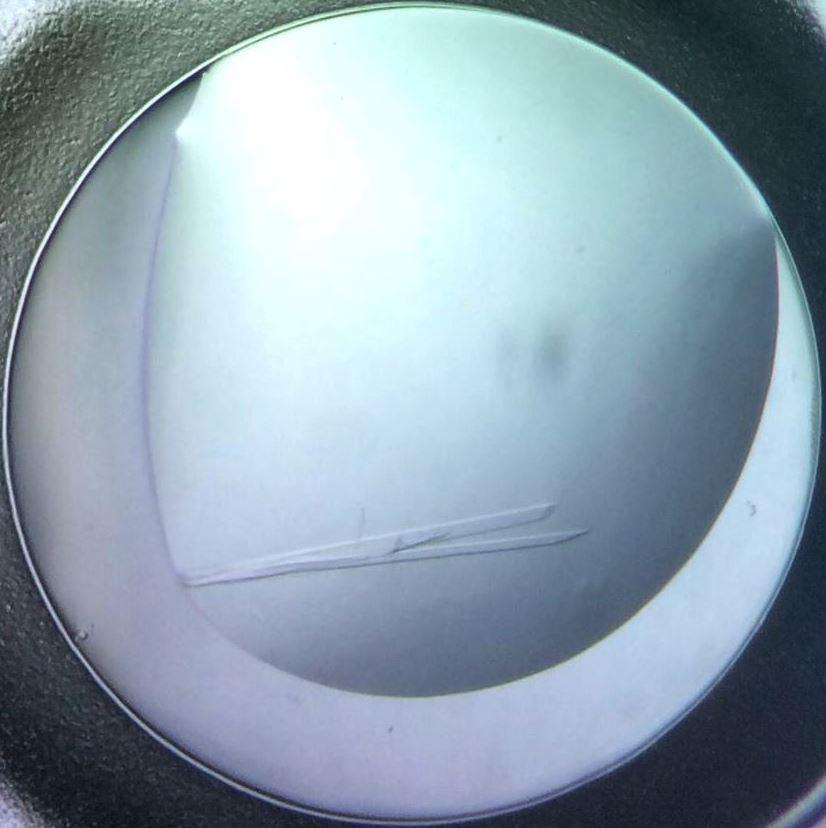
**FIG. 28.** HA-RBD crystals from sitting drop method.

As seen in **FIG. 28**, the HA-RBD crystals obtained from the experiment were very thin and not optimal for small compound soaking experiment. It is possible that misfolded proteins were also still present in the dissolved sample, which can interfere with the crystallization process. To optimize crystallization, refolded HA-RBD sample was further purified by size exclusion chromatography.



**FIG. 29.** Peak profile of refolded HA-RBD sample used in the crystallization screening in **FIG. 28**

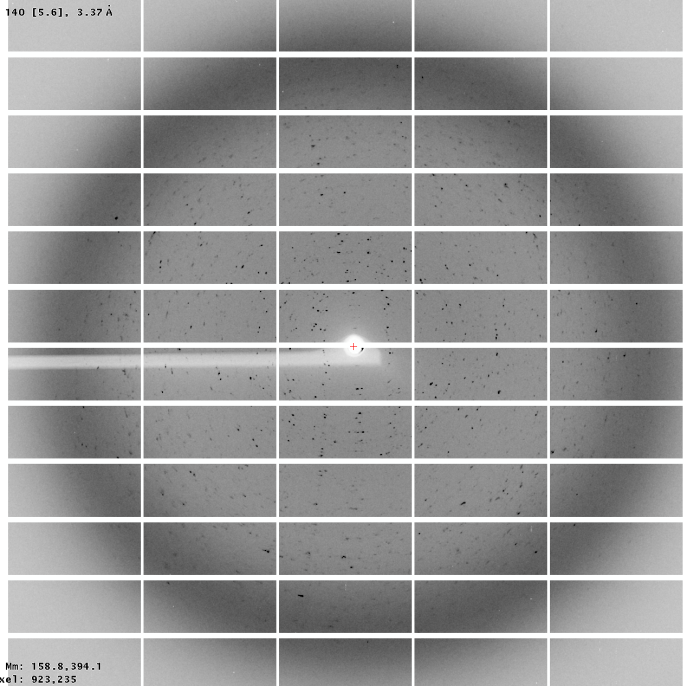
The sample obtained from peak area (around minute 50 in **FIG. 29**) was used for crystallization optimization experiment with several conditions based on the result of previous crystallization attempt (see **FIG. 4** for conditions).



**FIG. 30.** HA-RBD protein crystal after purification by size exclusion chromatography.

From **FIG. 30**, it can be seen that the crystal result from the optimization experiment (well 7C from **FIG. 4**) was a single, large crystal which is optimal to be used for diffraction experiment on an X-ray beamline.

## Crystal Diffraction



**FIG. 31.** Electron density map of protein crystal diffraction. Left image shows the protein as multiple crystals and the right image shows the protein as a single crystal.

|  |  |
| --- | --- |
| Diffraction source | Beamline P11, PETRA III, DESY Hamburg |
| Wavelength ‎(Å) | 0.97267 |
| Temperature (K) | 100 |
| Detector | PILATUS |
| Crystal-to-detector distance (mm) | 398.89 |
| Beam size (µm) | 20 |
| Rotation range per image (o) | 0.1 |
| Number of images | 2200 |
| Exposure time per image (ms) | 10 |
| Space group | P2(1) |
| *a, b. c* ‎(Å) | 39, 70, 70 |
| *α, β, γ* (o) | 90, 106, 90 |
| Multiplicity | 3.73 |
| Mosaicity (o) | 0.198 |
| Resolution range ‎(Å) | 38.86-1.74 |
| Total number of reflection | 141575 |
| Number of unique reflections | 37935 |
| Completeness (%) | 89.5 |
| *(I/σ(I))* | 11.58 |
| Rmeas (%) | 9.2 |
| Overall B factor from Wilson plot ‎(Å2) | 31.36 |

**FIG. 32.** Diffraction statistics for HA-RBD protein.

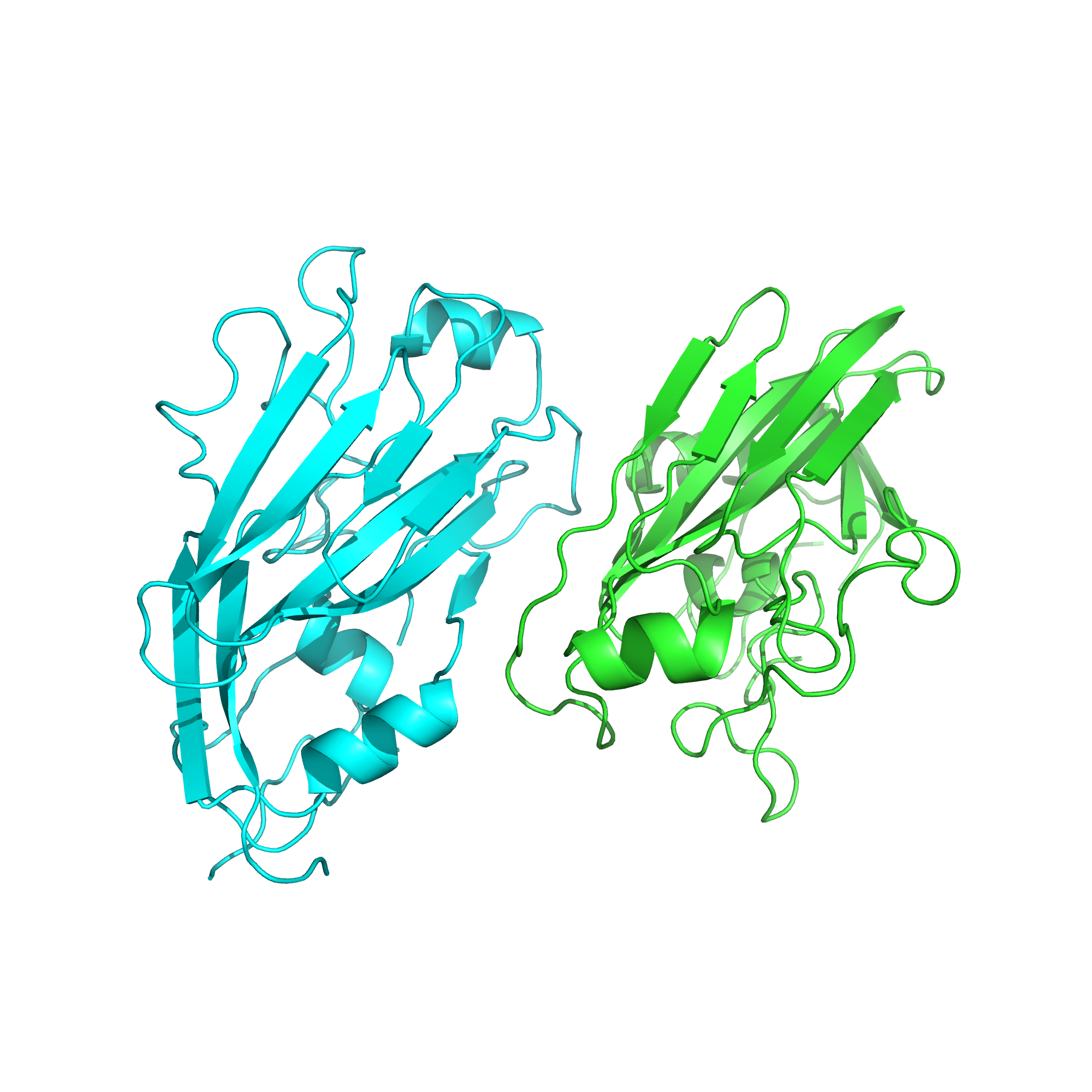
After analysis by XDS,(19) the structure was solved by molecular replacement using the Molrep package in CCP4 program. This demonstrates that we have successful produced high-resolution diffracting crystals of HA-RBD.

## Crystallization Screening with Dimethyl Sulfoxide (DMSO)

While it was found that stability of HA-RBD protein crystal was significantly decreased with the addition of 15% v/v DMSO. HA-RBD crystals are apparently stable in 10% DMSO, as the crystals did not show any noticeable change in difrraction quality after an overnight incubation in 10% DMSO.

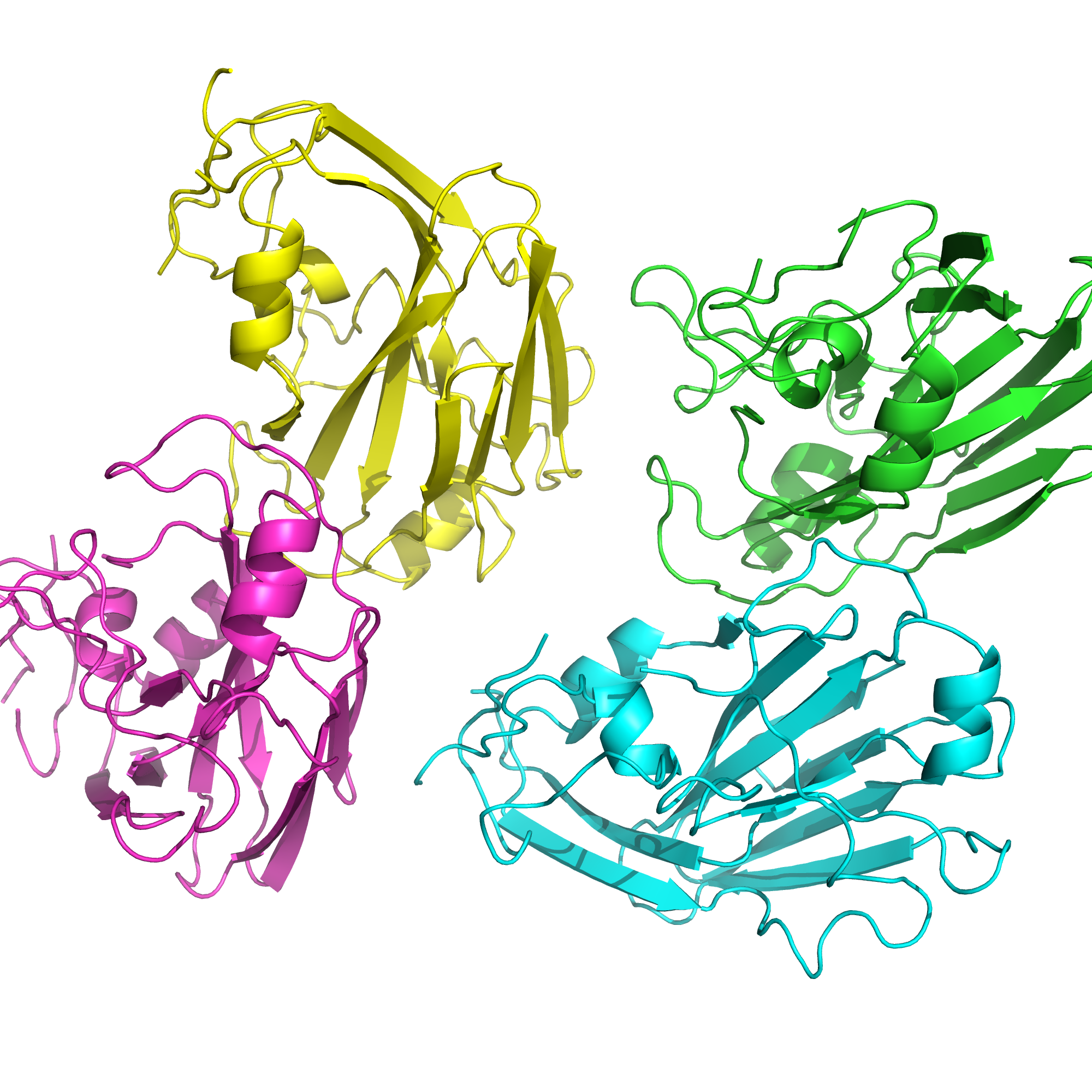
## Compound Soaking Experiment

The diffraction result of the crystal soaked in 2mM compound (TR-326, condition X in **FIG. 6**) was used to generate protein structure using XDS software. (19) As a starting model, the structure of HA-RBD protein (RCSB: 3MLH) was used to solve the structure by molecular replacement. Comparison between the observed structure from this experiment and HA-RBD model shows which part is only observed in the experiment, which is an early sign of the compound interaction with the HA-RBD protein.



**FIG. 33.** Structure solution of HA-RBD crystal soaked in small compound at low concentration.

Protein structure obtained from the higher concentration compound-soaking experiment was found to have a different molecular arrangement from the protein structure before being soaked in small compound solution.



**FIG. 34.** Structure solution of HA-RBD crystal soaked in small compound in high concentration.

In high concentration (**FIG. 34**), the change in molecular arrangement is even more visible. Although the small compound itself is not visible in the generated structure due to its small size, the changing in the orientation signifies an interaction between the small compound and HA-RBD protein, suggesting binding. As 1.36 M glycerol is used in the cryopreservation buffer and is known to bind to the sialic acid binding site, it is likely that the significant excess of glycerol will compete with the compound, which is present at 2 mM to bind to the HA-RBD protein.

# Conclusions

HA-RBD protein was expressed from *E. coli* as insoluble inclusion bodies. As inclusion bodies contain insoluble and misfolded HA-RBD, the protein must be unfolded first by denaturation with 8 M urea. To refold the protein back into its native state, several refolding methods were tested in small scale: fast dialysis, slow dialysis, IMAC-assisted refolding, and shock dilution. Additives arginine and βME were also added in the refolding process.Most samples were successfully refolded but based on the size exclusion chromatography peak profile, the best samples are obtained from IMAC-assisted refolding and shock dilution methods. However, the chromatogram result from IMAC-assisted refolding method also shows a dimer aggregation peak therefore the shock dilution method was used for larger scale experiments. It was also found that the IMAC-assisted refolding method was not suitable for use as a large amount of protein was not bound in Ni2+ beads due to interference bu the high concentrations of arginine.An optimal buffer condition for purification and crystallization was determined based on the protein thermal stability. Based on pH, salt, metal presence, and DMSO presence parameters, the buffer chosen for crystallization has a composition of 10 mM Tris-HCl, 200 mM NaCl, 10 mM Na-citrate at pH 8. Using this buffer, the protein was concentrated to 5.6 mg/mL. After 48 hours, thin crystals were observed in a solution containing 30% PEG at pH 8.5. The refolded HA-RBD sample was then purified by size exclusion chromatography and used in optimization crystallization experiments. A single, large crystal was observed in a solution containing 34% PEG-2000 at pH 8.8. The diffraction result of this crystal shows an electron diffraction pattern of a single crystal.

As the refolded HA-RBD is going to be co-crystallized with small molecules dissolved in DMSO, the stability of HA-RBD crystal in DMSO needed to be determined. After addition of 10% DMSO, the crystal did not show observable significant change. However, after addition of 15% DMSO, HA-RBD crystal was significantly damaged. After this optimization, HA-RBD crystal was successfully used in compound soaking experiments. The structure obtained from the compound soaking experiment shows a change in molecular arrangement suggesting an interaction between the compound and the HA-RBD protein. Further experiments are currently being performed in which the glycerol will be completely omitted from the purification and crystal cryopreservation. We are confident that this will enable us to visualize the mode of binding of the compound with HA-RBD and provide data for further rounds of optimization of a novel candidate for the treatment of influenza in humans.

# References

1. Abe, R. et al., 2009. Immobilized metal affinity chromatography in the presence of arginine. *Biochemical and Biophysical Research Communications*, 381(3), pp. 306–310.
2. Aguilar-Yáñez, J. M. et al., 2010. An Influenza A/H1N1/2009 hemagglutinin vaccine produced in Escherichia coli. *PLoS ONE*, 5(7), pp. 8–12.
3. Almarza, J. et al., 2009. Molecular mechanism for the denaturation of proteins by urea. *Biochemistry*, 48(32), pp. 7608–7613.
4. Baudin, F. et al., 1994. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *The EMBO Journal*, 13(13), pp. 3158–3165.
5. Bennion, B. J. & Daggett, V., 2003. The molecular basis for the chemical denaturation of proteins by urea. *Proceedings of the National Academy of Sciences,* 100(9), pp. 5142–5147.
6. Berman, H. M. et al., 2000. The Protein Data Bank. *Nucleic Acids Research*, 28, pp. 235–242.
7. Bertani, G., 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. *Journal of Bacteriology*, 186(3), pp. 595–600.
8. Burch, J., Corbett, M., Stock, C., et al. 2009. Prescription of anti-influenza drugs for healthy adults: a systematic review and meta-analysis. *Lancet Infectious Disease*, 9(9), pp. 537–545.
9. Burgess, R. R., 2009. Chapter 17 Refolding Solubilized Inclusion Body Proteins. *Methods in Enzymology*, 463, pp. 259–282.
10. Chen, J. et al., 2008. Different Effects of L -Arginine on Protein Refolding : Suppressing Aggregates of Hydrophobic Interaction, Not Covalent Binding. *Biotechnology Progress*, 24, pp. 1365–1372.
11. Collaborative Computational Project Number 4, 1994. The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallographica*, D50, pp. 760-763.
12. Das, A. & Mukhopadhyay, C., 2009. Urea-mediated protein denaturation: A consensus view. *Journal of Physical Chemistry B*, 113(38), pp. 12816–12824.
13. Das, K., 2012. Antivirals targeting influenza A virus. *Journal of Medicinal Chemistry*, 55(14), pp. 6263–6277.
14. Deng, T. et al., 2005. In vitro assembly of PB2 with a PB1-PA dimer supports a new model of assembly of influenza A virus polymerase subunits into a functional trimeric complex. *Journal of Virology*, 79(13), pp. 8669–8674.
15. DiPiro, Joseph. et al., 2008. Pharmacotherapy: A Pathophysiologic Approach. 7 Edition. McGraw-Hill Medical, pp. 1791–1793.
16. Dong, G. et al., 2015. Adamantane-Resistant Influenza A Viruses in the World (1902 – 2013): Frequency and Distribution of M2 Gene Mutations. *PLoS ONE*, 10(3), pp. 1-20.
17. DuBois, R.M. et al., 2011. The receptor-binding domain of influenza virus hemagglutinin produced in Escherichia coli folds into its native, immunogenic structure. *Journal of Virology*, 85(2), pp. 865–872.
18. He, F., 2011. Laemmli-SDS-PAGE. *Bio-protocol* Bio101: e80, http://www.bio-protocol.org/e80.
19. Kabsch, W., 2010. XDS. *Acta Crystallographica,*D66, pp. 125-132.
20. Liu, J. et al., 2009. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proceedings of the National Academy of Sciences of the United States of America*, 106(40), pp. 17175–17180.
21. Liu, J., 1985. Process for preparing rimantadine. United States.
22. Matsuoka, Y. et al., 2013. A comprehensive map of the influenza A virus replication cycle. *BMC Systems Biology*, 7(1), pp. 97.
23. Niesen, F. et al., 2007. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *National Protocol of Structural Genomics Consortium Botnar Research Centre*, 2(9), pp. 2212-21.
24. Nicholson, K. G. et al., 1998. *Textbook of Influenza*, Oxford: Wiley-Blackwell, pp. 46 – 47, 55.
25. Novagen, 1997. Protein refolding kit manual. *Novagen*, pp. 7–9.
26. Palmer, I. & Wingfield, P. T., 2004. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from *Escherichia coli*. *Current Protocols in Protein Science*, 38(6.3), unit 6.3.1–6.3.18.
27. Prichard, W. W., 1967. Adamantanes and tricyclo[4. 3. 1. 1 3.8] undecanes
28. Qiagen, 2008. Ni-NTA spin kit handbook. *Qiagen*, pp. 11–12, 27, 31.
29. Rossky, P. J., 2008. Protein denaturation by urea: slash and bond. *Proceedings of the National Academy of Sciences of the United States of America*, 105(44), pp. 16825–16826.
30. Samji, T., 2009. Influenza A: Understanding the viral life cycle. *Yale Journal of Biology and Medicine*, 82(4), pp. 153–159.
31. Santos, C. et al., 2012. A novel protein refolding protocol for the solubilization and purification of recombinant peptidoglycan-associated lipoprotein from Xylella fastidiosa overexpressed in Escherichia coli. *Protein Expression and Purification*, 82(2), pp. 284–289.
32. Schuldt, L. & Weiss, M. S., 2010. Stabilization of proteins for crystallization – how Thermofluor can help. *Biological Strucutres*, pp. 3–4.
33. Sigma-Aldritch, 2013. SYPRO®Orange Protein Gel Stain Product Information. *Sigma-Aldritch*. *Co. LLC,* pp. 1.
34. Singh, S. M. & Panda, A. K., 2005. Solubilization and refolding of bacterial inclusion body proteins. *Journal of Bioscience and Bioengineering*, 99(4), pp. 303–310.
35. Steinle, A. et al., 2001. Interaction of human NKG2D with its ligands MICA, MICB, and homologs of the mouse RAE-1 protein family. *Immunogenetics*, 53, pp. 279–287.
36. Stephenson, I. et al., 2009. Neuraminidase Inhibitor Resistance after Oseltamivir Treatment of Acute Influenza A and B in Children. *Clinical Infectious Diseases*, 48 (4), pp. 389–396.
37. Stevens, J. & Donis, R. O., 2007. Influenza virus hemagglutinin--structural studies and their implications for the development of therapeutic approaches. *Infectious Disorders Drug Targets*, 7(4), pp. 329–335.
38. Thomson, C. et al., 2012. A Simplified Method for the Efficient Refolding and Purification of Recombinant Human GM-CSF. *PLoS ONE*, 7(11), pp. 1–6.
39. Vagin, A., 1989. New translation and packing functions. *Newsletter on Protein Crystallography Daresbury Laboratory*, 24, pp. 117-121.
40. Wilson, I. A. et al., 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature*, 289(5796), pp. 366–373.
41. Wlodawer, A. et al., 2008. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS Journal*, 275(1), pp. 1–21.
42. Xu, R. et al., 2010. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science*, 328(5976), pp. 357–360.
43. Zambon, M. C., 1999. Epidemiology and pathogenesis of influenza. *Journal of Antimicrobial Chemotherapy*, 44(TOPIC B), pp. 3–9.

# APPENDIX A: HA-RBD Gene for Protein Expression

1 mkailvvlly tfatanadtl cigyhannst dtvdtvlekn vtvthsvnll edkhngklck

61 lrgvaplhlg kcniagwilg npeceslsta sswsyivets ssdngtcypg dfidyeelre

121 qlssvssfer feifpktssw pnhdsnkgvt aacphagaks fyknliwlvk kgnsypklsk

181 syindkgkev lvlwgihhps tsadqqslyq nadayvfvgs sryskkfkpe iairpkvrdq

241 egrmnyywtl vepgdkitfe atgnlvvpry afamernags giiisdtpvh dcnttcqtpk

301 gaintslpfq nihpitigkc pkyvkstklr latglrnvps iqsrglfgai agfieggwtg

361 mvdgwygyhh qneqgsgyaa dlkstqnaid eitnkvnsvi ekmntqftav gkefnhlekr

421 ienlnkkvdd gfldiwtyna ellvllener tldyhdsnvk nlyekvrsql knnakeigng

481 cfefyhkcdn tcmesvkngt ydypkyseea klnreeidgv klestriyqi laiystvass

541 lvlvvslgai sfwmcsngsl qcrici

Source: GenBank accession number ACQ99608.

# APPENDIX B: Composition of Refolding Buffers in Small Scale Refolding

AßT : Tris-HCl 50 mM, 150 mM, 0.4 M arginine, and 5 mM ßME at pH 8

AßP : NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, KH2PO4 1.8 mM, 0.4 M arginine, and 5 mM ßME at pH 8

A0T : Tris-HCl 50 mM, 150 mM, and 0.4 M arginine at pH 8

A0P : NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, KH2PO4 1.8 mM, and 0.4 M arginine at pH 8

0ßT : Tris-HCl 50 mM, 150 mM, and 5 mM ßME at pH 8

0ßP : NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, KH2PO4 1.8 mM, and 5 mM ßME at pH 8

00T : Tris-HCl 50 mM and 150 mM at pH 8

00P : NaCl 137 mM, KCl 2.7 mM, Na2HPO4 10 mM, and KH2PO4 1.8 mM at pH 8

# APPENDIX C: Thermofluor Buffer Conditions

|  |  |  |
| --- | --- | --- |
|  | Buffer Composition | Salt |
| A1 | 60 mM Citric acid pH 4.0 |  |
| A2 | 60 mM Na-acetate pH 4.4 |  |
| A3 | 60 mM Na-citrate pH 5.0 |  |
| A4 | 60 mM Na-citrate pH 5.5 |  |
| A5 | 60 mM Sodium phosphate pH 6.0 |  |
| A6 | 60 mM MES pH 6.2 |  |
| A7 | 60 mM MES pH 6.5 |  |
| A8 | 60 mM PIPES pH 6.7 |  |
| A9 | 60 mM Sodium phosphate pH 7.0 |  |
| A10 | 60 mM HEPES pH 7.0 |  |
| A11 | 60 mM MOPS pH 7.0 |  |
| A12 | 60 mM Ammonium acetate pH 7.3 |  |
| B1 | 60 mM HEPES pH 7.5 |  |
| B2 | 60 mM Imidazol pH 7.5 |  |
| B3 | 60 mM Imidazol pH 8.0 |  |
| B4 | 60 mM Tris-HCl pH 8.0 |  |
| B5 | 60 mM Tris-HCl pH 8.5 |  |
| B6 | 10 mM Tris-HCl pH 8 | 100 mM NaCl |
| B7 | 60 mM bis-Tris pH 6.0 |  |
| B8 | 10% DMSO, 10 mM Tris-HCl pH 8 | 100 mM NaCl |
| B9 | 5% DMSO |  |
| B10 | 10% DMSO |  |
| B11 | 15% DMSO |  |
| B12 | 20% DMSO |  |
| C1 | 60 mM Citric acid pH 4.0 | 250 mM NaCl |
| C2 | 60 mM Na-acetate pH 4.4 | 250 mM NaCl |
| C3 | 60 mM Na-citrate pH 5.0 | 250 mM NaCl |
| C4 | 60 mM Na-citrate pH 5.5 | 250 mM NaCl |
| C5 | 60 mM Sodium phosphate pH 6.0 | 250 mM NaCl |
| C6 | 60 mM MES pH 6.2 | 250 mM NaCl |
| C7 | 60 mM MES pH 6.5 | 250 mM NaCl |
| C8 | 60 mM PIPES pH 6.7 | 250 mM NaCl |
| C9 | 60 mM Sodium phosphate pH 7.0 | 250 mM NaCl |
| C10 | 60 mM HEPES pH 7.0 | 250 mM NaCl |
| C11 | 60 mM MOPS pH 7.0 | 250 mM NaCl |
| C12 | 60 mM Ammonium acetate pH 7.3 | 250 mM NaCl |
| D1 | 60 mM HEPES pH 7.5 | 250 mM NaCl |
| D2 | 60 mM Imidazol pH 7.5 | 250 mM NaCl |
| D3 | 60 mM Imidazol pH 8.0 | 250 mM NaCl |
| D4 | 60 mM Tris-HCl pH 8.0 | 250 mM NaCl |
| D5 | 60 mM Tris-HCl pH 8.5 | 250 mM NaCl |
| D6 | water |  |
| D7 | 60 Mm bis-Tris pH 6.0 | 250 mM NaCl |
| D8 | 0.1 M H3BO3 |  |
| D9 | 0.1 M CuCl2 |  |
| D10 | 0.1 M Na2MoO4 |  |
| D11 | 5% DMSO |  |
| D12 | water |  |
| E1 | 100 mM Ammonium chloride |  |
| E2 | 100 mM Rubidium chloride |  |
| E3 | 100 mM Potassium chloride |  |
| E4 | 100 mM Sodium chloride |  |
| E5 | 100 mM Lithium chloride |  |
| E6 | 100 mM Magnesium chloride |  |
| E7 | 100 mM Calcium chloride |  |
| E8 | 100 mM Manganese chloride |  |
| E9 | 100 mM Nickel chloride |  |
| E10 | 100 mM Cobalt chloride |  |
| E11 | 100 Mm Iron (III) chloride |  |
| E12 | 100 mM Zinc chloride |  |
| F1 | 100 mM Sodium citrate |  |
| F2 | 100 mM Sodium sulfate |  |
| F3 | 100 mM Sodium phosphate |  |
| F4 | 100 mM Sodium acetate |  |
| F5 | 100 mM Na-malonate |  |
| F6 | 100 mM Na-tartrate |  |
| F7 | 100 mM Sodium formate |  |
| F8 | 100 mM Sodium fluoride |  |
| F9 | 100 mM Sodium bromide |  |
| F10 | 100 mM Sodium nitrate |  |
| F11 | 100 mM Sodium iodide |  |
| F12 | 100 mM Sodium thiocyanate |  |
| G1 | 60 mM NaCl |  |
| G2 | 100 mM NaCl |  |
| G3 | 200 mM NaCl |  |
| G4 | 400 mM NaCl |  |
| G5 | 600 mM NaCl |  |
| G6 | 800 mM NaCl |  |
| G7 | 1000 mM NaCl |  |
| G8 | 5% Glycerol |  |
| G9 | 10% Glycerol |  |
| G10 | 20% Glycerol |  |
| G11 | 10% PEG400 |  |
| G12 | 10% PEG2000 |  |
| H1 | 3% D-glucose |  |
| H2 | 3% Sucrose |  |
| H3 | 30 mM L-arginine |  |
| H4 | 10% PEG3350 |  |
| H5 | 10% PEGMME5000 |  |
| H6 | 10% PEG6000 |  |
| H7 | 10% PEG8000 |  |
| H8 | 10% PEG20000 |  |
| H9 | 1 mM ßME |  |
| H10 | 2 mM ßME |  |
| H11 | 5 mM ßME |  |
| H12 | 10 mM ßME |  |