From the Institute of Biochemistry

of the University of Lübeck

Director: Prof. Dr. Thomas Krey

Exploring structure-based discovery of antivirals targeting *Human cytomegalovirus* (HCMV)

Dissertation

for Fulfillment of Requirements

for the Doctoral Degree – Dr. rer. nat –

of the University of Lübeck

from the Department of Natural Sciences

UNIVERSITÄT ZU LÜBECK

Submitted by

Dinesh Thiyagaraj

from Arani, India

Lübeck, 2025

First referee: Prof. Dr. Thomas Krey

Second referee: Prof. Dr. Jens Bosse

Date of oral examination: July 24, 2025

Approved for printing on July 28, 2025. Luebeck.

Zusammenfassung

Das Humane Cytomegalievirus (HCMV) verursacht weltweit verbreitete Infektionen, die typischerweise asymptomatisch verlaufen und in derem Rahmen das Virus in einem latenten Zustand im Wirt verbleibt. Unter Immunsuppression kann das Virus jedoch reaktiviert werden und schwere Erkrankungen hervorrufen. Die United States Food and Drug Administration (USFDA) hat sechs antivirale Wirkstoffe zur Prävention und/oder Behandlung von HCMV-Infektionen zugelassen: Ganciclovir, Valganciclovir, Cidofovir, Foscarnet, Letermovir (ausschließlich zur Prophylaxe) und Maribavir. Trotz ihrer Wirksamkeit ist die Anwendung dieser antiviralen Therapeutika durch eine oder mehrere der folgenden Herausforderungen eingeschränkt: Toxizität, eingeschränkte Effektivität sowie die Entstehung von Arzneimittelresistenzen. Dies unterstreicht den dringenden Bedarf an neuartigen antiviralen Wirkstoffen, die sowohl hochwirksam als auch sicherer sind und essentielle virale Proteine gezielt inhibieren, um das Risiko einer Resistenzentwicklung zu minimieren.

Ziel dieser Studie war es, das bestehende Spektrum an HCMV-Therapeutika durch die Etablierung einer struktur-basierten Arzneimittelentwicklungs-Pipeline für drei distinkte HCMV-Proteinziele zu erweitern: den pUL94/pUL99-Komplex, pUL98 und pUL77. Im Rahmen dieser Arbeit wurde die strukturelle Charakterisierung zweier dieser HCMV-Zielproteine verfolgt, wobei Kristallisationsmethoden in Kombination mit sogenannten „crystallization chaperones“ sowie alternative Verfahren wie der Einzelpartikelanalyse mittels Kryo-Elektronenmikroskopie (cryo-EM) zum Einsatz kamen. Die resultierenden atomaren Strukturen sollen ein virtuelles Hochdurchsatz-Screening für niedermolekulare Verbindungen ermöglichen, um potenzielle antivirale Kandidaten zu identifizieren. Darüber hinaus verfolgten wir das Ziel, eine teilautomatisierte virtuelle Screening-Workflow zu etablieren und spezifische Inhibitoren der Protein-Protein-Interaktion gegen pUL77 zu identifizieren.

Zunächst exprimierte ich pUL98 in E. coli und konnte es bis zur Homogenität reinigen. Da erste robotergestützte Kristallisationsversuche erfolglos blieben, wurden Sybodies und Megabodies als kristallisationsfördernde Chaperone für pUL98 eingesetzt. Leider führten diese Ansätze nicht zur Bildung von Kristallen mit ausreichender Qualität.

Für den pUL94/pUL99-Komplex standen zu Beginn der Arbeit bereits Sybodies zur Verfügung. Da sich auch hier keine geeigneten Kristalle erzeugen ließen, wurde eine Zusammenarbeit mit der Gruppe von Maya Topf am CSSB in Hamburg initiiert, um die Komplexstruktur mittels Einzelpartikelanalyse durch Kryo-EM zu untersuchen. Zur Verbesserung des Partikelkontrasts in der Kryo-EM entwickelte ich auf Basis der vorhandenen Sybodies sogenannte Legobodies, die an pUL94/pUL99 binden. Dennoch zeigte der Komplex eine ausgeprägte heterogene Zusammensetzung und Konformation, was die erreichbare Auflösung auf etwa 6 Å begrenzte.

Abschließend nutzte ich die publizierten Strukturen des pentameren und monomeren pUL77, um einen Consensus-Screening-Workflow für pUL77 zu etablieren. Dabei kamen mehrere Scoring-Funktionen, Docking-Tools sowie ergänzende Selektionskriterien zum Einsatz. Dies ermöglichte die Identifizierung übereinstimmender Liganden-Posen aus zwei Docking-Programmen und erhöhte damit die Wahrscheinlichkeit für Treffer (true positives). Durch den Einsatz von Clustering-Algorithmen konnten die Moleküle weiter analysiert und die Rate potenzieller Wirkstofftreffer sowie die Auswahl von Leitstrukturen verbessert werden. Die aus dieser Screening-Workflow resultierenden Hitkandidaten wurden bereits bestellt und sollen nun von unseren Kooperationspartnern in zellulären HCMV-Infektionsmodellen getestet werden.

Zusammenfassend wurde in dieser Arbeit eine effiziente Workflow für das virtuelle Screening niedermolekularer Substanzen mit verbesserten Trefferquoten etabliert, deren Ergebnisse über die konventionelle Bewertung mittels Docking-Scores hinausgehen. Dieser

Ansatz soll die Effizienz der struktur-basierten antiviralen Wirkstoffforschung gegen HCMV signifikant steigern.

Abstract

Human cytomegalovirus (HCMV) causes widespread infections globally, typically remaining asymptomatic and latent. However, under conditions of immunosuppression, HCMV can reactivate and cause severe disease. The United States Food and Drug Administration (USFDA) has approved six antivirals for prevention and/or treatment of HCMV infection: Ganciclovir, Valganciclovir, Cidofovir, Foscarnet, Letermovir (only prophylaxis), and Maribavir. Despite their efficacy, these antivirals face one or more of the following challenges: toxicity, efficacy, and the emergence of drug resistance. This highlights the urgent need for novel antivirals that are both effective and safer, targeting essential viral proteins to minimize resistance.

This study aimed to extend the available HCMV drugs by establishing structure-based drug design pipeline for three distinct HCMV protein targets, namely the pUL94/pUL99 complex, pUL98 and pUL77. I worked towards the structural characterization of two HCMV targets, employing crystallization methods in combination with crystallization enhancing scaffolds as "crystallization chaperones" as well as alternative methodologies like cryo-EM single particle analysis. The resulting atomic structures will pave the way for virtual screening of small molecules to identify potential antiviral candidates. Additionally, we aimed to establish a semi-automated virtual screening workflow and screen for protein-protein interaction inhibitors against HCMV pUL77.

I initially expressed pUL98 in *E.coli* and purified it to homogeneity. Since initial robotic crystallization experiments remained unsuccessful, I employed sybodies and megabodies as "crystallization chaperones" for pUL98. Unfortunately, crystallizing pUL98 with sybodies or megabodies did not yield diffraction-quality crystals.

For the pUL94/pUL99 complex sybodies were already available at the beginning of this thesis, and as crystallization also remained unsuccessful, we started a collaboration with the group of Maya Topf at CSSB; Hamburg, to characterize the complex structure using cryo-EM single particle analysis. To increase particle contrast in cryo-EM, I used the existing sybodies to engineer legobodies binding pUL94/pUL99, but the complex showed significant compositional and conformational heterogeneity, limiting the obtained resolution to ~6Å.

Finally, I used the reported cryo-EM structure of the pUL77 pentamer to establish a consensus virtual screening workflow for pUL77 using multiple scoring functions, docking tools and additional screening parameters. This allowed us to identify consensus poses from two docking tools and attempt to improve probability of true-positive hits. Clustering algorithms were used to scrutinize the molecules further, enhancing the hit identification rate and lead selection. The shortlisted molecules from this screening workflow have been ordered and will be tested by our collaboration partners in cell culture models for HCMV infection.

In conclusion, this work established a workflow for virtually screening molecules with improved hit rates, moving beyond reliance on docking scores alone. This workflow aims to enhance the efficiency of structure-based antiviral discovery targeting HCMV infection.

Table of Contents

1. Introduction

1.1 Identification of Human cytomegalovirus

German pathologist Hugo Ribbert first observed inclusion-bearing kidney cells in a stillborn in 1881 [Monto Ho, 2008]. In 1921, Goodpasture and Talbert were the first to suggest that the 'cytomegalia' disease could be caused by a viral agent [Harris D Riley JR., 1997]. In 1956, Smith and Rowe, and independently in 1957, Thomas Weller and his colleagues, isolated the virus causing the disease. In 1960, Thomas Weller named the virus 'Cytomegalovirus (CMV)' due to the characteristic enlargement in the size of infected cells [Alfred Grafe, 1991]. The modern history of cytomegalovirus began with the examination of stillborns and the shedding of the viral agent from infants with generalized diseases [Stern, 1965].

Based on post-mortem studies, two forms of cytomegalic inclusion disease have been observed: disseminated and localized. Disseminated disease is generally severe, where multiple organs are infected, and it often leads to birth defects in congenital CMV infection when the virus infects the fetus in-utero [Stern, 1965].

1.2 Characterization of CMV-associated disease

Characterization of human cytomegalovirus (HCMV) and its associated disease revealed crucial information that led to the addition of HCMV to the family of herpesviruses by the International Committee on Taxonomy of Viruses in 1971 [Wildy, 1971]. The characteristics of HCMV, including a spherical virion with a double-stranded linear DNA genome, an icosahedral capsid, a tegument, and a double-layered envelope, were primary criteria for its inclusion in the family Herpesviridae. Although the general morphology of viruses in Herpesviridae is consistent with these features, variations in overall size, replication cycle speed, host range, and latency establishment categorize herpesviruses into alpha, beta, and gamma subfamilies. CMV belongs to the Betaherpesvirinae subfamily [Davison, 2007].

HCMV has approximately 230 kilobases of linear double-stranded DNA—one of the largest genomes among known human-infecting viruses, about 50% larger than that of Herpes simplex virus-1 that causes cold sores. This genome is densely packed inside a capsid core, surrounded by a tegument containing proteins, intracellular fluid, RNA, and other biomolecules, all enveloped by a double-layered membrane with embedded glycoproteins forming the mature virion [Vescovini et al., 2016; Marti-Carreras & Piet Maes, 2019].

Until 1960s, HCMV was studied extensively in congenital and post-natal pathological contexts. However, research focus shifted towards the 'opportunism' of CMV infection—defined as an infection with limited or no pathogenicity under ordinary circumstances but causing serious disease due to the predisposing effects of another condition, disease, or its treatment [Symmers, 1965]—following the first successful renal transplant by Hume et al. in 1952, which highlighted the need for immunosuppressants to prevent organ rejection [Hume et al., 1952]. The successful use of the immunosuppressant Azathioprine in 1963 led to the observation of HCMV disease in patients administered with immunosuppressants [Oka and Yoshimura, 1996; Hong and Kahan, 2001].

Studies in the 1970s documented the seroconversion of 10 out of 12 seronegative kidney transplant recipients—who received organs from seropositive donors within 90

days post-transplantation [Ho et al., 1975]. This significant observation spurred further research into the establishment of latency and reactivation of HCMV under certain predisposing conditions, with seropositivity being an indicator of latent HCMV infection [Van Son and The, 1989; Stratta et al., 1992; Monto Ho, 2008; Crawford et al., 2022].

HCMV research remains active and extensive, with a growing number of scientists joining the community to understand the virus's life cycle and disease pathology and significant advancements have been made, including the understanding that HCMV establishes latency in myeloid lineage cells—specifically CD33+, CD34+ bone marrow cells, and peripheral blood mononuclear cells, including primary CD14+ monocytes [Marti-Carreras and Maes, 2019]. Reactivation is as common as reinfection, and immunocompromised patients, pregnant women, newborns, and the elderly are at high risk of severe consequences from HCMV infection or reactivation [Bale et al., 1996; Boeckh and Geballe, 2011].

1.3 Transmission and disease burden of CMV infection

Human cytomegalovirus (HCMV) transmission occurs via contact with body fluids such as saliva, tears, urine, genital secretions, nasal secretions, or breast milk. The primary sites of HCMV infection are typically epithelial cells at the entry site, followed by virus spread to various organs and cell types [Dupont and Reeves, 2016]. HCMV infects a wide variety of cells but establishes latency only in a selected subset of cell types [Marti-Carreras and Maes, 2019]. The selectivity of host cell types for HCMV latency is a characteristic of the virus. The triggers for HCMV reactivation are not well understood, but reactivation is observed in conditions such as pregnancy, natural or induced immunosuppression, certain viral diseases like AIDS, and in patients with Hodgkin's disease, malignant, non-Hodgkin's lymphoma and in critical care [De Carmo et al., 2017; Mehravaran et al., 2017].

Epidemiological studies estimate HCMV seroprevalence in diverse populations worldwide to be between 39% and 90%, with Europe having the lowest and the Middle East having the highest reported prevalence [Zuhair et al., 2019]. According to a longitudinal analysis in a German university hospital by Hoehl et al., the overall CMV seroprevalence in Germany stands at 56.48% based on data from 2009 to 2018. This represents a decline from 63.7% between 1988 and 1997 to 57.25% between 1998 and 2008. In the same study, CMV seroprevalence in patients with HIV was 92.93%, in oncology patients 60.64%, and in women of childbearing age 51.7% [Hoehl et al., 2020]. The same study also highlighted age and country of birth as the strongest risk factors for seropositivity [Hoehl et al., 2020]. This observation has been reported in multiple studies, with seropositivity being relatively lower in developed countries compared to developing and underdeveloped countries [Cannon et al., 2010; Lachmann et al., 2018].

The highest disease burden of all is estimated to be due to congenital CMV infection [Griffiths, 2012]. Congenital CMV infection occurs through vertical, intrauterine transmission of the virus from mother to fetus. The rate of infection in women of childbearing age is between 0.5% and 4.0% [Buxmann et al., 2017], and the clinically relevant congenital infection rate is between 0.008% and 0.04% in Germany [Rutten and Rissmann et al., 2017]. While congenital CMV infection is prenatal and severe if primary infection or reactivation occurs in the first trimester of pregnancy, CMV infection of neonates due to breastfeeding or contact with infected maternal secretions during delivery (perinatal) can also lead to long-term morbidities. In Germany, according to a

study by De Lepper and Stephan et al. in 2023, long-term sequelae of congenital and neonatal CMV infection include intrauterine growth retardation (42.6%), sensorineural hearing loss (SNHL) (38.9%), and motor development disorders (33.3%). Additionally, newborns also suffered from visual impairment due to retinitis [De Lepper and Stephan et al., 2023]. The rate of SNHL is similar for congenital and neonatal infection, but the severity is higher for congenital CMV infection [Hoehl et al., 2020]. According to a recent report by Stephan and De Lepper et al. in 2023, the mean annual healthcare costs per newborn between two cohorts of congenital CMV-infected newborns are €22,737 and €34,498 for the first year after birth [Stephan and De Lepper et al., 2023].

Other high-risk groups include patients with predisposed conditions and diseases, and solid organ transplantation recipients whose immune systems are suppressed to prevent graft rejection. In these conditions, infection or reactivation may cause pneumonitis, retinitis, and end-organ diseases including hepatitis and gastrointestinal ulceration as direct effects. Atherosclerosis, systemic inflammation, T-cell stimulation, and monocyte activation are indirect effects [Griffiths and Reeves, 2021]. Despite the severity in susceptible risk groups, CMV infection is generally asymptomatic in healthy individuals and may cause mild flu-like symptoms in about 5% to 20% of cases [Buxmann et al., 2017].

1.4 Existing options for intervention

Human cytomegalovirus (HCMV) infection can lead to significant disease progression and severe illness. Currently, there is no vaccine approved against HCMV that the interventions are generally using antivirals and immunoglobulins. Management strategies in transplantation settings include:

1. **Prophylaxis:** HCMV antiviral drugs are administered to transplant recipients who are at risk of HCMV infection [Yadav et al., 2022].
2. **Pre-emptive Therapy:** Transplant recipients are administered HCMV antivirals when active viral replication is detected, regardless of clinical signs and symptoms. One way of detection of active viral replication is by pp65 antigen tests [Razonable et al., 2003, Hasegawa et al., 2015].
3. **Therapy or Treatment:** Transplant recipients who developed HCMV disease are given HCMV antivirals [Yadav et al., 2022].

Currently, strategies to prevent HCMV infection in seronegative organ transplant recipients include standard medical precautionary testing, hygiene practices, and testing for HCMV IgG antibodies in donors and recipients. The presence of IgG antibodies does not preclude transplantation but helps initiate appropriate therapy regimes for the recipient [Ljungman et al., 2017; Ljungman et al., 2019; Stycynski, 2020].

Antivirals Ganciclovir, foscarnet, fomivirsen, valganciclovir, valaciclovir/valacyclovir, cidofovir, letermovir (only prophylaxis), and maribavir are approved either for treatment and prevention or only for prevention (prophylaxis) in pediatric patients above 12 years and adults by the European Medicines Agency (EMA) until now. However, fomivirsen was withdrawn from the market due to low demand. The United States Food and Drug Administration approved all these drugs for treating HCMV infection, except valacyclovir [Newman and Cragg, 2012; Griffiths and Reeves, 2021].

The effectiveness of each antiviral varies across different risk groups and treatment regimes. For example, letermovir is effective as prophylaxis in hematopoietic stem cell transplantation (HSCT) but is ineffective for pre-emptive therapy and general

treatment in all risk groups [Tan and Palen et al., 2024]. Maribavir is approved for treating pediatric and adult patients with HCMV infection refractory to ganciclovir, valganciclovir, cidofovir, or foscarnet [Connie Kang, 2022; Styczynski, 2020; Buxmann et al., 2017].

Additionally, Hyperimmune immunoglobulin from Biotest AG, marketed as 'Cytotect CP,' was approved for prophylactic therapy in hematopoietic stem cell or solid organ transplant recipients under immunosuppressants by EMA. These immunoglobulins are isolated from HCMV-positive patients and are specific to HCMV, providing passive, prophylactic immunity to recipients [Panesso et al., 2023; Karafin et al., 2018].

Though none of the currently available therapeutic options are approved for treating congenital CMV infection or CMV infection in newborns or pregnant women by the FDA or EMA, antivirals are used off-label for these cases. Valacyclovir has shown effectiveness in preventing intrauterine transmission of HCMV from mother to fetus [Shahar-Nissan, 2020]. The standard care for congenital CMV infection and antenatally acquired CMV infection in newborns involves using ganciclovir within 28 days after birth, followed by valganciclovir for at least six months, which may extend beyond two years depending on disease severity in the newborn [Gyu Hong Shim, 2023].

1.5 Need for discovery of novel antivirals

One of the primary reasons antivirals are not approved to treat HCMV infections in newborns and pregnant women is their toxicity [Pontes et al., 2024]. However, letermovir and maribavir are exceptions and are yet to be studied extensively in these risk groups. Ganciclovir and valganciclovir share similar side effects, as valganciclovir is the prodrug of ganciclovir. Common side effects include leukopenia, spermatotoxicity, and abnormal retinal function with long-term administration, and there is a clear relationship between dosage and toxicity [Hu et al., 2021; Jensen et al., 2021; Lee et al., 2021]. Additionally, ganciclovir and valganciclovir are shown to possess teratogenic characteristics. Cidofovir and foscarnet commonly cause nephrotoxicity, and cidofovir can also cause uveitis, leading to photophobia and blurred vision [Perruccio et al., 2021; Testi et al., 2020; Domingo et al., 2021; Ota and Hirata, 2021]. Information on the long-term effects of letermovir and maribavir is scarce, but they are considered relatively safer compared to other antivirals.

In addition to toxicity, existing antivirals can lead to resistance in HCMV strains with long-term administration. Valganciclovir (the prodrug of ganciclovir) was developed to address the low bioavailability of ganciclovir, with valganciclovir converting to ganciclovir during first-pass metabolism in the liver. Upon entry into host cells, ganciclovir is monophosphorylated by pUL97 (a viral cyclin-dependent kinase ortholog), while cidofovir, already monophosphorylated, skips phosphorylation by the viral kinase, and foscarnet, a pyrophosphate analogue, is not phosphorylated. Two additional phosphate groups are added to ganciclovir and cidofovir by cellular kinases. Both ganciclovir (a guanosine analogue) and cidofovir (a cytosine analogue) bind to the dNTPs binding site of viral DNA polymerase (pUL54), causing premature termination of DNA synthesis and reducing the rate of polymerization. While ganciclovir, valganciclovir, and cidofovir are nucleotide analogues, foscarnet targets the pyrophosphate binding site of the same viral DNA polymerase (pUL54). Letermovir targets pUL56 in the terminase complex, and maribavir targets the viral kinase (pUL97) [Marschall et al., 2024; Bottino et al., 2023; Gourin et al., 2023; Wild et al., 2023].

Drug resistance is suspected when the plasma viral load remains consistently high for several weeks after administering an appropriate dose of antivirals. Resistance usually results from single or multiple mutations in target proteins that affect antiviral binding. Sometimes, mutations rendering antivirals ineffective are associated with severe HCMV disease. Reports on letermovir resistance have recently emerged, while resistance due to mutations in viral DNA polymerase (pUL54) and viral kinase (pUL97) is relatively well-studied. Most remaining antivirals target either of these proteins [Limaye et al., 2000; Eckle et al., 2002; Lurain and Chou, 2010]. Since foscarnet and cidofovir do not require the activation step needed for ganciclovir, they are generally used as second-line antiviral agents when HCMV infection is refractory in patients treated with ganciclovir or valganciclovir. Moreover, because cidofovir and foscarnet do not interact with pUL97, mutations in pUL97 do not affect these antivirals' activity. However, mutations in pUL54 can render one or more of the nucleoside and pyrophosphate analogues ineffective, making letermovir and maribavir the drugs of choice [Chou et al., 2018; Perera et al., 2021].

The toxic and harmful side effects associated with ganciclovir, valganciclovir, foscarnet, and cidofovir, along with the restricted efficacy of letermovir and maribavir and the emergence of antiviral resistance in HCMV strains, highlight the urgent need for new antivirals. These new antivirals should be broadly effective, safer, and target novel viral proteins to mitigate the likelihood of resistance.

1.6 Drug discovery methodologies

The modern drug discovery and development process, which has evolved significantly since the revolution of molecular biology sciences in the 1980s, focuses on molecular medicine to identify and develop treatments for previously untreated medical conditions. Molecular medicine is defined as “a branch of medicine that develops ways to diagnose and treat disease by understanding the way genes, proteins, and other cellular molecules work. It is based on research that shows how certain genes, molecules, and cellular functions may become abnormal in diseases such as cancer” [Henry, 2019; Swinney, 2013].

This approach has shown success in the drug development process to a varied extent, although it has faced challenges. Traditionally, phenotype-based drug screening was the standard method-of-choice, where drugs were tested and evaluated against the phenotype of biological systems such as cells or animals. However, deconvoluting the specific target and describing the Molecular Mechanism of Action (MMOA) was a significant challenge. As a result, many approved drugs do not have a reported target protein but rather a likely disease pathogenesis pathway. Additionally, phenotype-based drug discovery is costly and time-intensive, with lead optimization based on a blind structure-activity relationship.

These challenges, coupled with the need to improve the success rate of the discovery process, motivated a shift towards target-based drug discovery. In this approach, the target involved in disease pathogenesis is identified and validated scientifically. Assays are then developed to screen molecules against the target to find hits and leads. However, lead optimization in this methodology faced similar challenges to phenotype-based drug discovery.

Measurable registrations of new molecules obtained through the target-based method encouraged scientists to focus on further development, leading to the structure-

based drug discovery methodology. This shift does not disregard testing drug molecules on phenotypic assays in cells or animals but rather promotes a more rational approach. This is supported by the analysis of Swinney and Anthony (2011), which showed that 37% of first-in-class drugs were identified through phenotypic screening compared to 23% from target-based screening between 1999 and 2008. Conversely, 51% of follower drugs were from target-based screening compared to 18% from phenotypic screening. This finding revives the relevance of phenotypic-based methodologies.

While the target-based approach shares the challenge of lead optimization, it also faces complexity in describing MMOA due to functional selectivity or cross-reactivity. Structure-based drug discovery, a subset of target-based drug discovery, attempts to address issues in lead optimization, reducing costs and the overall timeline of the drug discovery process. Modern technological advancements have improved the drug-to-clinic (discovery & pre-clinical research) timeline to as low as 15 months, according to a McKinsey report [Ondetti et al., 1977; Urban et al., 2006; Swinney, 2013; Moffat et al., 2017; Vincent et al., 2022].

1.7 Structure-based drug discovery as a strategy

Structure-based drug discovery began in the 1970s when the development of the first human angiotensin-converting enzyme (ACE) inhibitor utilized protein crystallography. This research led to the creation of the first-in-class ACE inhibitor, 'Captopril,' which was approved in 1981 (Rondeau and Schreuder, 2015).

The modern target-based drug discovery process involves several stages, including target identification and validation, hit identification, lead optimization, preclinical development, and clinical trials. These steps ultimately lead to obtaining market approval to launch the product [Chang et al., 2022]. Target identification is crucial for the success of the discovery process. Knowing the precise target to treat a disease allows for screening drug molecules to optimize treatment, improve selectivity, and reduce potential side effects. Ineffectiveness in modulating the disease and higher toxicity in the pharmacological profile are two major reasons for a drug to fail in clinical studies. Target identification significantly contributes to a drug's effectiveness. The term "target" broadly describes various biomolecule types, including enzymes, receptors, ion channels, transcription factors, and nucleic acids, where the drug binds in vivo to exert its action [Moumne et al., 2022]. A druggable target is accessible to the drug molecule and, upon binding, elicits a measurable and statistically significant biological response.

Target identification is a complex process requiring extensive omics and biomedical data mining. Once potential targets are identified, they must be validated. Target validation involves various in vitro biochemical, cell culture, and in vivo animal model experiments to evaluate changes in disease phenotype when the target function is diminished. A good target will show positive disease outcomes when inhibited or knocked out. Sometimes, target validation and hit identification steps are performed together, depending on the availability of information and resources [Chen and Du, 2007; Hughes et al., 2010; Tabana et al., 2023].

Hit identification is the process of screening drug molecules that can potentially modulate the target of interest. This process can be performed in various ways, each with its limitations and advantages. The choice of screening methodology depends on the availability of information on the target, resources, feasibility of assay development, associated costs, and the precision and accuracy of the screening data required to proceed

to further steps of drug discovery [Hughes et al., 2010]. Some hit identification methodologies are described below:

- **Physical screening:**
  - Physical screening involves screening of a library of drug molecules against target of interest through high-throughput in-vitro biochemical or biophysical assays or cell culture based phenotypic screening assays to identify initial hits.
  - Physical screening techniques are generally time consuming and requires the physical availability of a library of compounds that can be expensive.
- **Virtual (in-silico) screening:**
  - This involves using computational algorithms to screen for drug molecules that would bind to the target protein to identify initial hits. Virtual screening is followed by physical screening to validate the computationally identified hits. Virtual screening acts as a pre-screening step to reduce the number of drug molecules to test in physical screening.

Once potential hits are identified, choosing and optimizing the lead involves target-drug structural characterization to model the Structure Activity Relationship (SAR). This process includes modifying the structure and physicochemical properties of the drug molecule to improve binding potency and target modulation. This sub-domain is termed Computer-Aided Drug Design (CADD). Lead identification and optimization are beyond the scope of this thesis. One part of this thesis focuses on mining available literature for target identification and validation and performing virtual screening to identify hits and leads to target desired and novel biomolecules to treat HCMV infection, addressing the limitations of existing approved antivirals.

Currently, existing HCMV antivirals target pUL54 (Viral DNA polymerase), pUL97 (Viral Kinase), and pUL56 (Terminase subunit).

1.8 HCMV life cycle

Understanding the lifecycle and pathogenesis of HCMV is one of the conventional ways to hypothesize potential drug targets. The HCMV lifecycle begins with entry into the host cell, followed by genome replication, capsid assembly, virion maturation, and egress.

1.8.1 Entry into host cell and nuclear transport of genome

HCMV primarily uses two routes for entry into host cells: endocytosis and fusion with the host cell membrane. This process is facilitated by various viral glycoproteins present on the surface of the mature virion. Some notable viral glycoprotein complexes include the gB (pUL55) homotrimer, gH/gL/gO (pUL75/pUL115/pUL74) heterotrimer, gH/gL/pUL128/pUL130/pUL131 heteropentamer, gM/gN (pUL100/pUL73) heterodimer, and the recently characterized gH/pUL116 heterodimer complexes (Calo et al., 2016; Nguyen and Kamil, 2018; Rustandi et al., 2021; Wang and Shenk, 2005). While the precise molecular mechanisms of these entry modes are not fully understood, certain receptors such as platelet-derived growth factor receptor α (PDGFR α), transforming growth factor β receptor type III (TGF βRIII), and neuregulin-2 (NRG-2) are known to interact with the gH/gL/gO trimer complex. Additionally, the gM/gN dimer and gB trimer complexes play a role in the entry process [Wu et al., 2017; Nguyen and Kamil, 2018; Foglierini et al., 2019; Ye et al., 2020; Kschonsak et al., 2021]. The functional significance

of the gH/pUL116 complex is yet to be studied. The abundance of each glycoprotein complex on the surface of the virion varies among different strains of the virus and is suspected to modulate viral tropism towards different host cell types [Wang et al., 2007].

Upon entering the host cell cytoplasm, the viral capsid sheds its envelope and associates with microtubules through capsid-associated tegument proteins for transport to the nuclear pore complex, facilitating nuclear import of the genome. The tegument layer, located between the outer lipid envelope and the capsid, appears amorphous or unstructured under electron microscopy and contains proteins and other biomolecules [Chen et al., 1999; Ogawa-Goto et al., 2003; Kalejta, 2008]. Upon entry, some tegument proteins detach from the capsid and diffuse through the cytoplasm, termed outer tegument proteins, while those that remain attached to the capsid are called inner tegument proteins or capsid-associated tegument proteins. These terms were initially based on the presence of tegument proteins associated with purified virions [Baldick and Shenk, 1996; Mettenleiter et al., 2009; Radtke et al., 2010; Yang et al., 2022]. Major tegument proteins, such as pp71 (pUL82), pp150 (pUL32), pp65 (pUL83), and pp28 (pUL99), are phosphorylated. Phosphoproteins pp150, pUL47, and pUL48 are found associated with the capsid, while pp71 and pp65 translocate freely to the nucleus independently of capsids [Tomtishen, 2012]. Capsid docking over the nuclear pore complex is an important step in nuclear import of genome and the mechanism remains unclear.

1.8.2 Replication of genome

The tegument protein pp71 is suggested to control the expression of Immediate Early (IE) genes, which are crucial for the initiation of genome replication. At this stage, the virus might enter a latent state if there is an unfavorable and quick immune response. In contrast, during the lytic replication cycle, the immediate early proteins IE1 and IE2 facilitate the expression of early genes, thereby initiating genome replication.

Liquid-liquid phase separation (LLPS) mediated by viral proteins, including pUL112 and pUL113, form replication compartments (RC) and the genome replication occurs in the nucleus [Caragliano et al., 2022]. The temporal regulation of gene expression has resulted in classification of individual genes as Immediate Early (IE), Delayed Early (DE), and Late (L) genes. As discussed earlier, nuclear import initiates IE gene expression. IE genes strengthen the immune evasion responses of host cells and act as transcription factors for the expression of DE genes [Shenk and Alwine, 2014]. IE gene products, along with de novo proteins from the HCMV virion, target proteins responsible for host cell cycle regulation, such as cyclin A and E, APC/C ubiquitin ligase, Rb-E2F transcription factors, and hydrolase SAMHD1, to arrest the host cell cycle at the transitionary phase between the G1 and S phases of the cell cycle [Dittmer and Mocarski, 1997; Wiebusch and Hagemeier, 2001; Bogdanow et al., 2021; Turner and Mathias, 2022].

The DNA replication complex mainly consists of proteins from DE genes, including six core proteins: pUL70 (primase), pUL105 (helicase), pUL54 (DNA polymerase), pUL57 (single-stranded DNA-binding protein), pUL44 (DNA polymerase processivity factor), and pUL102 (primase-associated factor) [Manska and Rossetto, 2022].

Preliminary data from a co-immunoprecipitation experiment conducted by Strang et al. in 2009 identified interaction partners of pUL44 from HCMV-infected human foreskin fibroblast (HFF) cells using anti-pUL44 monoclonal antibody (MAb) [Strang et al., 2009]. They identified a new interacting partner, pUL84, co-eluting with pUL44. This

result, in addition to previous studies, confirms that the replication complex is recruited to the origin of lytic replication (oriLyt) using replication initiator proteins IE2 and pUL84 for genome replication [Xu et al., 2004; Strang et al., 2009]. The study also found pUL98 co-immunoprecipitating with pUL44, suggesting a functional role of pUL98 in viral replication [Strang et al., 2009]. The UL98 gene shares significant sequence homology to alkaline exonucleases from other herpesviruses, that pUL98 is suspected to be an alkaline exonuclease [Chee et al., 1990; Sheaffer et al., 1997].

Herpesvirus alkaline nucleases have the highest enzymatic activity at pH levels higher than 8.0 in in-vitro experiments, with 5'-3' exonuclease activity on both single and double-stranded DNA [Sheaffer et al., 1997]. However, the pUL98 ortholog in Kaposi's sarcoma-associated herpesvirus (KSHV), a shutoff exonuclease (SOX), is shown to endonucleolytically catalyze RNA in addition to DNA [Lee et al., 2017; Pardamean and Wu, 2021]. The function of an alkaline nuclease in the HCMV life cycle is unclear. In Herpes Simplex Virus -1 (HSV-1), the alkaline nuclease processes DNA intermediates of the replicated DNA before being packed into capsids [Martinez et al., 1996], while in KSHV, SOX facilitates host protein expression shut-off to support the viral life cycle [Hartenian et al., 2023].

Replacing HCMV alkaline nuclease with HSV-1 alkaline nuclease and recovering the HSV-1 lifecycle from the defect caused by the absence of HSV-1 alkaline nuclease reveals functional conservation [Gao et al., 1998]. The deletion of pUL98 in HCMV is shown to be lethal, suggesting pUL98 is essential for HCMV replication [Weller et al., 1990; Martinez et al., 1996; Yu et al., 2003; Kuchta et al., 2012]. Along with exonuclease activity, HCMV alkaline nuclease also possesses endonuclease activity with a limited catalytic rate [Sheaffer et al., 1997].

Modelling of the pUL98 structure along with mutational enzymatic assays suggest residues D254, E278 and K280 to be critical for exonuclease and endonuclease activity and residues R164 and S252 are proposed to bind 5' phosphate group of DNA [Kutcha et al., 2012]. 48% of predicted type II restriction endonucleases posses PD-(D/E)XK fold, and belong to DEK superfamily [Orlowski and Bujnicki, 2008]. Residues: D254, E278, and K280 from the structural model of pUL98, and the presence of endonuclease activity suggests pUL98 might possess similar structural fold as DEK superfamily of endonucleases where aspartic acid (D), glutamic acid (E) and lysine (K) exist predominatly in the active site [Ban and Yang, 1998]. It is notable that members of DEK superfamily functions as recombinases, integrases, in addition to endo and exo nucleases [Orlowski and Bujnicki, 2008; Wei Yang, 2011]. The precise function of pUL98 in HCMV lifecycle is still unclear. However, its effect, studied through deletion studies, on viral replication, suggests pUL98 to be a potential antiviral target in HCMV.

Over the past 30 years since early 1990s, the discovery and development of antivirals targeting pUL98 have been pursued using target-based drug discovery methodologies [Barnard et al., 1995]. Initial hits identified through in-vitro studies include emodin, atanyl blue-PRL, and thioxothiazolo[3,4-a]quinazoline derivatives [Alam et al., 2015; Zhang et al., 2023]. However, inhibitors such as emodin and the quinazoline derivative (AD-51) exhibit significant cytotoxicity, whereas atanyl blue-PRL, despite having a relatively lower efficacy (EC50 of approximately 6.4 μM compared to EC50 of 4.9 μM for emodin and 1 μM for AD-51) [Alam et al., 2015; Hsiang and Ho, 2008; Zhang et al., 2023], still poses challenges in balancing efficacy and toxicity.

DNA replication is followed by capsid assembly and genome packing.

1.8.3 Capsid assembly and genome encapsidation

While DNA replication is in progress, the proteins required for capsid assembly are transcribed, translated, and localized to the nuclear compartment, where they assemble into the capsid [Muller et al., 2021]. The HCMV capsid is icosahedral in shape, with 20 sides and 12 vertices [Yu et al., 2017; Li et al., 2021]. Capsids are primarily formed by the major capsid protein (MCP: pUL86), the triplex dimer (Tri2A – Tri2B), triplex monomer (Tri1), and the small capsid protein (SCP: pUL48.5) [Yu et al., 2017; Li et al., 2021]. The capsids consist of 11 pentons and 150 hexons formed by MCP protomers [Yu et al., 2017]. One of the twelve vertices of the capsid has a dodecameric portal main body formed by pUL104 through which genome passes for packaging into capsid [Li et al., 2021]. Capsid assembly requires scaffold proteins (pUL80 and pUL80.5) in addition to structural proteins. The interaction between scaffold proteins and structural proteins is critical for procapsid formation in the nucleus. However, mature capsids do not contain scaffold proteins, suggesting their release after capsid maturation [Muller et al., 2021]. HCMV capsids can be categorized as 'A', 'B', and 'C' capsids, representing procapsids that are empty ('A'), capsids that are mature but lack scaffold protein and DNA ('B'), and capsids containing DNA ('C'), respectively [Liu et al., 2021; Li et al., 2021; Borst et al., 2022]. 'C' capsids are present in mature infectious particles [Turner and Mathias, 2022].

The HCMV capsid portal consists of a main body, portal cap, turret (located between the main body and portal cap), and a 10-helix anchor that enhances the stability of the capsid portal structure [Li et al., 2021]. Capsid vertex-specific components (CVSCs) associate with the pentons of the capsid and the phosphoprotein pp150, which is associated with both pentons and hexons [Li et al., 2021]. This association increases the pressure towards the core of the capsid to withstand the high pressure created by packing the large genome inside the capsid [Yu et al., 2017; Li et al., 2021]. CVSCs include pUL77, pUL48, and pUL93. pUL77 placed around the portal vertex constitutes the portal cap that seals the portal after DNA is packed into the capsid, as shown by cryo-electron microscopy structures [Li et al., 2021; Jih et al., 2024]. Interaction with the phosphoprotein pp150 is suggested to stabilize the capsids during maturation, nuclear egress through secondary envelopment, and final maturation of the infectious virion [Tandon and Mocarski, 2008].

DNA replication produces head-to-tail concatemers of the replicated genome, which are processed before being packaged into capsids by the terminase complex. The terminase complex consists of pUL51, which recruits pUL56 and pUL89 to the replication compartment, with pUL56 as the large terminase subunit and pUL89 as the small ATPase subunit. Each component of the holo-terminase complex has different functions and cofactors. DNA packaging occurs in multiple steps. The holo-terminase complex, upon localizing in the nucleus, binds pac motifs (cis-acting packaging signal motifs) on the concatemeric genome and performs the first cleavage of the DNA. The entire complex, along with the DNA, is recruited to the portal vertex of the procapsid. The ATPase activity of the terminase complex facilitates the translocation of unit-length DNA into the procapsid using ATP as a cofactor. Once the unit-length DNA is packaged into the procapsid, the terminase complex cleaves the remaining DNA and detaches from the capsid portal vertex. Knockdown or deletion of any of the three components of the terminase complex severely impairs DNA packaging, specifically inhibiting pUL51 prevents the subcellular localization of pUL56 and pUL89 [Borst et al., 2013; Neuber et al., 2017; Ligat et al., 2018]. All three components of the terminase complex are potential

antiviral targets, with 'Letermovir' being an approved antiviral that inhibits HCMV viral replication by interfering with viral terminal kinase (also termed terminase) subunit pUL56 [Goldner et al., 2011; Bosworth et al., 2020]. pUL56 has been shown to interact with pac motifs within the 'a' sequence of the DNA and mediate the interaction between the DNA-terminase complex and the capsid portal protein pUL104 [Bogner et al., 1998; Dittmer and Bogner, 2005; Wang and McVoy, 2011].

Proteins pUL77 and pUL93 from CVSC are expressed as late proteins, localize to the nucleus and are shown to interact with each other [Borst et al., 2016; DeRussy et al., 2016]. pUL77 is also suspected to interact with the terminase complex and DNA [Meissner et al., 2011; Borst et al., 2016; Ligat et al., 2018]. Deletion of either of pUL77 or pUL93 does not have an effect on capsid assembly but leads to production of only 'B' capsids without DNA and is shown to be essential for viral replication in human fibroblast cells [Yu et al., 2003; Dunn et al., 2003; Borst et al., 2016; Borst et al., 2022]. The significance of interaction of pUL77 with pUL93 is not yet well studied. While it was shown by Meissner et al., in 2011 that N-terminal coiled-coil motif (CCM) of pUL77 is essential for its oligomerization and DNA binding properties, it is postulated that the pentamerization of pUL77 is crucial for stabilization of DNA packed capsids based on a study in KSHV with the orthologue protein pORF19 [Meissner et al., 2011; Naniima et al., 2021]. These findings that pUL77 is important for viral replication and that pentamerization might be one of the steps in the functioning of pUL77 leads to the hypothesis that inhibiting pentamerization of pUL77 would be a potential targeting strategy for development of novel antivirals against HCMV.

While pUL77 was described to be portal cap that presumably seals the portal, function of pUL93 is not well characterized. However, one study by DeRussy and colleagues [DeRussy et al., 2016], shows pUL93 interaccts with components of nuclear egress complex (NEC): pUL59, pUL53 and pUL97 [DeRussy et al., 2016].

This complex orchestrates a process termed 'nuclear egress' where the matured capsid egresses from the nucleus and consists of the primary envelopment and de-envelopment of capsids during crossing the nuclear membrane [Sonntag et al., 2016]. The nuclear egress complex (NEC) is composed of core proteins pUL50 and pUL53 [Milbradt et al., 2012; Sonntag et al., 2016]. The NEC forms a complex with pUL97 (viral cyclin-dependent kinase) via cellular proteins, including p32/gC1Qr, emerin, protein kinase C (PKC), and additional proteins [Sharma et al., 2016]. pUL97 phosphorylates nuclear lamin proteins to facilitate the redistribution of the nuclear lamina to allow for the initiation of nuclear egress [Marschall et al., 2005]. It was also shown that pUL97 phosphorylates the NEC in addition to lamins [Sharma et al., 2014]. The currently proposed mechanism involves the recruitment of mature capsids to the sites of nuclear egress on the inner nuclear membrane, followed by NEC-induced membrane invagination/deformation, scission and budding of transitionary perinuclear membrane vesicles, and the release of naked capsids into the cytoplasm where the secondary envelopment takes place [Bigalke et al., 2014; Zeev-Ben-Mordehai et al., 2015; Hagen et al., 2015; Sanchez and Britt, 2021].

1.8.4 Secondary envelopment and maturation of virions

HCMV assembly occurs in two main steps: capsid formation and DNA encapsidation in nucleus, followed by final assembly in cytoplasm. After the capsids bud through the nuclear membrane and reaches cytoplasm, the final assembly takes place where the capsids acquire the entire tegument layer along with host cellular biomolecules

and a double-layered outer envelope containing viral glycoproteins and cellular proteins. This process, known as secondary envelopment, is orchestrated through the hijacking of the host cytoskeleton, endocytic, and exocytic machineries in a specially formed, juxtanuclear compartment termed the assembly complex/compartment (AC), also referred to as the viral assembly complex/compartment (vAC) in the cytoplasm [Das et al., 2007; Mettenleiter et al., 2009; Tandon and Mocarski, 2012].

Infected cells produce not only mature infectious viral particles but also non-infectious enveloped particles (NIEPs) and dense bodies (DBs), which mature in the assembly complex and egress along with infectious particles [Craighead et al., 1972]. However, the spatiotemporal regulation is unclear. NIEPs do not contain DNA and are suspected to carry B capsids, while dense bodies lack capsids and instead carry a dense aggregation of biomolecules, predominantly the phosphoprotein pp65 (pUL83) [Schmolke et al., 1995]. The functions of DBs and NIEPs are not well understood, but they do not cause or spread infection [Irmiere and Gibson, 1983; Chevillotte et al., 2009; Ahlqvist and Mocarski, 2011].

The assembly compartment is formed by structural rearrangement of the endoplasmic reticulum, Golgi, and other membranous organelles, mediated by microtubules originating from the microtubule-organising center (MTOC) during the late stages of the viral replication cycle. This is observed as a characteristic cytoplasmic inclusion in HCMV-infected cells [Sanchez et al., 2000; Homman-Loudiyi et al., 2003; Das et al., 2007]. Secondary envelopment, the final stage of viral morphogenesis before egress from infected host cells, occurs in the assembly compartment and is itself a multi-step process [Alwine, 2012; Procter et al., 2018]. Steps in secondary envelopment include the acquisition of a partial tegument layer by the cytoplasmic capsid, wrapping the tegumented capsid with a double-layered glycoprotein-embedded membrane envelope, and the final scission of the membrane to produce mature virions [Schauflinger et al., 2013; Read et al., 2019].

It is proposed that the binding of tegument proteins to the capsid begins during capsid assembly and maturation in the nucleus, based on fluorescence microscopy localization studies and cell culture studies of knock-out/deletion of tegument proteins [AuCoin et al., 2006]. Some tegument proteins, including pp150, pUL26, and pUL48, localize to the nucleus and are functionally important [Hensel et al., 1995; Sampaio et al., 2005; AuCoin et al., 2006; Munger et al., 2006; Tandon and Mocarski, 2008; Brock et al., 2015]. Tegument protein pUL94, along with pp150 and pUL48, localizes to both the cytoplasm and nucleus and possesses bipartite nuclear localization signals (NLS1 and NLS2) and one nuclear egress signal (NES) [Liu et al., 2012]. However, the function of this nucleo-cytoplasmic shuttling is unclear. While it is reasonable to assume that tegumentation starts during capsid assembly or maturation in the nucleus, other studies have shown that tegument proteins (pUL32, pUL48, pUL94) and others (pUL71, pUL99, pUL103) are critical to secondary envelopment and localize to the vAC in the late stages of infection [Womack and Shenk, 2010; Smith et al., 2014; Dietz et al., 2018; Read et al., 2019; Butt et al., 2024].

The widely accepted model of HCMV assembly involves the coordination of host endoplasmic reticulum (ER), Golgi, trans-Golgi network (TGN), early and late endosomes from the endosomal pathway, and the viral proteins required for secondary envelopment and maturation of the infective virion. Localization of the Endosomal Sorting Complex Required for Transport (ESCRT) proteins to the vAC, along with the utilization of syntaxin-

3 (SNARE protein) and the endosomal pathway for the translocation of key proteins, further strengthens the current model of viral assembly [Das et al., 2007; Tandon et al., 2009; Das and Pellett, 2011; Cepeda and Fraile-Ramos, 2011; Flomm et al., 2022]. The vAC encloses parts of the endosomal pathway and the trans-Golgi network, contains markers of both pathways, and has been shown to facilitate the translocation of viral glycoproteins to the assembly site through interactions and commandeering of the host proteome and mechanisms [Crump et al., 2003; Krzyzaniak et al., 2009].

Tegument proteins such as pUL94, pUL99, pUL71, and pUL103 play an important role in secondary envelopment. Deletion of pUL71 or pUL103 leads to the accumulation of partially enveloped capsids in the cytoplasm of the infected host cell [Schauflinger et al., 2011; Ahlqvist and Mocarski, 2011; Phillips and Bresnahan, 2012; Das et al., 2014]. pUL99 is a true late protein and the localization of pUL99 to the vAC is mediated by myristoylation of the glycine residue at position 2 and phosphorylation of the protein [Sanchez et al., 2000; Britt et al., 2004; Seo and Britt, 2006]. Myristoylation is the addition of a 14-carbon myristic acid group to the protein, catalyzed by N-myristoyltransferases [Wang et al., 2021]. Although phosphorylation of pUL99 is essential for trafficking to the vAC, hypo-phosphorylated pUL99 exists in extracellular viral particles [Jones and Lee, 2004; Seo et al., 2020]. Deletion of pUL99 results in the accumulation of partially tegumented capsids in the cytoplasm [Silva et al., 2003]. It was also reported that pUL99 is essential for production of infectious viral particles [Seo and Britt, 2006]. These findings suggest that pUL99 is essential for secondary envelopment. The presence of tegument in the accumulated capsids reiterates that pUL99 is a true late protein and membrane-associated to be incorporated into the virion [Sanchez et al., 2000; Silva et al., 2003]. Mutation studies on pUL99 show that up to 60 amino acids in the amino terminal are sufficient for localization to the vAC and for secondary envelopment function [Jones and Lee, 2004; Seo and Britt, 2006; Phillips et al., 2012].

Another tegument protein, pUL94, also expressed with true late kinetics, is also found to be essential and interacts with pUL99 [Phillips et al., 2012]. Deletion of pUL94 exhibited aberrant localization of pUL99 in the cytoplasm, resulting in completely defective viral replication [Phillips et al., 2012]. It is also shown that pUL94 shuttles between nucleus and cytoplasm but the significance of this shuttling is not yet understood [Liu et al., 2012]. Mutational studies of pUL94 and pUL99 revealed that the amino terminal 50 residues of pUL99 are essential to bind to pUL94 [Phillips et al., 2012]. The absence of interaction between pUL94 and pUL99 disrupts the localization of both proteins and also is detrimental to production of infectious virions [Phillips et al., 2012]. Combined with the defect in viral replication in the absence of either protein, this suggests that pUL94 functions to direct pUL99 to the assembly complex to facilitate secondary envelopment [Phillips et al., 2012; Phillips and Bresnahan, 2012].

Mutational studies revealed that cysteines at position 250 and 252 of pUL94 are likely important for interaction with pUL99 [Phillips et al., 2012]. The N-terminal 50 amino acid residues of pUL99 are sufficient to bind to pUL94 and for the secondary envelopment. These findings that both pUL94 and pUL99 are essential for secondary envelopment and production of mature virions and that the interaction between pUL94 and pUL99 is a crucial facilitator for the localization and function of both the proteins, suggests, targeting the interaction between pUL94 and pUL99 would be a potential strategy for development of HCMV antivirals.

Enveloped viral particles utilize the host exosome pathway to release viral particles into the bloodstream, either as individual releases or intermittent bulk releases utilizing multivesicular bodies [Turner et al., 2020; Flomm et al., 2022].

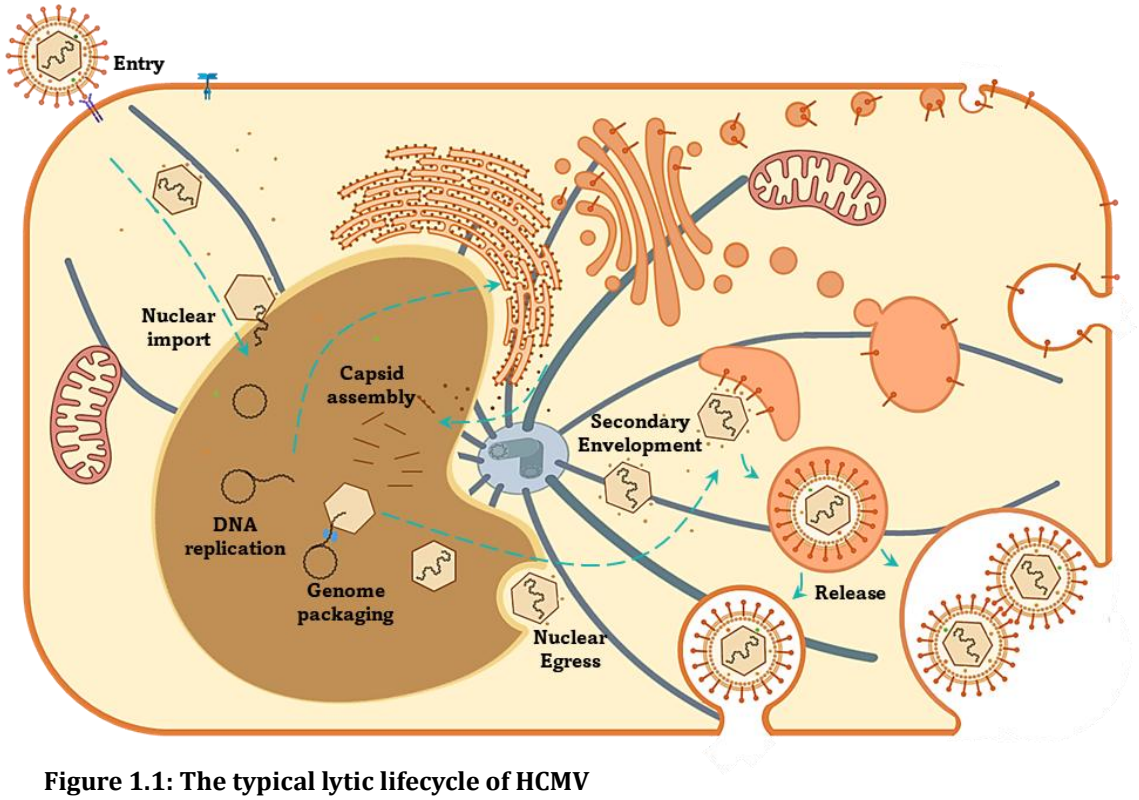


Figure 1.1: The typical lytic lifecycle of HCMV

This figure illustrates the lytic lifecycle of the human cytomegalovirus (HCMV). The lifecycle begins with the virus entering the host cell and concludes with the release (egress) of mature infectious virions. Initially, the virus attaches and fuses with the host cell membrane, releasing its genetic material into the host cell. Following entry, the viral capsid releases the viral genome into the host cell cytoplasm, and the genome is then transported to the host cell nucleus. Within the nucleus, the viral DNA is replicated, and viral genes are transcribed and translated to produce viral proteins. Subsequently, new viral capsids are assembled in the host cell nucleus, and these capsids are packaged with viral DNA, maturing into infectious virions. Finally, the mature virions are released from the host cell to infect new cells, completing the lifecycle.

2. Aim of the thesis

Several potential targets for HCMV antivirals have been described, primarily based on the effects of gene deletion on the phenotype of virus-infected cells in vitro. This thesis focuses on exploring the pUL94/pUL99 complex, pUL98 and pUL77 as targets for the development of antivirals against HCMV using structure-based drug discovery methodologies. The aim is to obtain an atomic structure employing x-ray crystallography and/or cryo-electron microscopy (cryo-EM), where it is not yet reported and subsequently identify hits and leads using computational screening of putative drug molecules against the target structures. The specific aims of the thesis are:

- To establish a semi-automated virtual screening workflow to screen for protein-protein interaction inhibitors against pUL77, with the goal of identifying hits and leads for the structure-based discovery of antivirals targeting HCMV.
- To structurally characterize the pUL94/pUL99 complex and pUL98 from HCMV by employing additional crystallization-enhancing moieties and cryo-EM single-particle analysis.

3. Methods

3.1 AutoDock Tools

AutoDock Tools v1.5.7 was employed to prepare the protein for docking in AutoDock Vina (version 1.2). The protein structure, obtained in PDB format from the Protein Data Bank, was opened in AutoDock Tools and cleaned by removing water molecules, unwanted ligands, and heteroatoms. AD4 atom types were assigned to the protein (Menu > Edit > Atoms > Assign AD4 type). Polar hydrogens were added (Menu > Edit > Hydrogens > Add > Choose only polar), and Gasteiger charges were computed and assigned (Menu > Edit > Charges > Compute Gasteiger). The prepared protein structure was then saved in PDBQT format, which is required for AutoDock Vina. Additionally, grid box parameters (Menu > Grid > Grid Box), including the center and dimensions, were calculated.

3.2 AutoDock Vina

A configuration file for AutoDock Vina was created, specifying the input file path (receptor), grid box centers along the x, y, and z axes (center_x, center_y, center_z), grid box dimensions (size_x, size_y, size_z), exhaustiveness, scoring function (scoring), and the number of CPUs (cpu). This file was saved as a text document. AutoDock Vina was installed and run on a Linux system via the terminal using the command:

vina --config /XXX.txt --ligand /YYY.pdbqt --out /output_directory_path

Here, the configuration file path, ligand PDBQT file path, and output directory path were provided to execute the docking process.

3.3 DiffDock

DiffDock is installed and executed on a Linux system via the terminal. An input CSV file is prepared, specifying the path to the target protein, the path to the ligand, and the name of the ligand. DiffDock is run using the following command:

python -m inference --protein_ligand_csv /protein_ligand.csv --out_dir /output --inference_steps 20 --samples_per_complex 10 --batch_size 10 --actual_steps 19 --no_final_step_noise

The command requires the path to the input CSV file, the output directory path, and other parameters. The parameters mentioned here represent the default configuration, and the template for the input CSV file can be found on the DiffDock GitHub page: https://github.com/gcorso/DiffDock.

3.4 Polymerase Chain Reaction (PCR)

To perform a polymerase chain reaction (PCR), a PCR mixture was prepared using the components listed in Table 3.1 and run in a thermocycler machine at specific temperatures and cycles, as described in Table 3.2. Additionally, an appropriate pair of primers must be designed to meet the specific requirements of the experiment for successful PCR amplification.

Table 3.1 PCR reaction mixture	
Components	**Composition (50 µl)**
10X HF buffer	5 µl
100% DMSO	1.5 µl
10 mM dNTPs	1 µl
100 µM Forward Primer and reverse primer	0.25 µl each
Phusion polymerase	0.5 µl
Template plasmid	10 ng
MilliQ	To make up the volume to 50 µl

Table 3.2 PCR reaction condition	
Steps	**Temperature**
Step 1: Initial Denaturation (2 minutes)	$95^{O}C$
Step 2: Denaturation (1 minute)	$95^{O}C$
Step 3: Annealing (30 Seconds)	Primers specific annealing temperature
Step 4: Extension (1-3 minutes depending on amplicon size)	$72^{O}C$
Step 5: Final Extension (5 minutes)	$72^{O}C$
Step 2 to 4 repeated for 30 cycles	

3.5 PCR product clean up

To purify PCR products using the NucleoSpin PCR Cleanup Kit (catalogue no. 740609.50), the PCR reaction mixture was first adjusted to binding conditions by adding 2 volumes of Buffer NTI to 1 volume of the PCR sample (e.g., 200 µL of NTI for 100 µL of PCR product). The mixture was thoroughly mixed by pipetting and then loaded onto a NucleoSpin column placed in a collection tube. The column was centrifuged at 11,000 x g for 30 seconds to bind the DNA to the silica membrane, and the flow-through was discarded. The column was washed with 700 µL of Buffer NT3 and centrifuged at 11,000 x g for 30 seconds. An optional second wash with 700 µL of Buffer NT3 was performed to ensure high purity, followed by centrifugation under the same conditions. To dry the silica membrane, the column was centrifuged at 11,000 x g for 1 minute. Finally, the column was placed in a clean microcentrifuge tube, and DNA was eluted by adding 15–30 µL of Buffer NE directly onto the membrane. After a 1-minute incubation at room temperature, the column was centrifuged at 11,000 x g for 1 minute to collect the purified DNA.

3.6 Transformation of *E.coli* chemical competent cells

For heat shock transformation of E. coli cells, 50 µL of chemically competent cells were thawed on ice for 10–15 minutes, followed by the addition of 1–5 µL of plasmid DNA (10–100 ng). The cells were gently mixed and incubated on ice for 20–30 minutes to facilitate DNA binding. Heat shock was performed by placing the tube in a 42°C Eppendorf Thermomixer C dry bath for 30–45 seconds. Immediately afterward, the tube was transferred back to ice for 2–5 minutes to stabilize the cells. Subsequently, 300 µL of SOC medium (without antibiotics) was added, and the cells were incubated at 37°C with shaking at 1000 rpm in an Eppendorf Thermomixer C dry bath for 45–60 minutes to allow recovery and the expression of antibiotic resistance genes. Finally, 100 µL of the transformed cells were spread onto LB agar plates containing the appropriate antibiotic, based on the resistance gene present in the plasmid. The plates were incubated upside-down at 37°C for 12–16 hours to enable colony formation.

3.7 Plasmid isolation from DH5α (*E.coli*) cells

Single colonies from an LB agar plate were inoculated into 5 mL of Terrific Broth medium with the appropriate antibiotics and incubated at 37°C with shaking at 200 rpm for 12–16 hours to isolate plasmid DNA using the NucleoSpin Plasmid Isolation Kit (catalogue no. 740588.50). Bacterial cells from the 5 mL overnight culture were harvested by centrifugation at 11,000 x g for 30 seconds. The resulting pellet was resuspended in 250 µL of Buffer A1 (Resuspension Buffer) by pipetting or vortexing. The resuspended cells were then transferred to 2 mL tubes, and 250 µL of Buffer A2 (Lysis Buffer) was added. The tubes were gently inverted 4–6 times to ensure complete lysis without damaging genomic DNA. The lysate was incubated at room temperature for up to 5 minutes, followed by the addition of 300 µL of Buffer A3 (Neutralization Buffer) with thorough mixing by inverting the tube 6–8 times. The lysate was centrifuged at 11,000 x g for 5–10 minutes to pellet cell debris and genomic DNA.

Table 3.3 List of primers

Primer	**Primer sequence (5' → 3')**
Thr2Tev_98_FP	CTTCCAGGGTGGCGAAATGTGGGGCGTCTCGAGT
Thr2Tev_98_RP	TACAGGTTTTCGCTGCCATGGTGGTGGTGGTGATG
98_mNeon_FP	GAAATAATTTTGTTTAACTTTAAGAAGGAGACTCGAGAATGTGGGGCGTCTCGAGTTTG
98_mNeon_RP	CCATTTGACCCTGGAAGTACAGGTTTTCAGCGCTGGGGCTCACCGG
98_mbp_FP	GACTAATGCCGCAGCAGTGGGTGTGTGGCATCTGTTGCGT
98_mbp_RP	GAGTCAGCATTTAATCGCCAACCACTATCAGTGGTGGTGGTGGTG
98_IDR_FP	CGGCGGCTCGGTGTACGAGGTAGCGCGC
98_IDR_RP	CCACCCGAACCCTTGGCGTCAAACACGATCAC
mbs_psb_FP	CAGCCCAGCCTGCAATGGCCCAAGTTCAATTAGTTGAATCTGGTG
mbs_psb_RP	CCAAGCTGGAGACCGTTTAAACTCAATGGTGGTGATGATGGTGC
mbl_psb_FP	CAGCCCAGCCTGCAATGGCCCAAGTACAACTTGTTGAATCAGGT
mbl_psb_RP	CCAAGCTGGAGACCGTTTAAACTCAATGGTGGTGATGATGGTGC
syb_mbs_FP	GAAGCGCATGTAACGACGTCAAAATACGGTTCGCTGCGTCTGA
syb_mbs_RP	GAGACCGTTTAAACTCAATGATGATGATGATGATGTGCGCTCACAGTCACTTG
94_328_FP	ATTTTGTTTAACTTTAAGAAGGAGATATACATATGCCCCAAGACGAAGCCAAACGAATCCTTGTC
94_328_RP	TCATTTGACCCTGGAAGTACAGGTTTTCAGGTTCCTCTACCTTGACCA
94_137_FP	GTCTGCATGTTGGTGCCCCAGGATGAAGCCAAG
94_137_RP	CTCGTAAATCACCATTGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTC
fabHC_FP	GGTTTAAACTCAGGGCCCTCATCAGCTGCCACAAGACTTAGGTTCCACTTTC
fabHC_RP	ATCCTCGAGGAATTCTCATCAGCTGCCACAAGACTTAGGTTCCACTTTC
fabLC_FP	CTTTGTTGGCCTCTCGCTCGGGGGAAATATCATGCTGACCCAGTCTCC
fabLC_RP	GGTTTAAACTCAGGGCCCTCATCAGCACTCGCCCCGGTTGAAAG
syb_lego_FP	GGAACACCACCACCACCACCACCACCACTGAGATCCG
syb_lego_RP	AGGCTGCTCACAGTCACTTGGGTACCTTGGC

The resulting clear supernatant was transferred to a NucleoSpin column placed in a collection tube, and the column was centrifuged at 11,000 x g for 1 minute to bind the plasmid DNA to the silica membrane. The flow-through was discarded, and the column was washed with 600 µL of Buffer A4 (Wash Buffer), followed by centrifugation at 11,000 x g for 1 minute after each wash. An additional centrifugation step at 11,000 x g for 2 minutes was performed to ensure complete drying of the membrane. The column was then placed into a clean 1.5 mL tube, and plasmid DNA was eluted by adding 30 µL of Buffer AE (Elution Buffer) directly onto the membrane. After a 1-minute incubation at room temperature, the tube was centrifuged at 11,000 x g for 1 minute to elute the plasmid DNA. The concentration of the isolated plasmid DNA was measured using a Denovix DS-11 NanoSpectrophotometer at 260 nm with Buffer AE as the blank. The plasmid DNA was then stored at -20°C.

3.8 Restriction-free (RF) cloning

To perform RF (Restriction-Free) cloning, primers were designed using the website rf-cloning.org by providing the insert and target plasmid DNA sequences. The platform calculates the appropriate primer sequences, including the insert and flanking regions complementary to the target plasmid, as well as the required proportions of megaprimer and plasmid. RF cloning involves two sequential PCRs. In the first PCR, the insert was amplified using high-fidelity Phusion polymerase to generate a "megaprimer," a large DNA fragment with overlapping regions that align with the target plasmid. The composition of the first PCR reaction mixture is outlined in Table 3.1, and the PCR was run under conditions described in Table 3.2, with an annealing temperature of 56°C. The resulting megaprimer was purified using the Nucleospin Gel and PCR Cleanup Kit (described in Section 3.5) and used in the second PCR along with the target plasmid. During the second PCR, the megaprimer annealed to complementary sequences on the target plasmid. This reaction extended the megaprimer, incorporating it into the plasmid and creating a complete plasmid with the insert cloned into the target plasmid. The composition of the second PCR reaction mixture, as detailed in Table 3.1, included the megaprimer instead of forward and reverse primers. The amounts of megaprimer and target plasmid were determined based on calculations provided by the RF cloning website for the specific primer pairs. PCR conditions were as described in Table 3.2, with an annealing temperature of 60°C. Following PCR, DpnI (FastDigest DpnI, catalogue no. FD1703) digestion was performed to remove the parental plasmid. This step involved incubating 5 µL of the PCR mixture with 0.5 µL of DpnI at 37°C for 1 hour. The DpnI-digested PCR mixture was directly used to transform E. coli DH5α chemically competent cells, which were then plated on LB agar plates containing the appropriate antibiotics, as detailed in Section 3.7.

3.9 Mutagenesis cloning using PCR

To perform deletions and substitutions in a gene construct, mutagenesis PCR was conducted using primers designed with the NEB primer design tool (www.nebasechangerv1.neb.com). This tool calculates the annealing temperature specific to the designed primers for use in the PCR. The PCR was performed as described in Section 3.4, with the specific annealing temperature determined by the NEB tool. After PCR, 2 µL of the PCR product was combined with the components listed in Table 3.4 for ligation. The mixture was incubated at 37°C for 1 hour without shaking in a 1.5 mL tube. Following incubation, 10 µL of the mixture was used to transform E. coli DH5α chemically competent cells. The transformed cells were then plated on LB agar plates containing the appropriate antibiotics, as outlined in Section 3.7.

Table 3.4 Ligation mixture for mutagenesis PCR product

Component	**Volume**
T4 DNA ligase (from Thermo Fisher, catalogue no. EL0011)	0.5 µL
T4 DNA ligase buffer	1 µL
DpnI (FastDigest DpnI, catalogue no. FD1703)	0.5 µL
T4 Polynucleotide kinase (from Thermo Fisher, catalogue no. EK0031)	0.5 µL
MilliQ water	5.5 µL

3.10 Preparation of TB media

To prepare Terrific Broth (TB) media, 12 g of tryptone, 24 g of yeast extract, 2.31 g of KH_2PO_4, and 12.54 g of K_2HPO_4 were added to approximately 900 mL of deionized water and thoroughly dissolved. Subsequently, 4 mL of glycerol were added, and the solution was mixed until completely homogeneous. The final volume was adjusted to 1 L using MilliQ water. The media was sterilized by autoclaving at 121°C for 15–20 minutes.

3.11 Preparation of ZYP5052 autoinduction media

To prepare 1 L of ZYP5052 media, 928 mL of ZY base media was combined with 50 mL of 20x NPS solution and 20 mL of 50x 5052 solution. To this mixture, 2 mL of 1 M $MgSO_4$ was added, along with 0.2 mL of 1000x trace elements, if required. The final volume was adjusted to 1 L using MilliQ water. Each component of the media was pre-autoclaved at 121°C for 15–20 minutes. The components of ZY base media, NPS solution, 5052 solution, and trace elements solution are detailed below.

Base Media (ZY):

- 10 g/L tryptone
- 5 g/L yeast extract

50x 5052 Solution:

- 25% glycerol (250 g/L)
- 2.5% glucose (25 g/L)
- 10% α-lactose (100 g/L)

20x NPS Solution:

- 0.5 M $(NH_4)_2SO_4$ (6.6 g/100 mL)
- 1 M KH_2PO_4 (13.6 g/100 mL)
- 1 M Na_2HPO_4 (14.2 g/100 mL)

1 M $MgSO_4$:

- 24.65 g $MgSO_4 \cdot 7H_2O$ dissolved in MilliQ water to make 100 mL

Trace Elements (optional):

- 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3, 2 mM H_3BO_3

3.12 Expression of recombinant protein in *E.coli* cells in TB media

An expression plasmid containing the gene of interest under the T7 promoter-terminator system was used to transform the desired E. coli expression strain, as described in Section 3.7. A single isolated colony from an LB agar plate was inoculated into 5 mL of Terrific Broth (TB) medium with appropriate antibiotics (based on the antibiotic resistance gene in the plasmid) and incubated at 37°C with shaking for 16 hours. One millilitre of this culture was reinoculated into 100 mL of TB medium in a 250 mL sterile Erlenmeyer flask containing antibiotics and incubated under the same conditions for another 16 hours. The following day, 10 mL of the grown culture was inoculated into 1 L of TB medium with antibiotics in a 5 L sterile Erlenmeyer flask. The culture was incubated at 37°C with shaking at 200 rpm until it reached an optical density (OD) of 0.8. Upon reaching the required density, the culture was induced with IPTG at a final concentration of 1 mM (1 mL of IPTG from a 1 M stock solution was added to the 1 L culture), and the temperature was reduced to 18°C. The incubation was continued for an additional 16

hours at 200 rpm in a shaker incubator. After the 16-hour incubation, the cells were harvested by centrifugation at 5,000 x g for 10 minutes at 4°C using an Avanti J25 centrifuge with a JLA8.1000 rotor and 1 L buckets. The supernatant was discarded, and the cell pellet was transferred to a 50 mL tube and stored at -80°C.

3.13 Expression of recombinant protein in *E.coli* cells in ZYP5052 autoinduction media

An expression plasmid containing the gene of interest under the T7 promoter-terminator system was used to transform the desired E. coli expression strain, as described in Section 3.7. A single isolated colony from an LB agar plate was inoculated into 5 mL of LB medium with the appropriate antibiotics (based on the antibiotic resistance gene present in the plasmid) and incubated at 37°C with shaking for 16 hours. One milliliter of this culture was reinoculated into 100 mL of Terrific Broth (TB) medium in a 250 mL sterile Erlenmeyer flask containing antibiotics and incubated under the same conditions for another 16 hours. The following day, 10 mL of the grown culture was inoculated into 1 L of ZYP5052 autoinduction medium (prepared as described in Section 3.11) with antibiotics in a 5 L sterile Erlenmeyer flask. The culture was incubated at 37°C with shaking at 200 rpm until it reached an optical density (OD) of 0.7. Once the culture reached the required density, the temperature was reduced to 18°C, and incubation continued for an additional 16 hours at 200 rpm in a shaker incubator. After this incubation period, the cells were harvested by centrifugation at 5,000 x g for 10 minutes at 4°C using an Avanti J25 centrifuge with a JLA8.1000 rotor and 1 L buckets. The supernatant was discarded, and the cell pellet was transferred to a 50 mL tube and stored at -80°C.

3.14 Preparation of Schneider 2 (S2) stable cell line

On **Day 0**, seed cells at a concentration of $1x10^6$ cells/ml in Schneider's Drosophila medium (Complete media). The next day, **Day 1**, transfect the cells using 1µg of each heavy and light chain pMT vectors. Transfection is performed using Effectene transfection reagent kit from Qiagen containing Buffer EC, Enhancer and Effectene. Begin by mixing 2 µg of plasmid DNA (if there are multiple plasmid DNAs, mix them in equal quantity such that the final amount is 2 µg) with 0.1 µg pCoPuro for Puromycin selection and top up the volume to 150 µl with Buffer EC. Add 16 µl of Enhancer, vortex the mixture, and incubate it for 5 minutes at room temperature. Then, add 20 µl of Effectene, vortex again, and incubate for another 15 minutes at room temperature. While the incubation is ongoing, prepare the cells by carefully pipetting off the media. Wash the cells by adding 2 ml of Complete media and carefully pipette it off without dislodging the cells, then add 4 ml of Complete media. After the incubation, add 850 µl of Complete media to the transfection mixture and transfer the entire mixture dropwise onto the cells, swirling the flasks gently to mix the contents. Incubate the cells for 16–24 hours at 28°C.

On **Day 2**, replace the media. Transfer the cells to a 15 ml Falcon tube, ensuring they are completely dislodged from the T25 flask. Spin the tube at 1100 rpm (approximately 200g) for 5 minutes and discard the supernatant. Resuspend the cells in 5 ml of Schneider's complete media. Transfer the resuspended cells back into the flask and incubate for two days.

On **Day 4**, add 8 µg/ml of Puromycin to the culture and incubate the cells for 72 hours. Following this, on **Day 7**, change the media by following steps similar to Day 2: transfer the cells, spin, discard the supernatant, and resuspend them in 5ml of Schneider's complete media. Culture the cells for another 48 hours.

On **Day 9**, repeat the media change steps as done on Day 7 and culture the cells again for 48 hours. On **Day 11**, begin adapting the cells to Insect cell express media with Puromycin. Transfer the cells (approximately 5 ml) to a T75 flask and add an equal volume of insect cell express media with Puromycin. Culture the cells for 72 hours.

On **Day 14**, add 5 ml of insect cell express media with Puromycin and continue culturing for 48 hours. Finally, on **Day 16**, Cells are counted and frozen at the density of 20 million cells per ml with freezing media (45% conditioned media, 45% fresh insect express media with puromycin and 10% DMSO) for backup. Remaining cells are passaged for cells maintenance and expansion into larger volume for expression.

3.15 Cytoplasmic expression of recombinant protein in Schneider 2 (S2) cells

Schneider 2 stable cells, containing the gene for expression of the target protein, were cultured at 28°C with shaking at 70 rpm in a shaker incubator until reaching a density of 3 million cells per milliliter in Insect Express media from Lonza. The culture was induced with cadmium chloride at a final concentration of 4 µM and incubated under the same conditions (28°C, 70 rpm) for 5 days. On day 5, the cells were harvested by centrifugation at 5,000 x g for 10 minutes using an Avanti J25 centrifuge with a JLA8.1000 rotor and 1 L buckets. The supernatant was discarded, and the cell pellet was transferred into 50 mL tubes and stored at -80°C.

3.16 Secretory expression of recombinant protein in Schneider 2 (S2) cells

Schneider 2 stable cells, containing the gene for expression of the target protein, were cultured at 28°C with shaking at 70 rpm in a shaker incubator until reaching a density of 4 million cells per milliliter in Insect Express media from Lonza. The culture was induced with cadmium chloride at a final concentration of 4 µM and incubated under the same conditions (28°C, 70 rpm) for 5 days. On day 5, the culture supernatant containing the target protein was harvested by centrifugation at 5,000 x g for 15 minutes using an Avanti J25 centrifuge with a JLA8.1000 rotor and 1 L buckets.

3.17 Preparation of S2 culture supernatant for purification

The cell culture supernatant obtained as described in Section 3.16 was concentrated using a tangential flow filtration (TFF) system, comprising a Masterflex peristaltic pump combined with a Sartorius Vivaflow 200 cross-flow cassette – PES (catalogue no. VF20P0). The supernatant was concentrated to 90 mL and supplemented with 10 mL of 1 M Tris (pH 8.0) buffer and 100 µL of 15 mg/mL Avidin to sequester excess biotin from the culture media. The concentrated protein solution was centrifuged at 50,000 x g for 30 minutes at 10°C in 50 mL tubes using an Avanti J25 centrifuge with a JA25.50 rotor. The resulting supernatant was filtered through a 0.45 µm filter.

3.18 Bacterial cell lysis by sonication

The bacterial cell pellet was resuspended in the appropriate lysis buffer at a weight-to-volume ratio of 1 g to 5 mL. The suspension was sonicated on ice using a

Branson W250D probe sonicator at 70% amplitude with a 1-second ON/OFF cycle, for a total ON time of 15 minutes (suitable for volumes between 200 mL and 350 mL). The resulting lysate was centrifuged at 50,000 x g for 30 minutes at 4°C using 50 mL tubes in an Avanti J25 centrifuge with a JA25.50 rotor. The supernatant was filtered through a 0.45 µm filter.

3.19 Insect cell lysis by sonication

The insect cell pellet (Schneider 2 cells) was resuspended in the appropriate lysis buffer at a weight-to-volume ratio of 1 g to 5 mL. The suspension was sonicated on ice using a Branson W250D probe sonicator at 40% amplitude with a 1-second ON/OFF cycle for a total ON time of 10 minutes (suitable for volumes between 100 mL and 250 mL). The resulting lysate was supplemented with 1.5 µg/mL Avidin and centrifuged at 50,000 x g for 30 minutes at 10°C using 50 mL tubes in an Avanti J25 centrifuge with a JA25.50 rotor. The supernatant was then filtered through a 0.45 µm filter.

3.20 Ni-affinity purification

The lysate obtained from the lysis of bacterial cells (as described in Section 3.18), containing recombinant protein with a polyhistidine tag and an Imidazole concentration of less than 60 mM, was loaded onto a 5 mL HisTrap FF column from Cytiva (catalogue no. 17531901). The column was pre-equilibrated with the same lysis buffer used for cell lysis (refer to Section 3.18). After the protein was loaded onto the column, it was washed with 50 mL of wash buffer and eluted with 50 mL of elution buffer, either through a 0%-100% gradient or by batch elution directly at 100% elution buffer. The eluted protein was collected in fractions of 3 mL. Ni-affinity purification was conducted using the AKTA pure 25 Fast Performance Liquid Chromatography (FPLC) system at 15°C.

3.21 Streptactin purification

The lysate (from cytoplasmic expression) or supernatant (from secretory expression) of insect cell culture containing recombinant protein with a Twin Strep tag was loaded onto a 5 mL Strep-Tactin XT 4Flow column from IBA Lifesciences (catalogue no. 25023001). The column was pre-equilibrated with Strep-Tactin wash buffer (refer to the buffer table). After the protein was loaded onto the column, it was washed with 50 mL of Strep-Tactin wash buffer (refer to the buffer table) and eluted with 50 mL of Strep-Tactin elution buffer (refer to the buffer table) via batch elution, directly at 100% elution buffer. The eluted protein was collected in 3 mL fractions. Strep-Tactin purification was conducted using the AKTA pure 25 Fast Performance Liquid Chromatography (FPLC) system at 15°C.

3.22 Size exclusion chromatography (SEC) purification in 26/600 column

Ten milliliters of protein solution were injected into a size exclusion chromatography 26/600 column (~320 mL volume) connected to the AKTA pure 25 FPLC system. The column was pre-equilibrated with 400 mL of SEC buffer. Following injection, chromatography was performed at a flow rate of 1.5 mL/min at 15°C using SEC buffer for a total of 400 mL. The eluted protein was collected in 3 mL fractions using fraction tubes.

3.23 Size exclusion chromatography (SEC) purification in 10/300 column

A 0.8 mL aliquot of protein solution was injected into a size exclusion chromatography 10/300 column (~23 mL volume) connected to the AKTA pure 25 FPLC system. The column was pre-equilibrated with 40 mL of SEC buffer. Following the injection, chromatography was performed at a flow rate of 0.5 mL/min at 15°C using SEC buffer for a total volume of 40 mL. The eluted protein was collected in 0.2 mL fractions into a 96-well block.

3.24 Chemical biotinylation of target protein

To prepare the target protein for sybody screening, chemical biotinylation was performed using EZ-Link NHS-PEG4-Biotin (catalogue no. 21330). The process began by adding the required quantity of EZ-Link NHS-PEG4-Biotin from a 20 mM stock solution to the target protein solution, achieving a final ratio of 22:1 (Biotin:Target Protein). The biotinylation reaction was incubated for 15 minutes at 4°C. Excess EZ-Link NHS-PEG4-Biotin was neutralized by adding Tris buffer (pH 8.0) to a final concentration of 100 mM. The biotinylated protein solution was subsequently injected into a Size Exclusion Chromatography (SEC) column (HiLoad 10/300GL Superdex 200 Increase) using SEC buffer. The biotinylated target protein eluted from the SEC was collected in 1 mL fractions.

3.25 HABA assay for estimation of extent of biotinylation

The HABA assay was conducted to quantify biotinylation levels in the biotinylated target protein. A HABA/Avidin solution was prepared by mixing 6.7 µL of a 15 mg/mL Avidin stock solution, 92.7 µL of SEC buffer, and 0.6 µL of a 10 mM HABA solution (Thermo Fisher, catalogue no. 1854180). Quantification was performed using the Denovix DS-11 NanoSpectrophotometer, with the HABA/Avidin solution serving as the blank. For the assay, 9 µL of the HABA/Avidin solution was mixed with 1 µL of biotinylated target protein and incubated in the dark at room temperature for 5 minutes. Absorbance was measured at 500 nm, and the resulting value was input into the Thermo Fisher HABA calculator to determine the number of biotin molecules per mole of target protein.

3.26 Screening of sybodies against target protein

The sybody binder screening process involved several steps, starting with ribosome display. mRNAs from concave, loop, and convex libraries, categorized by CDR3 length, were screened against the biotinylated target protein using PUREfrex2.0 with DS Supplements (catalogue no. GFK-PF201-0.1-EX). The ribosome-mRNA-sybody complex was introduced to the biotinylated target protein immobilized on Dynabeads MyOne Streptavidin T1 beads (catalogue no. 65601). After binding, the complexes were eluted, and the mRNA was converted into cDNA through reverse transcription PCR using Agilent AffinityScript reverse transcriptase (catalogue no. 600105). The cDNA was then amplified with Phusion DNA polymerase and cloned into the pDXinit vector (Addgene, #110101) via FX cloning using the BspQ1 enzyme (NEB, catalogue no. R0712L). The pDXinit phagemid vector was subsequently employed for phage display with M13 bacteriophage.

Electrocompetent SS320 cells (Biocat, catalogue no. 60512-1-LU) were transformed by electroporation using the Biorad Gene Pulser electroporation system at 2000 V, 25 µF capacitance, 200 Ω resistance, and 4°C, with pDXinit plasmids. The transformed cells were infected with M13KO7 helper phage (catalogue no. N0315S) for

30 minutes at 37°C and incubated overnight to produce phages. The phages were purified from the culture supernatant using PEG/NaCl precipitation and screened against the biotinylated target protein immobilized on immunoplates (48 wells per library). Bound phages were eluted with 0.25 mg/mL trypsin, reinfected into SS320 cells, and used for subsequent rounds of phage display. For later rounds, Dynabeads MyOne Streptavidin C1 beads (catalogue no. 65001) were used for immobilization of the biotinylated target protein. The eluted phages were reintroduced into SS320 cells for phagemid recovery. Isolated phagemids were cloned into the pSBinit vector (Addgene, #110100) via FX cloning for sybody expression. MC1061 cells (Biocat, catalogue no. 60514-1) transformed with pSBinit plasmids were plated onto LB agar, and 93 colonies per library were cultured in 96-well blocks. Sybody expression was induced with a final concentration of 0.02% L-arabinose and incubated overnight at 22°C. Expressed sybodies in the periplasm were lysed using a hypotonic solution and tested for binding on immunoplates coated with anti-cMYC antibody (Merck, catalogue no. M4439). The biotinylated target protein was added alongside a biotinylated maltose-binding protein as a negative control. ELISA signals were developed using streptavidin-HRP and measured at 650 nm. Sybodies showing signal intensities >1.5-fold above the negative control were identified as hits. Of these, 36 plasmids were sequenced to exclude sybodies with minimal CDR similarity.

The selected sybodies were expressed in MC1061 cells cultured in 50 mL Terrific Broth medium. Following induction and overnight incubation at 22°C, the cells were lysed, and the sybodies were purified using nickel-affinity resin in gravity flow columns. Elution was carried out using a buffer containing 50 mM Tris (pH 7.5), 150 mM NaCl, and 500 mM Imidazole. Purified sybodies were subjected to SEC (using a Sephax 10 column) to isolate monomeric variants in 50 mM Tris (pH 7.5) and 150 mM NaCl. For further details, refer to Zimmermann et al., 2020.

3.27 Single-point kinetics measurement using Bio-layer Interfermetry (BLI)

To estimate single-point kinetic values using Octet Red96e bio-layer interferometry, 500 nM of sybodies with a C-terminal polyhistidine tag (ligand) were coated onto HIS1K-Anti-Penta-HIS biosensors. The biotinylated target protein, serving as the analyte, was prepared at a concentration of 200 nM. The assay setup, as outlined in Table 3.5, guided the movement of the biosensors between wells. Baseline and dissociation steps were performed using BLI buffer, composed of SEC buffer components supplemented with 0.1% Tween 20 and 0.05 mg/mL BSA. This buffer was also used to dilute both the sybodies and the biotinylated target protein. The data obtained from the assay were analyzed using Octet data analysis software. Key steps included baseline subtraction using a no-loading negative control, sensogram alignment to the start of the association phase, and fitting a one-to-one binding model to calculate kon, koff, and KD values. Each well was filled with 200 µL of the corresponding solution.

Table 3.5 BLI assay parameters for single-point kinetics measurement

Steps	**Time in seconds**
Baseline	30
Loading	60
Baseline	30
Association	150
Dissociation	300

3.28 Cross-competition assay using Bio-layer Interfermetry (BLI)

A cross-competition assay was conducted using Octet Red96e bio-layer interferometry to identify sybodies competing for overlapping antigenic regions on the target protein. The assay involved loading 150 nM of biotinylated target protein onto a Streptavidin SA biosensor. Two solution sets were tested: one containing 500 nM of a single sybody and the other containing 500 nM of the same sybody combined with 500 nM of an additional sybody, extending this approach to include all sybody pairs.

Sensograms were analyzed using Octet data analysis software. The analysis included subtracting baseline readings from a no-loading negative control, aligning sensograms to the start of the association step, and comparing association differences between wells corresponding to each sybody pair. Each well was filled with 200 µL of the tested solution. The experimental steps were outlined in Table 3.6. The BLI buffer, which consisted of SEC buffer components, was supplemented with 0.1% Tween 20 and 0.05 mg/mL BSA and was used to dilute the sybodies, target protein, and as the baseline buffer.

Table 3.6 BLI assay setup for cross-competition assay

Steps	**Time in seconds**
Baseline	30
Loading	60
Baseline	30
Association - first sybody	120
Association – first and second sybody	120

3.29 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) was performed to identify optimal buffer conditions for ensuring protein stability and minimizing aggregation. The experiment utilized the SpectroQ instrument (Xtal Concepts), which required samples to be loaded into MRC 96-well under-oil plates (catalogue no. UO96T-UVP). Buffer screens were prepared in a deep-well block, and the DLS plate was set up using the Crystal Gryphon LCP system (Art Robbins Instruments), combining 0.4 µL of protein solution with 0.6 µL of buffer in each well. To prevent evaporation, a layer of 30 µL of paraffin oil was added, and the plate was placed in the SpectroQ instrument, which was maintained at 20°C and equipped with a plate hotel for automated storage and retrieval.

The DLS plate was equilibrated for 30 minutes in the plate hotel before measurements, which involved 10 runs (10 seconds per run) per well. Drop search and fine search steps were conducted to accurately focus the laser on the protein-buffer drop.

Scattering data were analyzed using the Spectro Crystal application, employing an auto-correlation function to generate curves for estimating the hydrodynamic radius and polydispersity index. The resulting radius distribution plots provided insights into hydrodynamic radius and sample heterogeneity. The experiment was repeated over 3–5 days, and data from all days were compiled into a single radius distribution plot. The viscosity was set to 1.0 mPa·s to reflect the theoretical viscosity of the buffer solutions.

3.30 Thermal Shift Assay (TSA)

The Thermal Shift Assay (TSA) was conducted to determine the melting temperature (Tm) of the target protein under different buffer and additive conditions, using SYPRO Orange dye to monitor protein thermal stability. SYPRO Orange binds to

hydrophobic regions exposed during protein unfolding, resulting in increased fluorescence. A final concentration of 3 µM target protein solution in 100 mM screening buffer was prepared in a total volume of 28.8 µL by mixing stock solutions of the target protein and buffer components in MilliQ water. Subsequently, 1.2 µL of 50X SYPRO Orange dye (Merck, catalogue no. S5692) was added to the mixture. The dye, being light-sensitive, was handled in the dark to prevent degradation.

The experiment was performed using a StepOne qPCR machine (Applied Biosystems). The temperature was incrementally raised from 25°C to 95°C in 1°C steps, each lasting 30 seconds, while fluorescence was measured at each step as the protein unfolded. The data generated a sigmoidal curve, with the point of inflection indicating the Tm. Alternatively, the lowest point on the negative first-order derivative of the sigmoidal curve also provided the Tm value.

3.31 Limited proteolysis experiment

Limited proteolysis was performed to identify stable domains within the target protein. The proteases and their respective concentrations are listed in Table 3.7. Protein-protease solutions were prepared and incubated at 37°C for 1 hour. Following the incubation, the reactions were terminated by adding 5x SDS-PAGE sample loading dye, and the samples were analyzed using a 15% SDS-PAGE gel.

Table 3.7 Proteases used for limited proteolysis experiment

Name of the Protease	Stock concentration of proteases	Volume of protease used	Concentration of target protein	Total volume made up with target protein buffer
Thermolysin	1 mg/mL	1 µL	180 µM	12 µL
Proteinase K	1 mg/mL	1 µL	180 µM	12 µL
Streptomyces protease	10 mg/mL	1 µL	180 µM	12 µL
Licheniformis protease	10 mg/mL	1 µL	180 µM	12 µL
Papain	10 mg/mL	1 µL	180 µM	12 µL

3.32 Surface lysine methylation

Surface lysine methylation was performed to chemically modify lysine residues on the surface of the target protein via covalent addition of methyl groups to the lysine side chains. One milliliter of protein solution (1 mg/mL) was mixed with 20 µL of 1 M dimethylamine-borane complex (Merck, catalogue no. 180238) and 40 µL of 1 M formaldehyde, followed by incubation at 4°C for 2 hours. This step was repeated with the same additions and incubation for another 2 hours. Finally, 10 µL of 1 M dimethylamine-borane complex was added to the solution, and it was incubated overnight at 4°C.

3.33 Size exclusion chromatography – multiangle light scattering (SEC-MALS)

Multiangle light scattering (MALS) is a static light scattering technique used to estimate molecular weight, radius of gyration, second virial coefficient, and other parameters. In this setup, a MALS device (miniDAWN from Wyatt) was connected in series with an RI detector (Shodex RI-501), an in-line UV spectrophotometer (variable wavelength detector from Agilent), and a High-Performance Liquid Chromatography

(HPLC) system (Agilent 1260 Infinity II HPLC system) equipped with a Superdex 200 Increase 10/300GL size exclusion chromatography column.

A protein sample of 100 µg, in a volume not exceeding 100 µL, was injected into the HPLC system connected to the SEC column, UV detector, RI detector, and MALS device. The HPLC was operated using the Agilent HPLC dashboard application, with the flow rate set to 1.0 mL/min. Sample injection, as well as UV, RI, and scattering measurements, were managed and recorded using ASTRA software version 8.0 (Wyatt). Method setup included inputting the sample concentration and extinction coefficient.

To standardize and analyze the results, 25 µL of bovine serum albumin (BSA) at a concentration of 2 mg/mL was used. Following the analysis, the software provided molecular weight and polydispersity index values for the target protein.

3.34 Tobacco Etch Virus (TEV) protease digestion

One milligram of the target protein was mixed with 0.25 mg of TEV protease and supplemented with Dithiothreitol (DTT) to a final concentration of 1 mM. The protein solution was poured into a 10 kDa cut-off SnakeSkin dialysis bag (Thermo, catalogue no. 88245). Both ends of the dialysis bag were clamped, and the bag was dialyzed against TEV_dialysis buffer (as detailed in Table 3.8) for 16 hours. The dialysis was conducted in a beaker containing a magnetic bead, placed on a magnetic stirrer set at 100 rpm. The dialysis bag was hung using a thread to ensure it did not come into contact with the spinning magnetic bead at the bottom of the beaker.

After 16 hours, the digested protein solution was collected and loaded onto a 5 mL HisTrap FF column pre-equilibrated with TEV_dialysis buffer. The TEV protease and uncleaved target protein bound to the column resin, while the cleaved protein was recovered in the flow-through solution. Following the protein solution, an additional 10 mL of TEV_dialysis buffer was passed through the column, and the flow-through contained the target protein without the affinity tag.

3.35 Expression of sybodies in MC1061 *E.coli* cells in TB media

An expression plasmid containing the gene of interest for sybody, in the pSBinit vector under the araBAD promoter system with a pelB periplasmic expression signal, was used to transform the MC1061 E. coli expression strain, as described in Section 3.7. Transformation was performed in the presence of chloramphenicol (35 µg/mL). A single isolated colony from an LB agar plate was inoculated into 5 mL of TB medium with chloramphenicol and incubated with shaking at 37°C for 16 hours.

One milliliter of this culture was reinoculated into 100 mL of TB medium in a 250 mL sterile Erlenmeyer flask containing chloramphenicol and incubated under the same conditions for another 16 hours. The following day, 10 mL of the grown culture was inoculated into 1 L of TB medium with chloramphenicol in a 5 L sterile Erlenmeyer flask. The culture was incubated at 37°C with shaking at 200 rpm until it reached an optical density (OD) of 0.8.

Once the required density was reached, the culture was induced with L-arabinose at a final concentration of 0.02% (2 mL of 10% L-arabinose per 1 L of culture), and the temperature was reduced to 18°C. The incubation was continued for an additional 16 hours at 200 rpm in a shaker incubator.

After the 16-hour incubation, the cells were harvested by centrifugation at 5,000 x g for 10 minutes at 4°C using an Avanti J25 centrifuge with a JLA8.1000 rotor and 1 L buckets. The supernatant was discarded, and the cell pellet was transferred into a 50 mL tube and stored at -80°C.

Table 3.8 List of buffers

Name	**Composition**
98_lysis1	50 mM Tris pH8.0, 300 mM NaCl, 10% Glycerol, 40 mM Imidazole, 2mM ATP, 5mM $MgCl_2$, 10 µg/mL DnaseI, 0.5x Protease inhibitor cocktail, 0.2% Triton-X-100.
98_wash1	50 mM Tris pH8.0, 300 mM NaCl, 10% Glycerol, 40 mM Imidazole.
98_urea	50 mM Tris pH8.0, 300 mM NaCl, 10% Glycerol, 40 mM Imidazole, 1 M Urea.
98_elution1	50 mM Tris pH8.0, 300 mM NaCl, 10% Glycerol, 500 mM Imidazole.
tev_dialysis	10 mM Tris pH8.0, 300 mM NaCl, 5% Glycerol.
98_sec	10 mM HEPES pH8.0, 300 mM NaCl, 5% Glycerol.
sybody_lysis	50 mM Tris pH8.0, 300 mM NaCl, 5% Glycerol, 30 mM Imidazole, 10 µg/mL DnaseI, 0.5x Protease inhibitor cocktail.
sybody_wash	50 mM Tris pH8.0, 300 mM NaCl, 5% Glycerol, 30 mM Imidazole.
sybody_elution	50 mM Tris pH8.0, 300 mM NaCl, 5% Glycerol, 500 mM Imidazole.
sybody_sec	20 mM Tris pH7.4, 150 mM NaCl, 5% Glycerol.
98_mbs_sec	20 mM HEPES pH7.4, 150 mM NaCl.
9499_lysis	100 mM Tris pH8.0, 150 mM NaCl, 1 mM EDTA.
streptactin_wash	10 mM Tris pH8.0, 150 mM NaCl, 1 mM EDTA.
streptactin_elution	10 mM Tris pH8.0, 150 mM NaCl, 1 mM EDTA, 50 mM D-Biotin.
9499_sec	20 mM HEPES pH7.4, 150 mM NaCl, 5% Glycerol.
strep_lego_wash	50 mM HEPES pH8.0, 150 mM NaCl, 1 mM EDTA, 5 mM D-maltose.
strep_lego_elution	50 mM HEPES pH8.0, 150 mM NaCl, 1 mM EDTA, 5 mM D-maltose, 50 mM D-Biotin.
lego1_9499	20 mM HEPES pH7.4, 150 mM NaCl, 2 mM D-maltose.
lego2_9499	10 mM CHES pH9.0, 50 mM NaCl, 100 mM $MgCl_2$, 2 mM D-maltose.

4. Results

4.1 Virtual screening of hits against pUL77 using consensus docking, clustering and ranking pipeline

4.1.1 Consensus docking, clustering and ranking (DoCR) pipeline to identify hits

I developed a pipeline for consensus Docking, Clustering, and Ranking, termed 'DoCR.' DoCR integrates several modules written in Python, including docking (using AutoDock Vina and DiffDock), modeling (with AlphaFold3), filtering, clustering, scoring, and ranking.

Usage of AutoDock Vina allowed to dock molecules over the chosen hotspot residues, while DiffDock only facilitates global docking across the entire surface of the target protein. AlphaFold3 was utilized to model the protein-ligand complex. Combining local and global docking tools with a modeling tool – Alphafold3 – enhances the probability of identifying true positive hits via an alignment of poses derived from all three tools.

Docking Tools:

- **AutoDock Vina** utilizes a Monte-Carlo sampling algorithm combined with the BFGS gradient-based optimizer and an empirical scoring function to evaluate poses. The sampling space in Vina is defined by the user as a box around selected amino acid residues on the target protein [Eberhardt et al., 2021].
- **DiffDock** applies a diffusion generative AI model to generate docked poses and a deep-learning scoring model to evaluate them. It is a global docking tool, sampling small molecules across the entire protein surface. Although higher scores in DiffDock suggest better poses, they do not correspond to binding energy [Corso, Stark, Jing et al., 2023].

Superposition of poses obtained with the local docking tool (AutoDock Vina) and the global docking tool (DiffDock) directly compares the results from the two algorithms, thereby allowing to identify matching poses with an increased likelihood of an accurate pose.

DoCR clusters small molecules based on the similarity of their chemical substructures interacting with the hotspot residues, utilizing Butina and F-clustering algorithms. Additionally, DoCR scores molecules based on the average atom-to-atom root mean square distance across pairs of poses from the three tools and the proportion of the target interface covered by the selected residues.

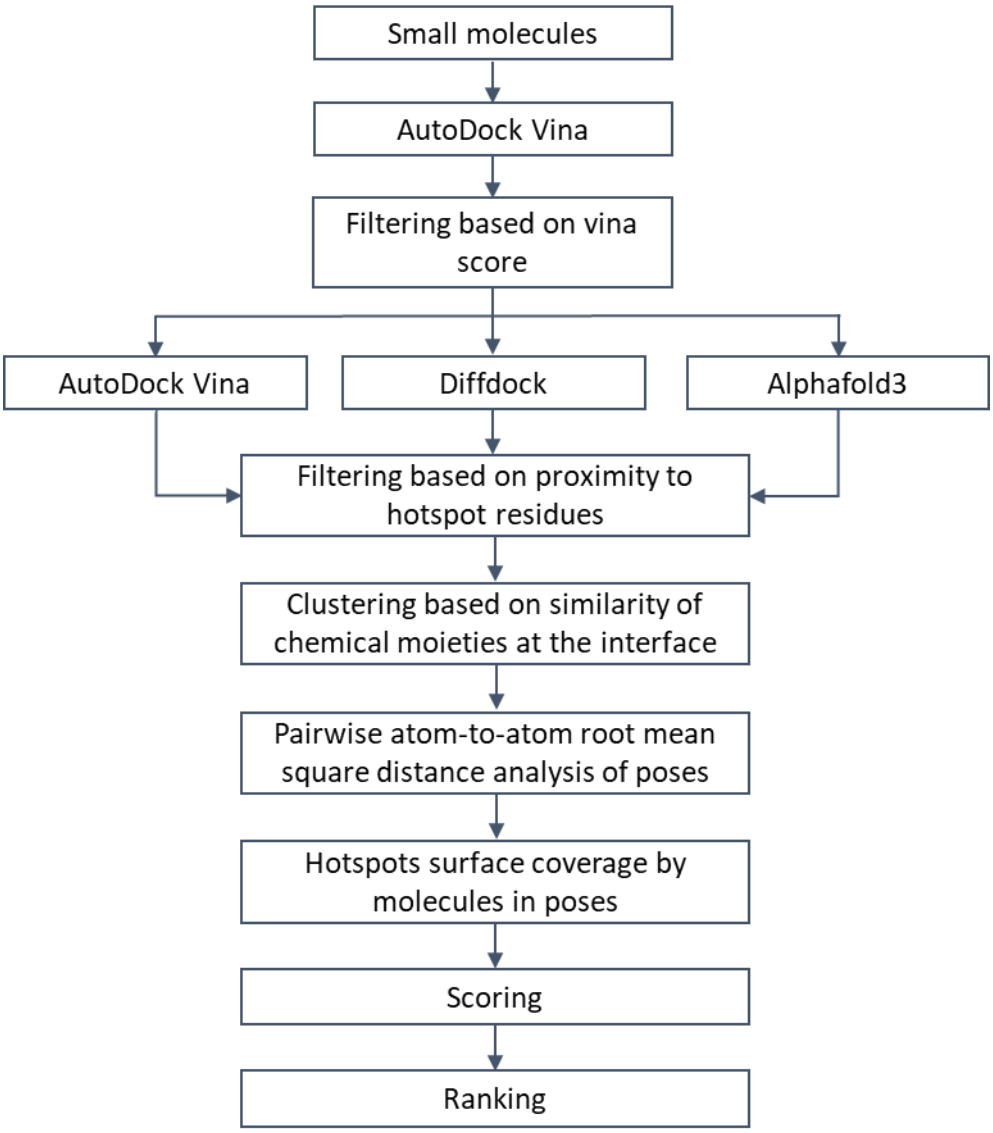


Figure 4.1.1: Workflow of DoCR pipeline

4.1.2 Python function in DoCR to identify hotspot residues

Protein-protein interactions (PPIs) are increasingly targeted for therapeutic interventions due to their critical roles in numerous diseases, including cancers, neurological disorders, and metabolic diseases. Estimates of the human proteome suggest the presence of over 600,000 PPIs [Stumpf et al., 2007; Sun and Zhao, 2010; Zhao et al., 2014]. Unlike traditional drug targets that feature well-defined binding pockets, PPIs present unique challenges. These include large interaction surfaces (1500 $Å^2$ to 3000 $Å^2$), hydrophobic interfaces, the absence of clear binding pockets, and shallow surfaces. Furthermore, the high affinity of protomer interactions often makes it difficult for small molecules to compete effectively. Additionally, the larger size of such molecules (often exceeding 500 Da) poses challenges in adhering to Lipinski's "rule of 5" [Lipinski et al., 2001; Arkin and Wells, 2004; Cheng et al., 2007; Buchwald, 2010; Smith and Gestwicki, 2012; Ivanov et al., 2013].

The discovery that certain residues, termed "hotspots," contribute significantly to the binding energy of PPIs [Shangary and Wang, 2009], has opened avenues for small

molecule inhibitors. These hotspots typically contribute more than 2 kcal/mol to binding energy [Thorn and Bogan, 2001; Moreira et al., 2007; Janin et al., 2008; Geppert et al., 2011]. Small molecules targeting these hotspots can effectively disrupt PPIs, creating a new class of therapeutic targets. Moreover, molecules binding to non-interface pockets (allosteric sites) can modulate interactions indirectly, resulting in two mechanisms of inhibition: orthosteric (direct hotspot interaction) and allosteric (binding to allosteric sites).

The DoCR pipeline includes a Python function designed to identify hotspot residues that mediate protein-protein interactions within a dimeric protein complex, utilizing the REF2015 score function from the PyRosetta package. The process involves the following steps:

1. **Interface Detection:** Interface atoms within a threshold distance of 8 Å are identified.

2. **Baseline Energy Calculation:** The potential energy of the entire dimer complex is computed to serve as a reference.

3. **Residue Analysis**: Each residue at the interface is individually analyzed by mutating all other interface residues to glycine, while keeping the residue of interest intact. The potential energy of this mutated complex is then calculated.

4. **Energy Difference Estimation:** The difference in potential energy between the wild-type dimer complex and the glycine-mutated complex is recorded for each interface residue. Residues causing the highest energy difference are considered hotspots, as they confer the most stability to the complex. While the potential energy difference does not provide absolute values, it allows for a relative comparison among residues.

Hotspots, in this context, are defined as residues with the lowest relative potential energy difference and are critical for complex stability. Figure 4.1.2 illustrates the Python functions used to mutate residues to glycine and estimate the potential energy of the complex.

```
def mutate_residue(pose, residue_index, new_residue):
    """Mutate a residue to a new amino acid in the given pose."""
    mutator = rosetta.protocols.simple_moves.MutateResidue(residue_index, new_residue)
    mutator.apply(pose)

def calculate_binding_energy_between_residues(pose, residue1, residue2, interface_residues,
score_function):
    """Calculate binding energy between two specified residues."""
    original_pose = pose.clone()
    original_score = score_function(original_pose)

    # Mutate all interface residues except residue1 and residue2 to glycine
    for resA, resB in interface_residues:
        if resA != residue1 and resA != residue2:
            mutate_residue(original_pose, resA, 'GLY')
        if resB != residue1 and resB != residue2:
            mutate_residue(original_pose, resB, 'GLY')

    mutated_score = score_function(original_pose)
    binding_energy_contribution = mutated_score - original_score
    return binding_energy_contribution

def analyze_binding_energies(pdb_file, chainA, chainB, pose):

    # Identify interface residues
    interface_residues = find_interface_residues(pose, chainA, chainB)
    print(f'Interface residues for chains {chainA} and {chainB}: {interface_residues}')

    # Define the REF2015 score function
    score_function = pyrosetta.create_score_function('ref2015')

    # Calculate binding energies for interface residues
    binding_energies = {}
    for resA, resB in interface_residues:
        binding_energy = calculate_binding_energy_between_residues(pose, resA, resB,
        interface_residues, score_function)
        binding_energies[(resA, resB)] = binding_energy

    return binding_energies
```

Figure 4.1.2: Python functions for mutating residues to glycine and estimating potential energies
This illustration showcases the Python functions used to mutate specific residues to glycine and estimate the potential energies of the mutated target protein dimer complex.

4.1.3 pUL77 has two non-overlapping regions with protein-protein interaction hotspots

Interface analysis of the pUL77 dimer, derived from the published pentameric structure of pUL77 (PDB: 8TEP), was conducted using the DoCR pipeline. An 8 Å distance threshold was applied to account for all potential long-range forces. The analysis identified 20 amino acid residues at the interface within this threshold distance, with the interactions localized into two clusters of amino acids (refer to Figure 4.1.3).

From the 20 interface residues, those exhibiting the lowest potential energy difference when all other interface residues were mutated to glycine were classified as ideal hotspot residues. The interface and hotspot analysis were performed using the *hotspot_analysis.py* script from the DoCR pipeline. In this analysis (see Figure 4.1.4), residue HIS196 showed the lowest binding energy difference among all residues.

Furthermore, residues ASN205, PHE208, and GLU209 from Cluster 1, along with residue GLN251 from Cluster 2, formed a shallow pocket with a surface area of 2500 $Å^2$, as calculated using PISA (see Figure 4.1.5). This shallow pocket represents part of the total protein surface area of approximately 19,586 $Å^2$.

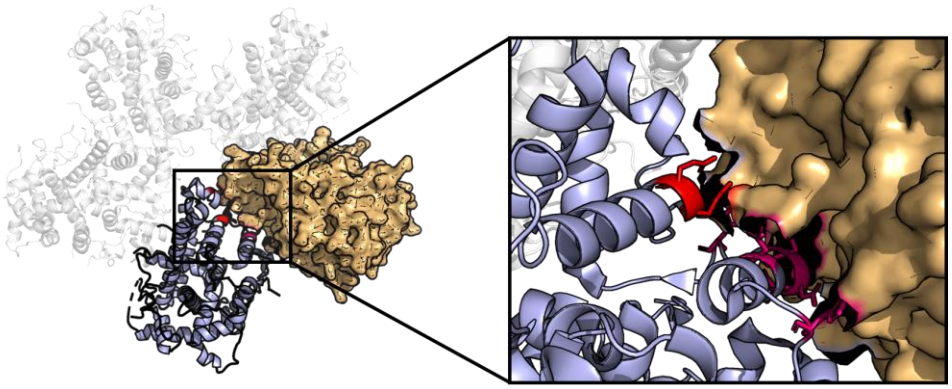


Figure 4.1.3: pUL77 pentameric structure (PDB: 8TEP) with interface residues highlighted
The pUL77 pentameric structure (PDB: 8TEP) is depicted, with interface residues in hotspot clusters highlighted. Residues within these critical clusters are colored in hotpink and red.

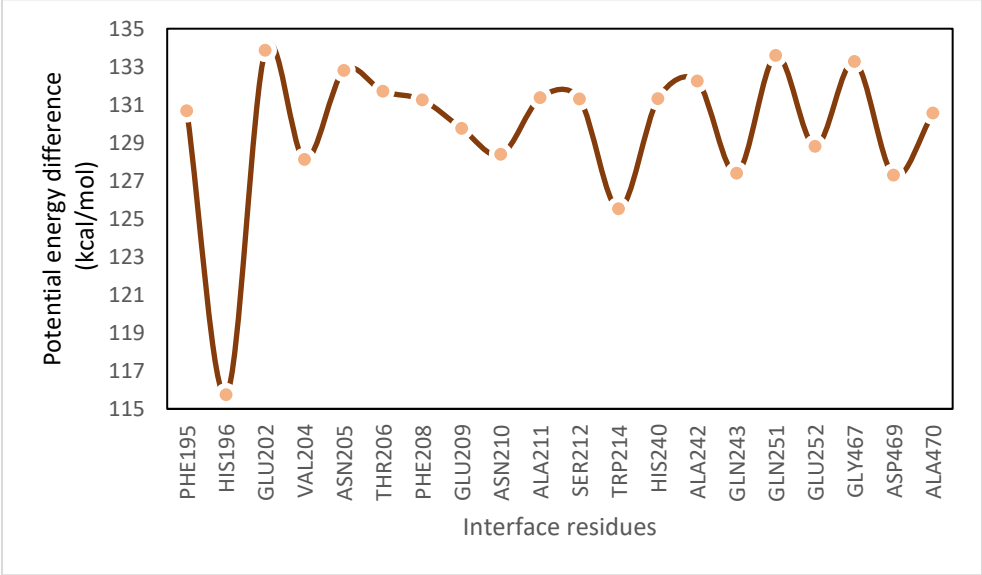


Figure 4.1.4: Plot of interface residues vs. potential energy difference
This plot showcases the relationship between the interface residues of the pUL77 dimeric structure and the potential energy differences (in kcal/mol) caused by mutating other interface residues to glycine.

Based on their potential energy differences and localization within the shallow pocket, residues ASN205, PHE208, GLU209, and GLN251 were selected for docking and subsequent downstream analysis.

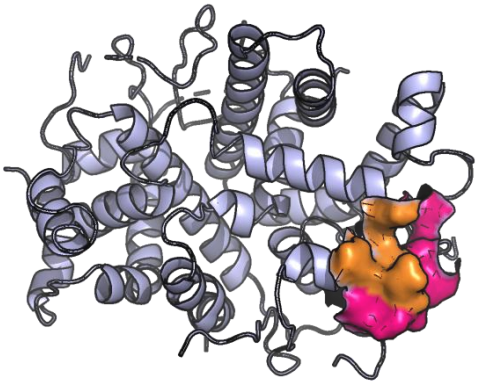


Figure 4.1.5: pUL77 (C-terminal) monomeric structure (PDB: 7NXP)
This figure highlights the surface area of the pUL77 (C-terminal) monomeric structure (PDB: 7NXP), with key regions colored in light orange and hotpink. The light orange area indicates a shallow pocket formed by residues ASN205, PHE208, GLU209, and GLN251.

4.1.4 Initial hits from AutoDock Vina

The pUL77 monomer crystal structure (PDB: 7NXP) was utilized for docking studies. The input PDBQT structure of pUL77 was prepared as outlined in Section 3.1 using AutoDock Tools. Grid parameters for AutoDock Vina were determined and are provided in Table 4.1.

A total of 1.1 million small molecules were sourced from the ZINC22 library, focusing on the subcategories of lead-like, readily purchasable molecules with neutral charge. These molecules were downloaded in SDF format. Using the *sdf_to_pdbqt.py* script from the DoCR pipeline, the small molecules were converted from SDF to PDBQT format for compatibility with AutoDock Vina.

Docking runs were performed on all 1.1 million small molecules in AutoDock Vina with an exhaustiveness setting of 8 and employing the vinardo scoring function. The docking workflow was executed using the bash script script_run_vina.sh. The results were analyzed to extract binding energies, and 6,830 small molecules with binding energies less than or equal to -6 kcal/mol were sorted for further investigation.

Table 4.1 Grid parameters for pUL77

Name of the parameter	**Value**
center_x	-53
center_y	75
center_z	6
size_x	28.0
size_y	26.0
size_z	30.0

4.1.5 Extensive screening of initial hits in AutoDock Vina and DiffDock

The sorted 6,830 molecules underwent docking in AutoDock Vina at an exhaustiveness setting of 32, utilizing the script script_run_vina.sh. All grid parameters, as described in Table 4.1, were kept unchanged, and the vinardo scoring function was applied.

These sorted molecules were converted back to SDF format in a two-step process: first by converting them to PDB format using the *pdbqt_to_pdb.py* script, followed by conversion to SDF format using the *pdb_to_sdf.py* script. Subsequently, the molecules were

docked globally using DiffDock with the script_run_diffdock.sh. The pUL77 monomer crystal structure, supplemented with polar hydrogen atoms in PDB format (prepared using AutoDock Tools, as detailed in Section 3.1), served as the target protein. The DoCR pipeline included a script to automate the DiffDock run by generating a CSV file that listed all 6,830 molecules alongside the target protein in the required format.

The top-ranked docking poses from AutoDock Vina and DiffDock were extracted using the script *script_extract_energy.py*, which exported the scores into a CSV file for downstream analysis.

4.1.6 Analysis for poses from Vina and DiffDock

To identify small molecules with similar poses, a binary vector was created for each molecule based on its interaction with the hotspot residues within a 4 Å distance from the pose using the script *pose_analysis.py*. A binary vector is an array of 0s and 1s, where 1 indicates an atom interacting with one of the hotspot residues, and 0 indicates no interaction.

These binary vectors were analyzed to sort small molecules that have at least four atoms interacting with the hotspot residues within a 4 Å distance using the script *pose_sorting.py*. This analysis identified 298 small molecules from DiffDock and 204 small molecules from AutoDock Vina, with 48 molecules common to both tools.

Visual inspection confirmed that these 48 molecules occupy the same pocket, albeit with varying degrees of angular and spatial deviation in their poses (see Figure 4.1.6). Nevertheless, these 48 molecules displayed overlapping poses and were considered consensus molecules.

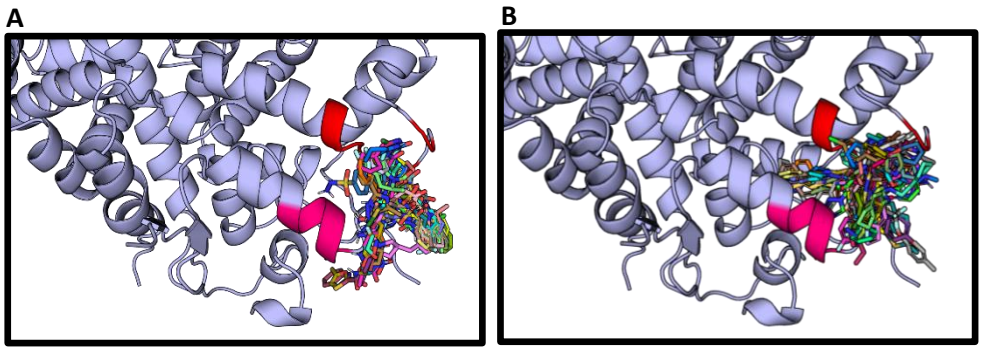


Figure 4.1.6: Poses of 48 molecules from Vina and DiffDock
Panel (A) displays the poses of 48 molecules as determined by AutoDock Vina, while Panel (B) shows the poses of the same 48 molecules as determined by DiffDock. Hotpink region refers to cluster 1 and red region refers to cluster 2.

4.1.7 Clustering of consensus hits

To group the hits based on the similarity of chemical moieties interacting with the hotspot residues of pUL77, the clustering program in the DoCR pipeline first extracted distinct chemical moieties or substructures from the entire molecule (see Figure 4.1.7 for illustration and Figure 4.1.8 for the Python function). The proximity of each substructure to the hotspot residues was estimated in terms of distance (in Ångstroms, Å).

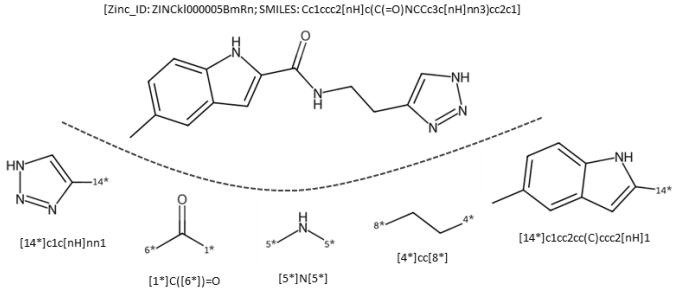


Figure 4.1.7: Fragmenting of one of the 48 molecules (ZINC_ID: ZINCkl000005BmRn)

Substructures interacting with the hotspot residues were fingerprinted using the RDFingerprint generator function from the RDKit package (see Figure 4.1.8 for the Python function). Fingerprinting is a computational method that encodes molecular characteristics such as atom types, functional groups, or substructures into binary vectors (or strings). In this pipeline, each fingerprint was represented by 1024 bits. Individual fingerprints of the substructures were then merged to generate a comprehensive fingerprint for each small molecule, ignoring substructures that did not interact with the hotspot residues.

```
def fragment_molecules(molecules):
    frags = BRICS.BreakBRICSBonds(molecules)
    fragments = Chem.GetMolFrags(frags, asMols=True)
    return fragments

def generate_fingerprint(fragments):
    generator = rdFingerprintGenerator.GetRDKitFPGenerator(maxPath=5, fpSize=4096)
    fingerprints = [generator.GetFingerprint(frag) for frag in fragments]
    return fingerprints
```

Figure 4.1.8: Python functions to generate substructures of molecules and to generate fingerprints of sub-structures

The fingerprints of each small molecule were used to calculate the Tanimoto distance coefficient. This coefficient was determined by generating a matrix of distance values for all possible combinations of small molecules using the formula shown in Figure 4.1.9 (see Figure 4.1.10 for the Python function).

$$\text{Tanimoto similarity coefficient} = \frac{\textit{No. of Similar Fingerprints}}{\textit{No. of total unique Fingerprints}}$$

$$\text{Tanimoto distance coefficient} = \textbf{(1 – Tanimoto similarity coefficient)}$$

Figure 4.1.9: Formula for tanimoto similarity and distance coefficient

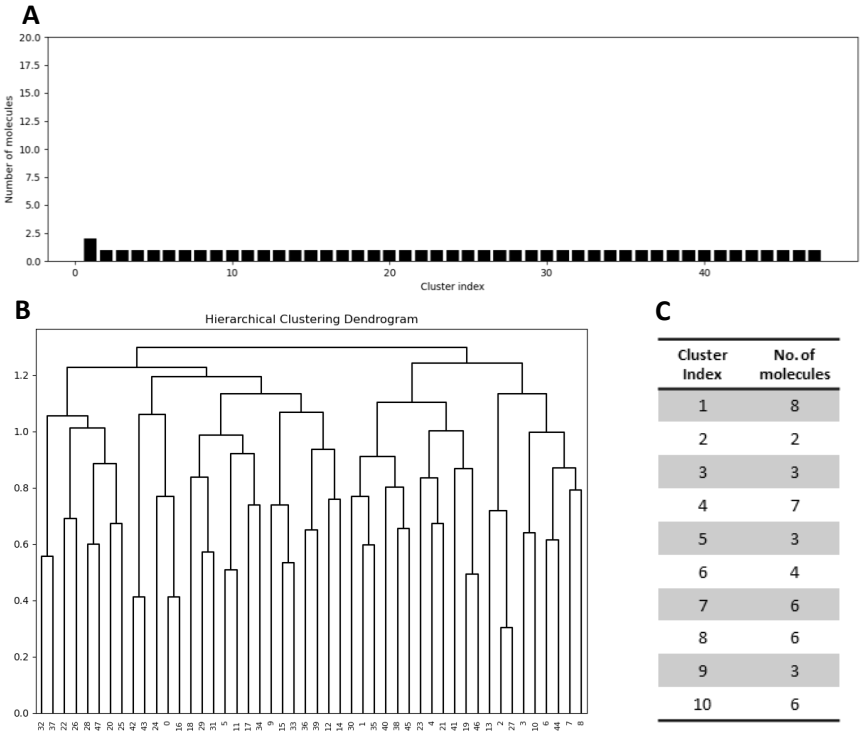
```python
def calculate_tanimoto_matrix(fp_list):
    dissimilarity_matrix = []
    for i in range(1, len(fp_list)):
        similarities = DataStructs.BulkTanimotoSimilarity(fp_list[i], fp_list[:i])
        dissimilarity_matrix.extend([1-x for x in similarities])
    return dissimilarity_matrix

def perform_hierarchical_clustering(matrix):
    linkage_matrix = linkage(matrix, method='ward')
    clusters = fcluster(linkage_matrix, t=10, criterion='maxclust')
    return clusters, linkage_matrix

def butina_clustering(tanimoto_distance_matrix, molecules_fingerprints, cutoff=0.2):
    clusters = Butina.ClusterData(tanimoto_distance_matrix, len(molecules_fingerprints),
    cutoff, isDistData=True)
    clusters = sorted(clusters, key=len, reverse=True)
    return clusters
```

Figure 4.1.10: Python functions to calculate tanimoto matrix and perform butina and F-clustering
This figure demonstrates the Python functions used to calculate the Tanimoto matrix and to perform clustering using Butina and F-clustering algorithms.

The Tanimoto distance coefficients were then employed for clustering small molecules based on their similarity using two algorithms: the Butina algorithm from the RDKit package and the F-clustering algorithm from the SciPy package. A Tanimoto distance cut-off of 0.4 was chosen for the Butina algorithm, based on performance analysis across threshold values ranging from 0.1 to 0.9. The cut-off value of 0.4 indicates that molecules with 60% or more similarity were grouped into the same cluster. The Butina clustering results revealed that 47 hits were distinct, while one hit shared 60% or more similarity with other hits (see Figure 4.1.11-A).

The F-clustering algorithm performs hierarchical linkage analysis using the previously calculated Tanimoto distance values. The number of clusters is an arbitrary input provided to the algorithm. For this analysis, the small molecule hits were grouped into 10 clusters, with the linkage cut-off automatically determined at 1.07 (see Figures 4.1.11-B and 4.1.11-C).

Clustering of the small molecule hits was executed using the *clustering.py* script from the DoCR pipeline. The script outputs Excel files listing the small molecules in each cluster, along with plots visualizing the results for both clustering algorithms.



Cluster Index	No. of molecules
1	8
2	2
3	3
4	7
5	3
6	4
7	6
8	6
9	3
10	6

Figure 4.1.11: Clustering analysis of molecules using butina and F-clustering methods
Panel (A) presents a bar graph showing the number of clusters versus the number of molecules in each cluster from Butina clustering. Panel (B) displays a hierarchical dendrogram from F-clustering, illustrating the linkage between molecules based on their index and linkage distance. Panel (C) provides a table of 48 molecules clustered into 10 clusters using F-clustering.

4.1.8 Pose prediction and analysis using Alpahfold3

All 48 consensus small molecules were analyzed using AlphaFold3 alongside the amino acid sequence of pUL77. For the first small molecule, an input FASTA file containing the pUL77 protein sequence and SMILES code was used. The multiple sequence alignment (MSA) generated from the first run was utilized for the subsequent AlphaFold3 runs. Results from AlphaFold3 were analyzed similarly to Section 4.1.6 (Analysis of poses from Vina and DiffDock) by employing the scripts *af3_pose_analysis.py* and *af3_pose_sorting.py*. These scripts relied on binary vectorization techniques to disregard predictions where the small molecule failed to interact with the desired hotspot residues of pUL77. This analysis revealed that all 48 molecules were predicted by AlphaFold3 to not interact with the hotspot residues of pUL77.

In another project, where the pipeline was used to screen for hits against a target protein from KSHV, 4,500 consensus hits were analyzed using AlphaFold3. This resulted in 430 predicted interactions that show similarities with the results obtained with Vina and DiffDock (data not shown here).

4.1.9 Orthogonal analysis of poses from Vina, DiffDock and Alphafold3 for scoring and ranking of hits

To extract small molecule coordinates from the AlphaFold3 results, a series of steps were performed. First, the results were converted from mmCIF to PDB format using

the script mmcif_to_pdb.py for easier analysis. Next, the results were superimposed with the pUL77 PDB structure using superimpose_pdb.py. Finally, the coordinates of small molecules were extracted using the script extract_chains.pdb. Since the coordinates of structures predicted by AlphaFold3 might differ even when using the same input protein sequence, superimposition corrects for these discrepancies, enabling analysis of poses from Vina and DiffDock.

SDF and PDBQT files of small molecule hits from Vina and DiffDock were converted to PDB format using the scripts sdf_to_pdb.py and pdbqt_to_pdb.py, respectively. This allowed streamlined analysis alongside the small molecule files generated by AlphaFold3.

A scoring model was developed based on two criteria: average atom-to-atom root mean squared distance (RMSD) and solvent-accessible surface area coverage of the small molecule hits (refer to Figure 4.1.12 for details on the scoring function). Average atom-to-atom RMSD was calculated for each small molecule by pairwise alignment of poses from Vina, DiffDock, and AlphaFold3. Additionally, solvent-accessible surface area coverage over the hotspot residues of pUL77 was estimated for all small molecule poses derived from Vina, DiffDock, and AlphaFold3.

$$\text{score} = \left(\sum_{x=1}^{N} \frac{2}{3N\left(\sqrt{1/n(\sum_{i=1}^{n}(a_i - a'_i)^2)}\right)} \right) + \sum_{x=1} \left(\frac{H}{12A}\right) vina, diffdock + \left(\frac{H}{6A}\right) af3$$

Figure 4.1.12: Scoring model to score the hits

N – no of tools used (3 – Vina, DiffDock and Alphafold3)
n – no. of atoms in the small molecule being analysed
α_i – Coordinates of atom from the first tool
α'_i – Coordinates of atom from the second tool
H – Solvent accessible surface area (in $Å^2$) of hotspot residues covered by small molecule
A – Total Solvent accessible surface area (in $Å^2$) of hotspot residues

The scoring model assigned two-thirds weight to the atom-to-atom RMSD term and one-third weight to the solvent-accessible surface area coverage term. Within the RMSD term, equal weight was distributed across all pairwise comparisons (Vina-DiffDock, DiffDock-AlphaFold3, and Vina-AlphaFold3). For the solvent-accessible surface area coverage term, one-fourth weight was allocated to poses from Vina and DiffDock, while AlphaFold3 predictions received one-half weight. The ideal poses, achieving 100% similarity, would yield a score of 1. The scores for hits identified against pUL77 ranged between 0.45 and 0.2, with higher-scoring small molecules ranked higher for shortlisting.

Since the predicted poses from Alphafold3 for small molecules with pUL77 protein were not consistent with results from Vina and Diffdock, the scores were calculated from the results from Vina and Diffdock. The highest-ranked small molecules within each F-cluster (based only on poses from Vina and Diffdock), a total of 16 small molecules, that are currently available for purchase have been ordered for functional studies through collaboration.

4.2 Structural characterization of pUL98

4.2.1 Multiple sequence alignment of pUL98 with orthologues and protein disorder prediction reveals intrinsic disordered region (IDR)

A multiple sequence alignment of the 584 residues of pUL98 from the AD169 strain (Uniprot ID: P16789), performed using Clustal Omega (v1.2.4) [Sievers and Higgins, 2017], revealed conservation of the putative active site residues (D254, E278, and K280) among orthologs of alkaline nucleases from the human herpesvirus family (highlighted in red in Figure 4.2.1–B). The residues from positions 88 to 119, which are not conserved among herpesviruses (refer to Figure 4.2.1–A), were designated as the non-conserved region (NCR). Additionally, residues from positions 400 to 440 in HCMV were found to be absent in other alkaline nucleases (refer to Figure 4.2.1–C).

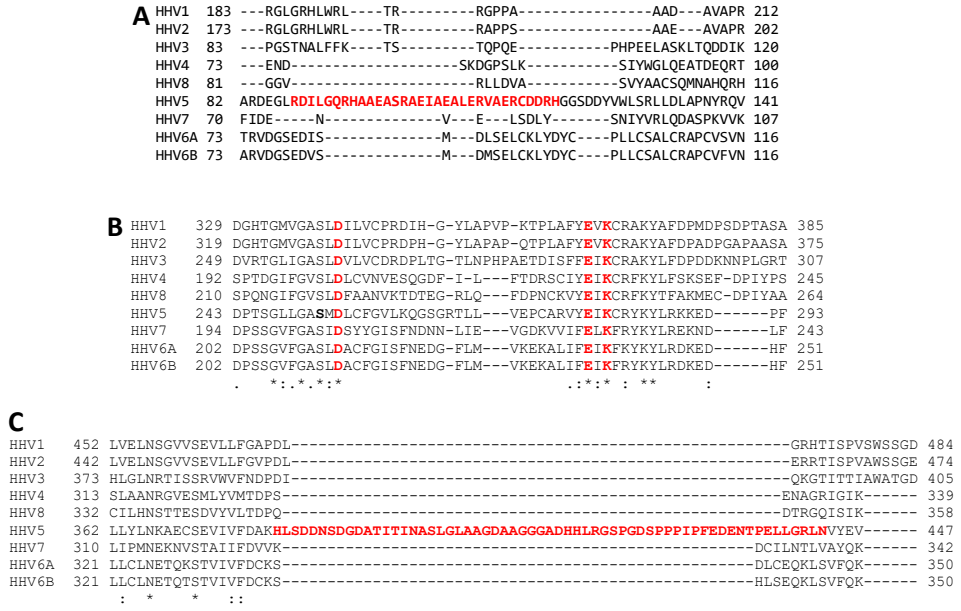


Figure 4.2.1: Multiple sequence alignment of pUL98 with other human Herpesviruses
The multiple sequence alignment of pUL98 was conducted using Clustal Omega. In panels (A) and (C), the amino acid sequences highlighted in red are not conserved among alkaline nucleases in other human herpesviruses. In panel (B), the residues highlighted in red constitute the putative active site residues, which are apparently conserved across alkaline nucleases of human herpesviruses.

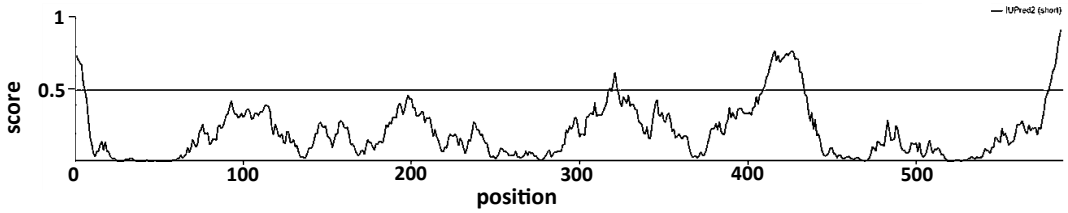


Figure 4.2.2: Bokeh plot of amino acid residues of pUL98 versus IUPred2

To gain further insights into the domains present in pUL98, the amino acid sequence was subjected to protein functional domain analysis using InterPro. The analysis confirmed the presence of the alkaline exonuclease domain and identified an intrinsic disordered region (IDR) between residues 410 and 430. This aligns with the stretch of

amino acids identified in the multiple sequence alignment that was absent in any orthologs and was thus labelled as an intrinsically disordered region (IDR).

Further evaluation using the intrinsic disorder prediction tool, IUPRED2, supported the findings from InterPro and confirmed the presence of an intrinsically disordered region within the same stretch. IUPRED2 predicted a score above 0.5 for the amino acid residues in this region, indicating intrinsic disorder (Chee et al., 1990; Meszaros et al., 2018; Erdos and Dosztanyi, 2020; Necci et al., 2021; Blum et al., 2024; Piovesan et al., 2024; The UniProt Consortium, 2024).

4.2.2 Design and cloning of construct without intrinsic disordered region

Long intrinsically disordered regions (IDRs) are generally detrimental to protein crystallization [Deller et al., 2015; Kojima et al., 2024]. Correlation between sequence analysis and published experimental structures of alkaline nuclease orthologs from Kaposi's sarcoma-associated herpesvirus (KSHV) Shutoff exonuclease (SOX) – PDB: 3POV, 3FHD, and 5HSW – and Epstein-Barr virus (EBV) alkaline exonuclease – PDB: 2W45 and 2W4B – revealed that the IDR is unlikely to influence or affect the remaining structural features of pUL98. Sequence and structural analyses further suggested that deleting the non-conserved region (NCR) might not compromise the integrity of the rest of the structure.

Table 4.2: Detailed overview of pUL98 constructs designed at the beginning of the project

Construct identifiers	**Details**	**Description**
Cons.1	N-His-TEV-UL98(aa1-aa584)-C	Wild type
Cons.2	N-His-TEV-UL98(aa1-aa584 [Δ88-119, §382-442:GGGS]-C	Intrinsically disordered region substituted with linker 'GGGS' and the non-conserved region is deleted
Cons.3	N-His-TEV-UL98(aa1-aa584 [Δ88-119, Δ379-417, Δ431-446]-C	Intrinsically disordered region substituted with 14 residues from the middle of itself and the non-conserved region is deleted
Cons.1g	N-His-TEV-UL98(aa1-aa584)[§382-442:GGGS]-C	Only the intrinsic disordered region is deleted
Cons.4	N-MBP-UL98(aa157-aa584)[§382-442:GGGS]-TEV-mNeonGreen-6xHis-C	Maltose-binding protein (MBP) fusion at amino terminus of truncated UL98 gene without IDR with mNeonGreen fluorescent protein at carboxy terminus.

* Δ – Deletion, § - Substitution

Consequently, five gene constructs of pUL98 were initially generated for recombinant expression in Escherichia coli (E. coli). Constructs 1, 2, and 3 were originally designed by Dr. Louisa Ströh. The details of these constructs are provided in Table 4.2. Synthetic genes for constructs 1, 2, and 3 were already available at the beginning of this project. The initial construct contained a thrombin cleavage site, which I substituted with a Tobacco Etch Virus (TEV) protease cleavage site using primers Thr2Tev_98 (forward and reverse) listed in Table 3.3, following the procedure described in Section 3.9 with an annealing temperature of 58°C. Initially, only constructs 1, 2, and 3 were cloned.

As the project progressed, additional constructs were deemed necessary due to reasons outlined in subsequent subsections, prompting cloning of the remaining constructs described in Table 4.2. Construct 1g was cloned via substitution PCR using primers 98_IDR, following the method outlined in Section 3.9 at an annealing temperature of 56°C. Construct 4 was cloned using a restriction-free (RF) cloning strategy in two steps: first, cloning the UL98 construct 1g gene into a plasmid containing the gene for mNeonGreen using primers 98_mNeon, and subsequently cloning the UL98 gene with mNeon into a vector containing MBP using primers 98_mbp, as described in Section 3.8.

The plasmids were introduced into DH5α competent cells following the procedure in Section 3.6, and isolated as described in Section 3.7. They were then sent for sequencing to confirm successful cloning, using the Sanger sequencing method outsourced to LGC Genomics. The sequencing results were analyzed using the sequence analysis tool Benchling.com and confirmed that the constructs were correct.

4.2.3 Expression of pUL98 in E.coli

pUL98 constructs 1, 2, 3, and 4 were transformed into Rosetta2(DE3)pLysS cells following the procedure outlined in Section 3.6. The expression of these constructs was carried out as described in Section 3.12 using kanamycin at a final concentration of 50 µg/mL. The expression parameters mentioned in Section 3.12 were optimized after multiple trials involving variations in temperature, incubation time post-induction, induction optical density (OD), and IPTG concentration. Expression induced with 1 mM IPTG at 0.8 OD of the culture, followed by incubation in shaker incubator at 18°C for 16 hours.

pUL98 construct 1g was transformed into Lemo21(DE3) cells following the method in Section 3.6 and expressed as detailed in Section 3.13. Expression was conducted using ZYP5052 autoinduction media supplemented with kanamycin (50 µg/mL).

ZYP-5052 autoinduction media is designed to automatically induce recombinant protein expression in *E. coli* without the need for IPTG. It works through diauxic growth, where bacteria first consume glucose for rapid growth, then switch to lactose metabolism, which activates the lac promoter and induces protein expression. This method allows for high protein yields without requiring constant monitoring

4.2.4 Affinity purification of pUL98

All constructs of pUL98 were designed with a poly-histidine affinity tag at the amino terminus. After expression, the cell pellets were lysed as described in Section 3.18, using the buffer 98_lysis1 (refer to Table 3.8), and purified through Ni-affinity chromatography following the protocol outlined in Section 3.20. The purification process utilized buffers 98_lysis1, 98_wash1, and 98_elution1, as listed in Table 3.8.

In addition to the wash step described in Section 3.18, a urea wash was performed using buffer 98_urea (60 mL) to remove impurities bound to the protein after the loading step. Elution was carried out using a gradient, up to 100% 98_elution1 buffer and 98_wash1 buffer. Following Ni-affinity chromatography, the protein underwent polyhistidine tag cleavage using TEV protease, as detailed in Section 3.34. Figure 4.2.3 shows nickel affinity chromatography and SDS-PAGE gel for pUL98 cons.1 protein.

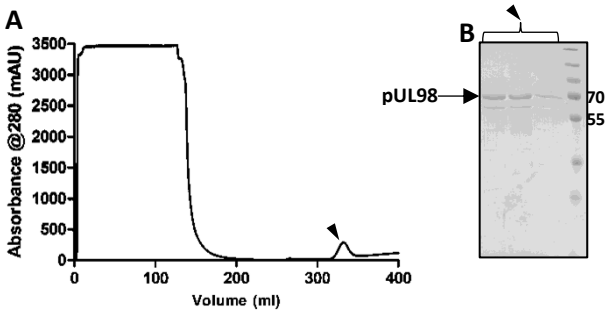


Figure 4.2.3: Nickel affinity chromatogram and SDS-PAGE Gel for pUL98 Cons.1 protein
Panel (A) shows Nickel affinity chromatogram for pUL98 cons.1 protein. Panel (B) 15% SDS-PAGE Gel loaded with fractions from nickel affinity chromatography.

4.2.5 Dimerization equilibrium of pUL98 protein

The affinity-tag-free protein solutions were concentrated to 10 mL at a concentration of 1 mg/mL and injected into SEC using a Superdex 200 26/600 column, as described in Section 3.22, with 98_sec buffer (refer to Table 3.8).

The optimized purification buffer compositions were determined after multiple trials of soluble protein purification. For constructs 2 and 3, a single peak was observed in the SEC chromatogram, corresponding to a molecular weight greater than 600 kDa when compared with standards. For constructs 1, 1g, and 4, three major peaks were observed in the SEC chromatogram. Based on the column calibration data, the first peak was determined to have a molecular weight exceeding 600 kDa, while the second and third peaks represented the dimer and monomer populations of pUL98, respectively. Figure 4.2.4 (A and B) illustrates the SEC chromatogram for the pUL98 construct 1 protein (molecular weight: 67 kDa) containing three peaks, along with the SDS-PAGE gel where pUL98 construct 1 protein was analyzed.

The area under the curve for the second and third peaks varied across all constructs and between different batches of the same pUL98 construct (results not shown). Attempts to disrupt dimerization by increasing salt concentrations and using detergents were unsuccessful. When the protein sample was injected at concentrations higher than 1 mg/mL into SEC, the dimer population increased relative to the monomer population. The molecular weights of the protein populations in the monomer and dimer peaks were confirmed using SEC-MALS (refer to Figure 4.2.4–C). SEC-MALS also revealed that, at an injection concentration of 4 mg/mL, the proportion of the dimer population was significantly higher than that of the monomer population. Protein injection into SEC-MALS was performed following the procedure described in Section 3.33.

Table 4.3 presents the protein yield obtained from 1 liter of culture.

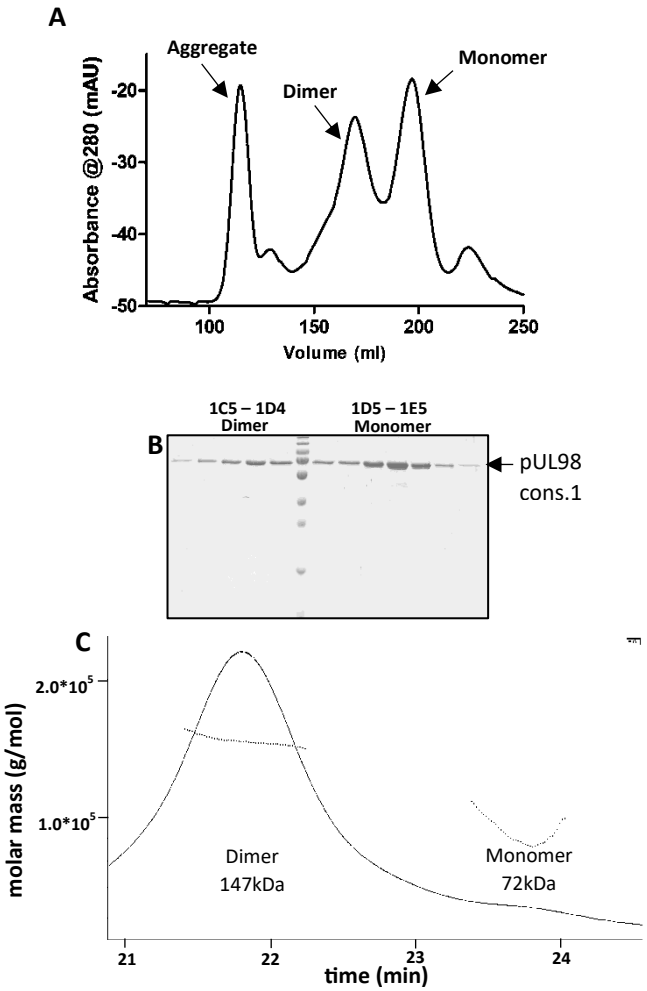


Figure 4.2.4: SEC chromatogram and SDS-PAGE Gel for pUL98 Cons.1 protein
Panel (A) shows SEC chromatogram with aggregate, monomer and dimer populations of pUL98 cons.1. Panel (B) 15% SDS-PAGE Gel loaded with fractions from both monomer and dimer peaks. Panel (C) shows chromatogram and molar mass for two peaks (monomer and dimer) from SEC-MALS with their corresponding molecular weight.

Table 4.3 Yield of protein constructs of pUL98 (monomer)	
Constructs	**Yield from 1L of culture**
Cons.1	0.4mg
Cons.2	0.25mg
Cons.3	0.25mg
Cons.1g	0.4mg
Cons.4	0.4mg

4.2.6 NCR deletion impacts pUL98 protein folding but not enzymatic activity

As described in Section 4.2.5, constructs 2 and 3 exhibited a single peak at a volume corresponding to a molecular weight greater than 600 kDa, based on calibration standards from the chromatogram. The protein from construct 3 was analyzed using SEC-MALS, following the procedure outlined in Section 3.33. The results showed two peaks, with the molecular weight of the first peak determined to be 1.7 MDa and the second peak measured at 837 kDa (refer to Figure 4.2.5). Additionally, the second peak accounted for more than 90% of the total protein.

A low-resolution X-ray diffraction study revealed the presence of a large ring-shaped GroEL complex (data not shown), which correlates with the observed molecular weight of 837 kDa in SEC-MALS. However, the protein was found to retain enzymatic activity in a fluorescence-based assay (conducted through collaboration; data not shown). The co-purification of GroEL chaperone protein with the enzymatically active pUL98 construct 3 suggests that, while retaining enzymatic activity (despite potential changes in the activity rate), the protein was either partially or improperly folded and is thus associated with the GroEL chaperone.

Constructs 2 and 3 are highly similar, differing only in the linker sequence used to replace the IDR. Based on these observations, it is reasonable to hypothesize that the deletion of the NCR negatively impacts protein folding. However, neither the deletion of the NCR nor the deletion of the IDR entirely abolishes the enzymatic activity of pUL98.

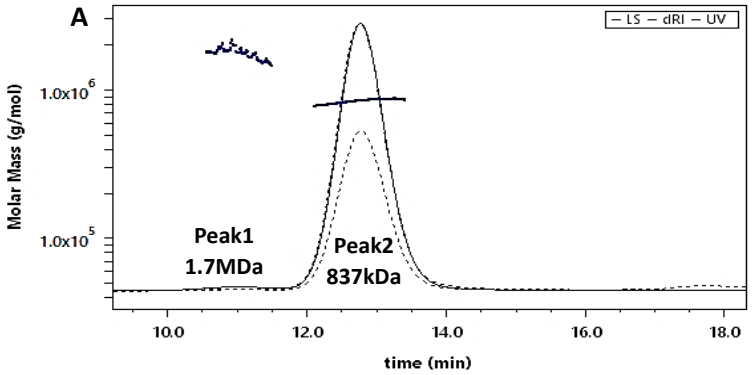


Figure 4.2.5: SEC-MALS analysis of pUL98 cons.3
Panel (A) presents the results from Size exclusion chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) for the pUL98 cons.3 protein.

4.2.7 The IDR is not essential for dimerization

To estimate the molecular weight of two peaks from SEC run with protein pUL98 cons.1g, the sample was injected into SEC-MALS as described in section 3.33. The result (see figure 4.2.6 - A) revealed monomer dimer equilibrium (similar to the results described in section 4.2.5). Although the dimer population was less in proportion compared to monomer population, the presence of dimer population suggests that the IDR is not essential for dimerization of pUL98. This was further confirmed with dynamic light scattering (DLS) experiment with sample at 10 mg/mL concentration where the radius distribution plot revealed the hydrodynamic radius of 4.89 nm corresponding to molecular weight of 129kDa suggesting that the majority of the population are dimers at this concentration and that the pUL98 construct without IDR forms dimers.

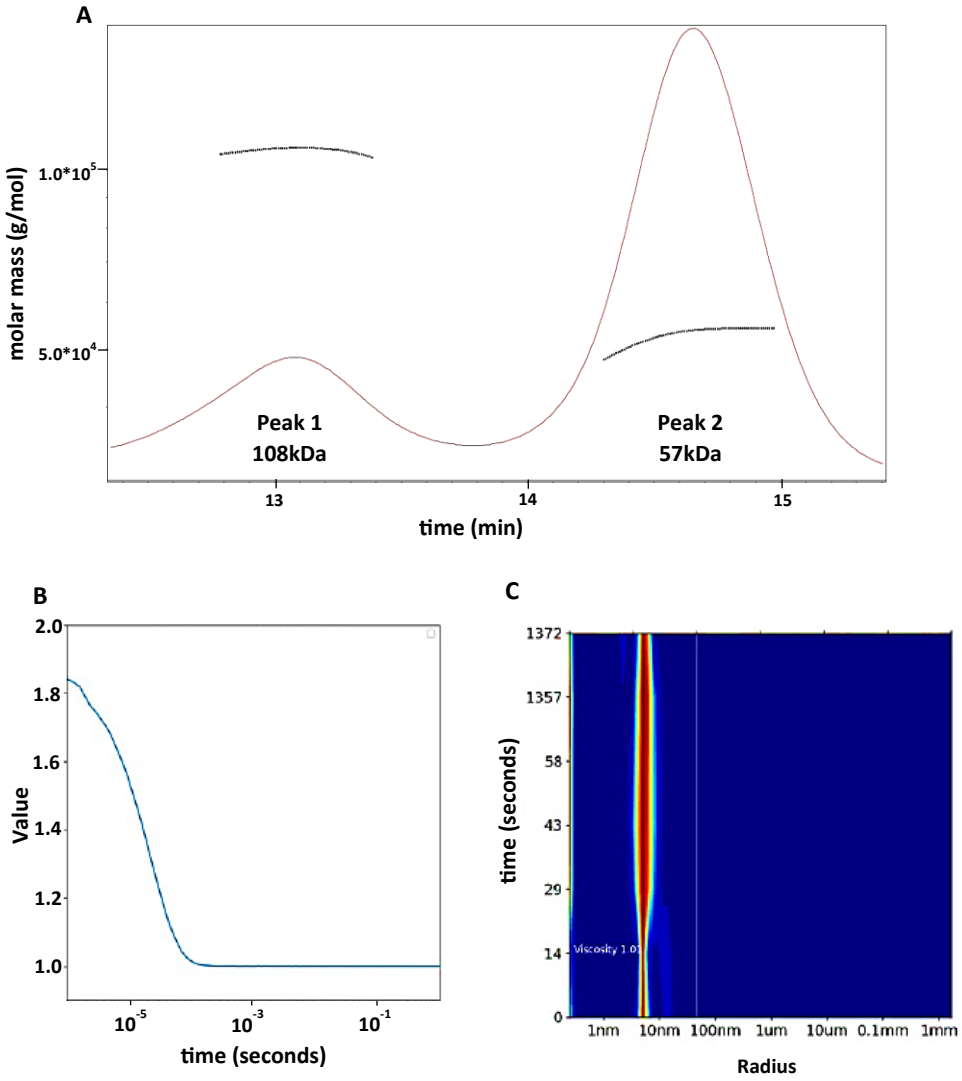


Figure 4.2.6: Analysis of pUL98 cons.1g using SEC-MALS and DLS
Panel (A) presents the SEC-MALS result for pUL98 cons.1g, showcasing two distinct peaks for monomer and dimer populations. Panel (B) is the autocorrelation function for the DLS raw data. Panel (C) is the radius distribution plot showing the distribution of hydrodynamic radius over time.

4.2.8 Crystallization of pUL98

Extensive crystallization trials were conducted for pUL98 constructs 1, 1g, and 4. Protein concentrations ranging from 2 mg/mL to 19 mg/mL were tested, along with heterologous random microseed matrix screening (rMMS) using seeds from lysozyme crystals at various dilutions and different buffer conditions. These trials primarily utilized the sitting-drop vapor diffusion crystallization technique. Commercial crystallization screens from Jena Biosciences were employed to identify potential conditions for crystal formation.

Despite all these efforts, no diffraction-quality crystals were obtained for any of the three constructs.

4.2.9 Thermal shift assay to determine pUL98 stability with oligos

Attempts to crystallize pUL98 constructs 1, 1g and 4 were not successful. As an alternative approach, co-crystallization of pUL98 with DNA oligonucleotides was

attempted, since oligonucleotides are substrates of nucleases. To screen for oligonucleotides that bind to and stabilise pUL98, thermal shift assay was performed for pUL98 protein in the presence of oligonucleotides. Two pairs of oligonucleotides were designed (refer to Table 4.4), with the oligonucleotides in each pair being complementary to one another. All four single-stranded oligonucleotides and two double-stranded oligonucleotides were incubated with pUL98 construct 1 protein in the presence of 2 mM $MgCl_2$ at a 1:1 ratio (3 µM each) in a total volume of 28.8 µL for 1 hour.

A thermal shift assay was performed as detailed in Section 3.30, and the resulting melting temperatures for each oligonucleotide were plotted in Figure 4.2.7. The melting temperature of pUL98 construct 1 protein alone with 2 mM $MgCl_2$ was recorded as 48°C. The results indicated that samples with double-stranded oligonucleotides exhibited a melting temperature at least 2°C higher than their corresponding single-stranded oligonucleotides. Furthermore, the single-stranded oligonucleotides displayed a minimum increase of 2°C in melting temperature.

However, this experiment did not account for factors such as the duration required for oligonucleotides to be cleaved by pUL98, if cleavage would occur.

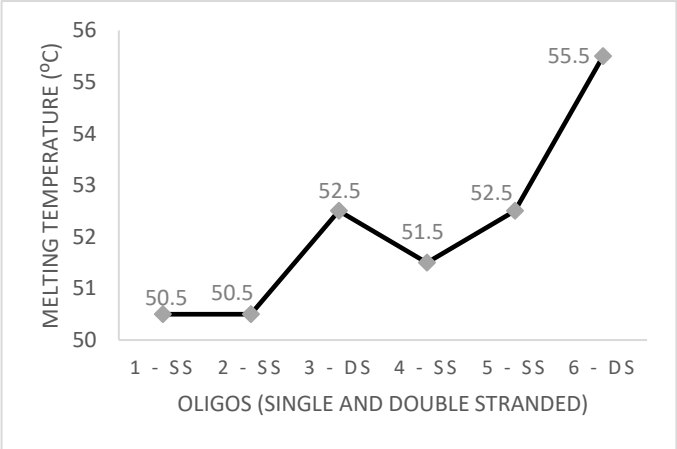


Figure 4.2.7: Melting temperature of pUL98 cons.1 in complex with different oligonucleotides

Table 4.4 List of single stranded oligos tested against pUL98	
Oligo no.	**Sequence (5' to 3')**
1	GGGGATCCTCCCAGTCGACC
2	GGTCGACTAGGAGGATCCCC
4	CTACGAGATCAAGTGCCGCTAC
5	GTAGCGGCACTTAATCTCGTAG

4.2.10 Co-crystallization of pUL98 cons.1 with oligo6

The thermal shift assay of pUL98 construct 1 with oligo6 revealed an increase in melting temperature by 7.5°C compared to pUL98 alone. This finding prompted me to attempt co-crystallization of pUL98 with oligo6 and MgCl2. A concentration of 75 µM oligo6 was incubated with 75 µM pUL98 (5 mg/mL) for 30 minutes before setting up sitting-drop vapor diffusion crystallization plates. Commercial screens from Jena Biosciences were employed to identify potential crystallization conditions. Crystals were obtained under the condition: 40% MPD, 100 mM Tris (pH 8.5), and 200 mM NH4Cl.

However, these crystals failed to diffract at both low and high oscillations. Extensive optimization efforts yielded crystals of similar morphology (refer to Figure 4.2.8), but they did not diffract X-rays in an X-ray diffraction experiment. Consequently, the strategy was shifted to using crystallization chaperones.

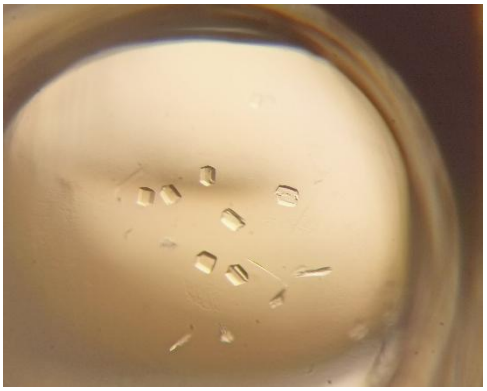


Figure 4.2.8: Crystal hits for pUL98 cons.1 with oligo2 at 40% MPD, 100mM Tris pH8.5, 200mM NH4Cl

4.2.11 *In-situ* proteolysis crystallization of pUL98

The *in-situ* proteolysis crystallization method involves adding trace amounts of proteases to the protein drop during crystallization. This approach aims to cleave flexible regions of the protein using suitable proteases [Dong et al., 2007]. To identify a suitable protease for in situ proteolysis crystallization, I tested five different proteases for limited proteolysis of pUL98 construct 1 protein, aiming to identify a stable domain resistant to trace amounts of proteases. The testing was performed using SDS-PAGE, following the protocol described in Section 3.31, with the proteases thermolysin, proteinase K, streptomyces protease, licheniformis protease, and papain.

Figure 4.2.9 displays the SDS-PAGE gel results, revealing two intense bands after digestion with thermolysin at concentrations below 0.01 mg/mL. In contrast, the other proteases produced multiple bands of relatively similar intensities, suggesting that thermolysin partially digests pUL98 construct 1 while yielding two stable domains at the tested concentration. This observation was made after 1 hour of incubation. The expected outcome was one intense band in the SDS-PAGE gel, representing a single stable domain.

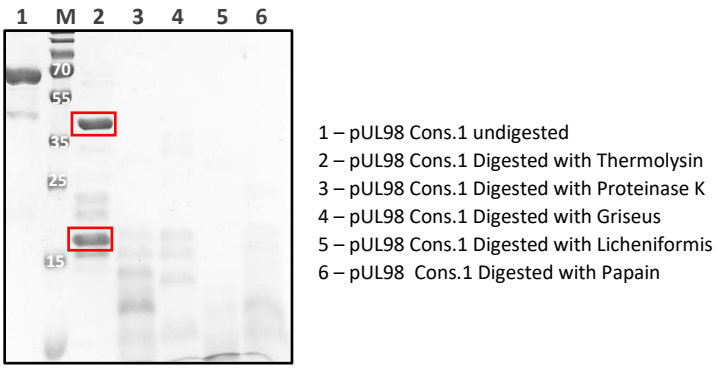


Figure 4.2.9: Identification of stable domains in pUL98 cons.1 protein through limited proteolysis with various proteases. Bands corresponding to stable pUL98 domains are highlighted with a red box

Despite the presence of two bands in after digestion in SDS-PAGE gel, thermolysin at final concentrations of 0.1 mg/mL and 0.05 mg/mL was added to pUL98 construct 1 protein (at a concentration of 12 mg/mL) in crystallization screening using sparse matrix crystallization screens from Jena Biosciences. Unfortunately, these attempts did not result in the formation of crystals.

4.2.12 Surface lysine methylation

Methylation of the primary amine group of lysine side chains has been shown to rescue crystallization in certain cases [Walter et al., 2006; Sledz et al., 2010]. pUL98 construct 1 protein was methylated following the protocol detailed in Section 3.32. The methylated pUL98 construct 1 protein was then concentrated and subjected to crystallization trials at various concentrations. However, no crystals were obtained from these attempts.

4.2.13 Biotinylation of pUL98

pUL98 construct 1 protein was biotinylated following the protocol outlined in Section 3.24, using 700 µL of 1.959 mg/mL protein solution and 27.4 µL of 20 mM EZ-Link NHS-PEG4-Biotin.

The extent of biotinylation was estimated using the HABA assay, as described in Section 3.25. It was determined that 1.8 moles of biotin were present per mole of protein, resulting in a biotin-to-protein ratio of 1.8:1. The biotinylated pUL98 construct 1 protein was prepared for use in the screening of sybodies against pUL98.

4.2.14 Screening of sybodies against pUL98

Following unsuccessful attempts at crystallization of the pUL98 protein, sybodies were screened against pUL98 to serve as crystallization chaperones and enhance its crystallization. Sybodies are synthetic nanobodies, originally described in 2018 by Zimmerman et al. Nanobodies are variable heavy-chain-only (VHH) domains derived from special heavy-chain-only antibodies found in camelids and certain shark species [Feng et al., 2025; Jin et al., 2023]. The traditional method of generating nanobodies involves immunizing camelids with the target protein, followed by extracting B-cells encoding heavy-chain-only antibodies [Pardon et al., 2014]. This approach is both time-consuming and expensive.

To overcome these challenges, the Seeger lab developed a platform for screening synthetic nanobodies, termed sybodies, from synthetically generated mRNA libraries consisting of 10^{12} nanobodies. This platform significantly reduces the timeline for generating sybodies to just 4 weeks [Zimmerman et al., 2018].

I received training from Dr. Tânia Custodio in Dr. Christian Löw's lab on this platform for screening sybodies against pUL98. Screening was conducted as described in Section 3.26, using three libraries of sybodies (concave, loop, and convex). Following ELISA screening, 14, 12, and 10 sybodies were shortlisted from the concave, loop, and convex libraries, respectively, for analysis in a Sepax SRT-C SEC100 column (7.8 mm x 300 mm, catalogue no. 215100-7830 from Sephax Technologies). Monomeric sybodies eluted at 7 mL, while oligomeric sybodies eluted earlier. SEC runs identified 23 out of 36 sybodies to be monomeric.

Single-point kinetic data were measured to determine koff values for interactions between sybodies and pUL98 construct 1 protein, as detailed in Section 3.27. The eight

sybodies with the lowest koff values were shortlisted (refer to Table 4.5). A summary of sybody enrichment during the screening process is provided in Table 4.6. Sybodies with single-point apparent KD values of less than 100 nM were successfully obtained during the screening process.

Table 4.5 Single-point binding kinetics for sybodies and pUL98 interaction

SYBODY NO	REFERENCE	LIBRARY	k_{OFF} (1/S)	k_{ON} (1/MS)	KD (nM)
9	SE1	Small	3.212	204700	15.69
10	SE11	Small	2.082	84010	24.78
11	SF2	Small	0.1129	216000	0.5225
14	SH1	Small	1.466	183200	8.004
19	MC12	Medium	9.574	121300	78.91
22	MF2	Medium	9.058	103600	87.47
30	LH7	Long	9.251	132800	69.64
31	LH8	Long	0.007988	11550	0.6913

Table 4.6 Summary of sybody selection against pUL98

Description	Concave (or small)	Loop (or medium)	Convex (or Long)
Enrichment after phage display round 1	6.8	4.3	3.3
Enrichment after phage display round 2	448.4	115.7	306.7
No. of positive ELISA Hits (>1.5 fold signal over control) out of total no. of sybodies analysed	89/93	84/93	54/93
No. of sybodies with unique sequence of CDRs out of total no of sybodies sequenced	14/14	9/12	9/10
No. of monomeric sybodies identified using analytical size exclusion chromatography (Sepax 10 column) out of total no. of sybodies analyzed	11/14	4/9	8/9
No. of sybodies that are shortlisted based on single point kinetic data from bio-layer interferometry (BLI) out of total no. of monomeric sybodies	4/11	2/4	2/8
No. of unique antigenic regions that sybodies from each library bind to, based on cross-competition assay performed using BLI, out of total no. of sybodies shortlisted from each library	1/4	1/2	2/2

4.2.15 Expression and purification of sybodies

Sybodies were expressed as detailed in Section 3.35. The resulting cell pellet was resuspended in sybody_lysis buffer and lysed using a sonicator, as described in Section 3.18. Ni-affinity chromatography was performed on the lysate following the procedure outlined in Section 3.20, using sybody_lysis, sybody_wash, and sybody_elution buffers. Elution from the HisTrap FF column was conducted as a gradient, with imidazole concentrations increasing up to 100%. The eluted sybodies were concentrated to a final volume of 10 mL and subsequently injected into a SEC Superdex 75 26/600 column, as described in Section 3.22. The sybodies eluted at a volume of approximately 200 mL. Figure 4.2.10 shows ninta chromatogram, SEC chromatogram and SDS-PAGE gel loaded with protein from peak fraction of the SEC.

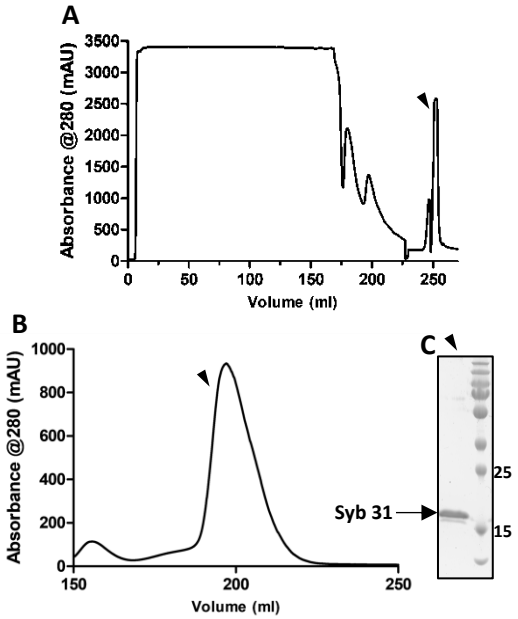


Figure 4.2.10: Nickel affinity chromatogram, SEC chromatogram and SDS-PAGE Gel for Sybody 31 that bind to pUL98

Panel (A) shows Nickel affinity chromatogram for Sybody 31 protein. Panel (B) shows SEC chromatogram for Sybody 31 protein. Panel (B) 15% SDS-PAGE Gel loaded with a fraction from SEC.

4.2.16 Cross-competition assay to categorise sybodies binding non-overlapping antigenic regions of pUL98

A cross-competition assay was performed using Biolayer Interferometry (BLI) to identify sybodies that recognize and bind to non-overlapping antigenic regions of the pUL98 protein. The assay revealed four distinct non-overlapping antigenic regions, each recognized and bound by one of the eight shortlisted sybodies. Based on their binding specificity to these antigenic regions, the sybodies were categorized and binned (refer to Table 4.7).

During the screening process, each sybody's ability to bind the target protein in the presence of other sybodies was assessed, allowing the determination of unique

binding sites. This procedure was conducted following the method described in Section 3.28.

Table 4.7 Sybodies grouped based on binding to unique non-overlapping antigenic regions of pUL98

Group A	Group B	Group C	Group D
9	19	30	31
10	22		
11			
14			

4.2.17 Sybodies binding pUL98 cons.1g

Sybody screening was initially performed with pUL98 construct 1 (wild-type protein). To identify sybodies that bind to pUL98 construct 1g, the pUL98 construct 1g protein was incubated with sybodies 9, 19, 30, and 31 (one sybody from each group identified in Section 4.2.15) at a 1:2 molar ratio of protein to sybodies. The mixture was injected into a Superdex 200 Increase 10/300 GL column, following the procedure described in Section 3.23. The eluted proteins were then analyzed on a 15% SDS-PAGE gel (refer to Figure 4.2.11–B).

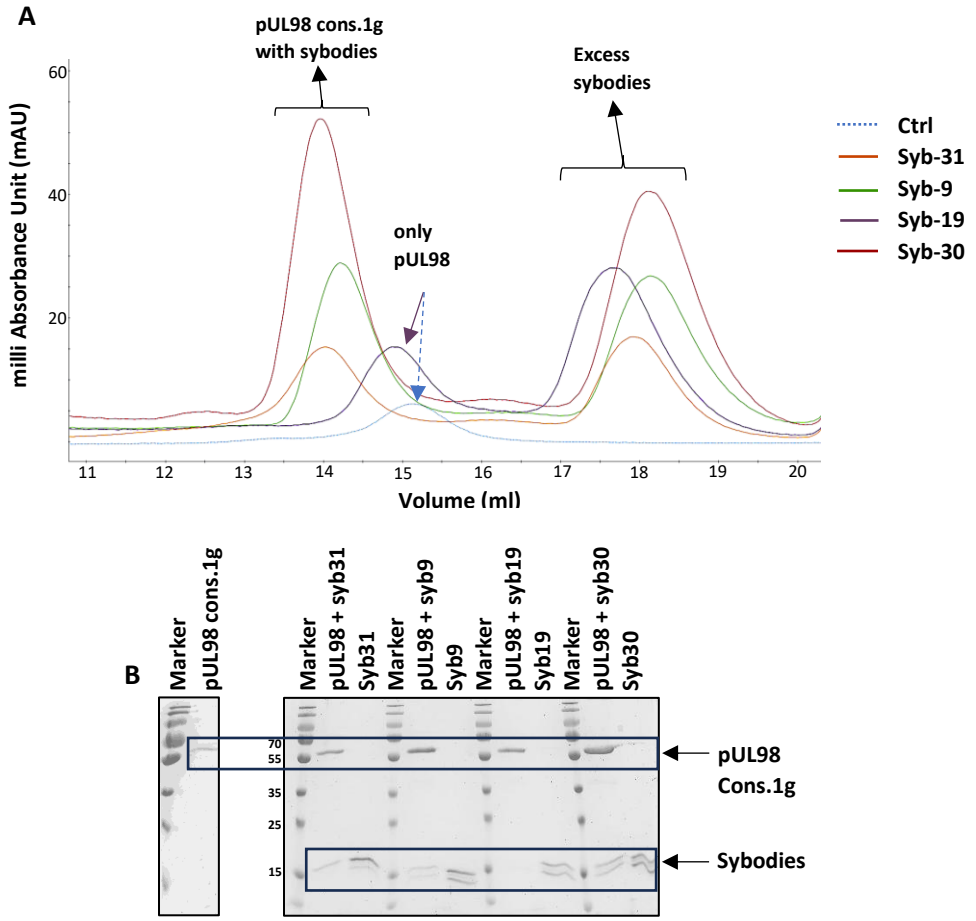


Figure 4.2.11: SEC chromatogram and SDS-PAGE Gel for pUL98 Cons.1 protein in complex with sybodies

Panel (A) shows overly of SEC chromatograms of pUL98 cons.1 with sybodies 9, 19, 30, 31 and pUL98 cons.1g alone. Panel (B) shows the SDS-PAGE gel with proteins from all the peaks from chromatogram shown in panel (A).

In the SEC chromatogram (Figure 4.2.11–A), peaks corresponding to pUL98 complexes with sybodies 9, 30, and 31 were observed between 14 mL and 14.3 mL, whereas peaks corresponding to pUL98 construct 1g alone and in the presence of sybody 19 eluted at 15 mL. This indicates that sybody 19 did not bind to pUL98 construct 1g. This observation was further confirmed by the SDS-PAGE gel (Figure 4.2.11–B), where the lane corresponding to pUL98 and sybody 19 did not display a band for sybody 19, while lanes for the other peaks showed sybody bands.

The key difference between constructs 1 and 1g is the substitution of the IDR with a GGGS linker in construct 1g. The observation that sybody 19 binds to construct 1 but not to construct 1g suggests that sybody 19's binding to pUL98 is dependent on the presence of the IDR.

4.2.18 Co-crystallization of pUL98 cons.1 and cons.1g with sybodies

pUL98-sybody complexes were formed by incubating pUL98 with sybodies at a 1:2 molar ratio for 12 hours at 4°C. The solutions were then injected into a Superdex 200 Increase 10/300 GL column, as described in Section 3.23, using 98_sec buffer. The eluted pUL98-sybody complexes were analyzed via SDS-PAGE to confirm the presence of both components of the complex. The complexes were then concentrated to the required concentrations (4.5 mg/mL and 9 mg/mL).

Co-crystallization was attempted for pUL98 construct 1 with sybodies 9, 19, 30, and 31, and for pUL98 construct 1g with sybodies 9, 30, and 31. These experiments were carried out using sparse matrix screens from Jena Biosciences in sitting-drop vapor diffusion crystallization plates at the two concentrations mentioned above.

Co-crystallization trials were also conducted for pUL98 constructs 1 and 1g in combination with two sybodies. However, both strategies failed to yield diffraction-quality crystals, with visible aggregation observed in the crystallization drops. Additional trials using different concentrations of the complexes were also attempted but were similarly unsuccessful.

4.2.19 Expression and purification of megabodies

Following the unsuccessful attempts to obtain co-crystals of pUL98 in complex with sybodies, it was hypothesized that a larger crystallization chaperone might enhance the crystallizability of pUL98 constructs 1 and 1g. As a result, megabodies were engineered using the existing sybodies as crystallization chaperones.

Megabodies represent a novel class of nanobody-fused chimeric molecules designed to address challenges in cryo-electron microscopy (cryo-EM), particularly for small and difficult particles, and are also applicable as crystallization chaperones. These molecules are created by grafting single-domain antigen-binding constructs such as nanobodies (~15 kDa) or monobodies (~10 kDa) onto larger scaffold proteins like HopQ (a β-turn between S3–S4 of the adhesin domain of Helicobacter pylori, 45 kDa) and YgjK (a β-turn between A'S1–A'S2 of glucosidase from E. coli, 86 kDa). The resulting megabodies form stable, rigid, and efficiently folding monomeric chimeras [Uchanski et al., 2021]. The structures of megabody scaffolds fused with an anti-GFP nanobody (anti - Green Fluorescent Protein) have been solved using X-ray crystallography (PDB: 5LP2 and 3W7T).

Synthetic gene fragments for both HopQ and YgjK scaffolds fused to a template sybody were ordered from Twist Biosciences. These gene fragments were cloned into the pSBinit plasmid under the araBAD promoter with the pelB signal peptide using RF cloning, as described in Section 3.8. Primer pairs mbs_psb were used to clone HopQ, while mbl_psb were used to clone YgjK. These cloned plasmids served as base plasmids for cloning other sybodies onto them. Sybodies against pUL98 were cloned into the megabody base plasmids using RF cloning, as detailed in Section 3.8, with syb_mbs primers. All cloned plasmids were confirmed via Sanger sequencing (outsourced to LGC Genomics).

Megabodies were expressed in the periplasm of MC1061 cells following the protocol described in Section 3.35. The resulting cell pellet was lysed using sybody_lysis buffer, as outlined in Section 3.18. The lysate was purified by Ni-affinity chromatography using sybody_wash and sybody_elution buffers, with elution performed as a gradient up to 100% starting from 40 mM.

The protein eluted from the HisTrap column during Ni-affinity chromatography was concentrated and injected into SEC using a Superdex 75 26/600 column, as described in Section 3.22, with sybody_sec buffer. Initial purification revealed that the yield of the megabody YgjK protein was approximately 600 µg from a 3-liter culture, while the yield of the megabody HopQ protein was approximately 30 mg from the same 3-liter culture. Based on these results, the HopQ megabody scaffold was selected for the expression of additional megabodies fused with sybodies against pUL98. Figure 4.2.12 shows nickel affinity chromatogram, SEC chromatogram and SDS-PAGE gel for megabody_small (HopQ)_sybody 30 (Mbs_30)

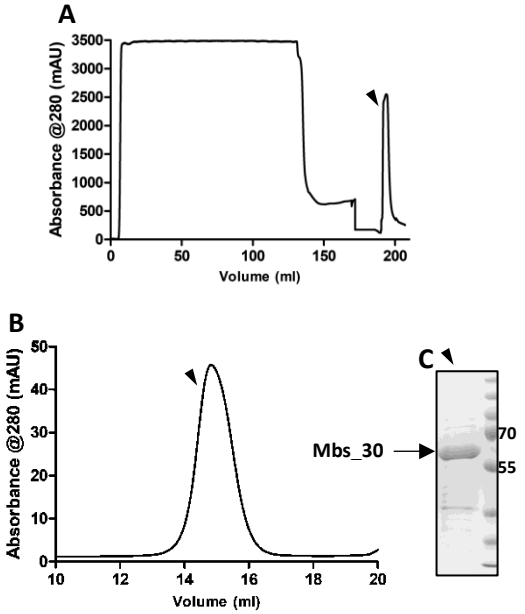


Figure 4.2.12: Nickel affinity chromatogram, SEC chromatogram and SDS-PAGE Gel for megabody_small (HopQ)_Sybody 30 (Mbs_30) that bind to pUL98
Panel (A) shows Nickel affinity chromatogram for Mbs_30 protein. Panel (B) shows SEC chromatogram for Mbs_30 protein. Panel (C) 15% SDS-PAGE Gel loaded with a fraction from SEC.

4.2.20 Megabodies as crystallization chaperone.

A 1:2 molar ratio of pUL98 constructs 1 and 1g to megabodies with sybodies 9, 19, 30, and 31 was incubated in a 10 mL solution for 12 hours at 4°C. The solutions were then injected into a Superdex 200 26/600 SEC column, as described in Section 3.22, using 98_sec buffer. The eluted proteins were concentrated to the required concentrations to set up sitting-drop vapor diffusion crystallization plates. Each drop contained 0.15 µL of protein and 0.15 µL of reservoir solution from sparse matrix reservoir solutions provided by Jena Biosciences, using a Gryphon instrument. Crystallization trials were conducted at two different concentrations for the four megabodies (HopQ megabodies with four distinct sybodies) and the two pUL98 constructs (constructs 1 and 1g). Despite these efforts, no diffraction-quality crystals were obtained. Figure 4.2.13 shows the final SEC chromatogram and SDS-PAGE gel for pUL98 cons.1 and Mbs_30 complex.

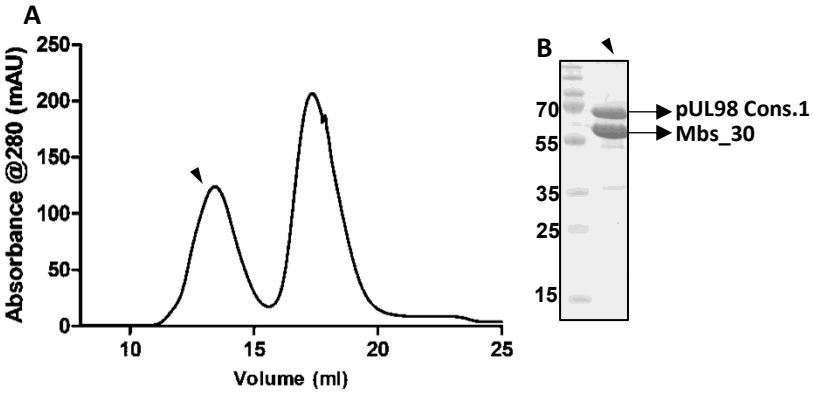


Figure 4.2.13: SEC chromatogram and SDS-PAGE Gel for Megabody (Mbs_30)_pUL98 Cons.1 complex
Panel (A) shows SEC chromatogram for Megabody (Mbs_30)_pUL98 Cons.1 complex. Panel (B) 15% SDS-PAGE Gel loaded with protein from a fraction from SEC.

4.2.21 Megabody pUL98 cons.1 complex for cryo-electron microscopy single particle analysis (cryoEM SPA)

Megabody scaffolds are versatile and can be used in cryoEM single-particle analysis (SPA) to form complexes with target proteins that are too small for cryoEM SPA on their own. In this case, the molar mass of pUL98 construct 1 is 67 kDa, which, although theoretically sufficient for SPA, is practically challenging for structural determination. As a solution, the megabody_sybody30 complex was utilized for cryoEM SPA.

The complex was prepared following the procedure described in Section 4.2.19 (similar to the preparation for crystallization), with adjustments to the protein concentration and the column used. For cryoEM SPA, the sample, at a concentration of ~1 mg/mL, was injected into a Superdex 200 Increase 10/300 GL SEC column, as outlined in Section 3.23, using 98_mbs_sec buffer. The eluted protein was collected in fractions of 100 µL. The fraction corresponding to the midpoint of the peak, with a concentration of 0.17 mg/mL as observed in the chromatogram, was immediately used for vitrification and grid preparation.

Dr. Guido Hansen vitrified the sample onto grids using a Vitrobot Mark IV plunge freezer and collected data using the Talos Arctica cryoEM. (The sample vitrification and data collection process are not described here.)

4.2.22 CryoEM image processing

CryoSPARC was utilized for image processing of the cryoEM SPA dataset. A total of 3,531 movies were collected with an electron exposure of 30 e-/Å^2 and a pixel size of 0.94 Å. The dataset was preprocessed using patch motion correction and patch contrast transfer function (CTF) correction jobs in CryoSPARC. Micrographs were manually curated to exclude those with a CTF fit greater than 5 Å, resulting in 3,100 micrographs for particle picking.

A map of the pUL98 and megabody_sybody30 complex was generated in ChimeraX v1.7 using an AlphaFold2-predicted model of the complex as a template. Template-based particle picking in CryoSPARC, using this predicted model, identified over 3 million particles. These particles were classified in 2D space using a 2D classification job, followed

by the selection of appropriate classes to run ab-initio reconstruction with 2.5 million particles, generating initial volumes.

The volumes generated from ab-initio reconstruction were further refined using five rounds of heterogeneous refinement, resulting in 259,600 particles in one class with a relatively better map. This class was subsequently refined using a homogeneous refinement job, yielding the final map with a Fourier Shell Correlation (FSC) of 0.143 at 5.84 Å resolution (refer to Figure 4.2.14–E).

The structure of pUL98 in complex with megabody_sybody30 (refer to Figures 4.2.14–C and D) was fitted to the refined map using ChimeraX. This fitting (refer to Figures 4.2.14–A and B) demonstrated that the overall shape and volume closely matched the predicted structure. However, the map quality was insufficient to precisely fit secondary structures and lacked complete volume for the megabody. The conical Fourier shell area ratio of 0.37 indicated slight bias in particle orientation, and the orientation chart (refer to Figure 4.2.14–F and G) revealed significant particle misalignments.

Although these results are not sufficient for atomic-resolution reconstruction, they suggest that the structure of pUL98 in complex with megabody_sybody30 could potentially be resolved using cryoEM. Further optimization of sample preparation, grid vitrification, and data collection parameters is planned.

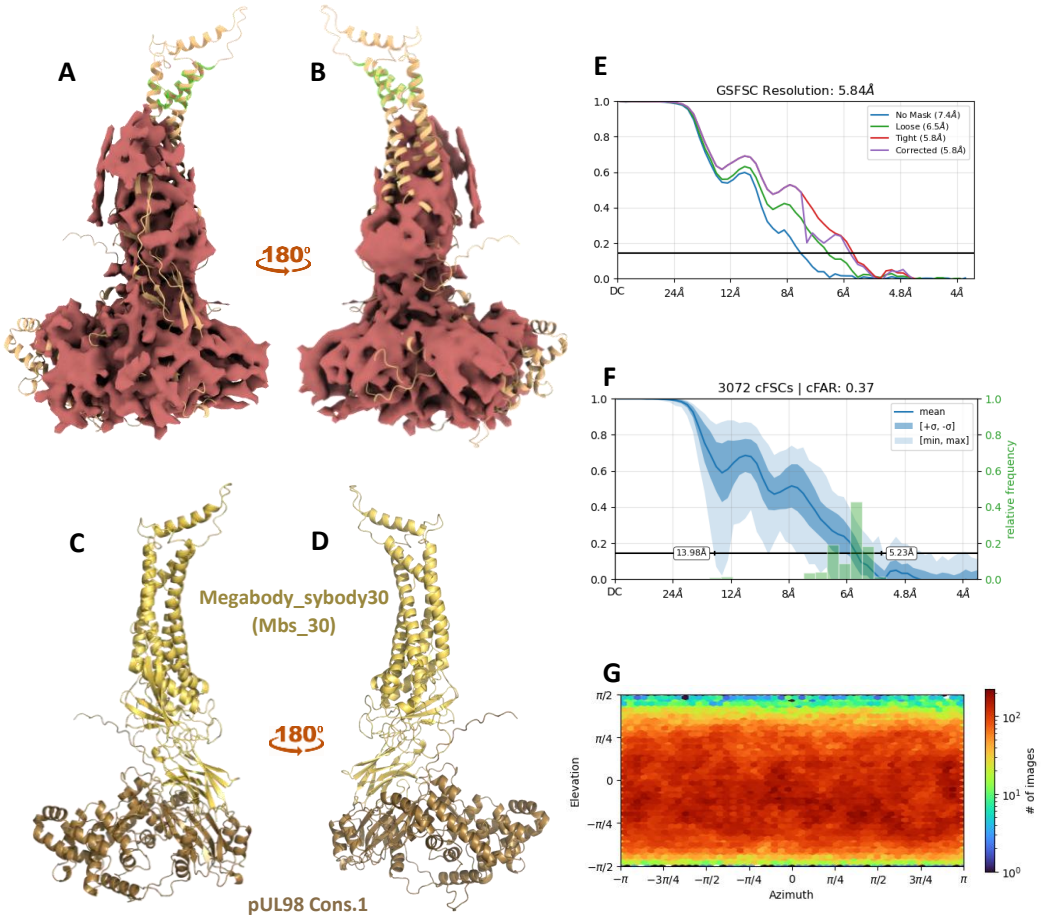


Figure 4.2.14: Analysis of pUL98 in complex with megabody (HopQ)_sybody30 (Mbs_30) In this figure, Panel (A) shows the homogenous reconstruction cryo-EM map of pUL98 in complex with Megabody (HopQ)_sybody30 (Mbs_30) overlaid with the Alphafold2 multimer model of the pUL98_Mbs_30 complex. Panel (B) presents a 180° rotated projection of the cryo-EM map fitted with Alphafold2 multimer model. Panel (C) displays the Alphafold2 multimer model of pUL98 and Mbs_30, with Panel (D) providing a 180° rotated projection. Panel (E) shows the gold-standard Fourier shell correlation (FSC) plot. Panel (F) presents the conical Fourier shell correlation plot with a conical Fourier area ratio of 0.37. Panel (G) shows orientation distribution.

4.3 Structural characterization of pUL94/pUL99

4.3.1 Previous work performed on pUL94/pUL99

Dr. Giacomo Castoro, a colleague and alumnus of our lab, worked on the crystallization of pUL94/pUL99. When the 345 amino acids pUL94 protein was recombinantly expressed in E. coli, it aggregated and formed inclusion bodies. However, when expressed solubly in Schneider's 2 (S2) insect cells, biophysical characterization revealed that the protein still formed aggregates.

As described in Section 1.8.4, the N-terminal 60 amino acids of pUL99 are sufficient for its perceived function. Consequently, a peptide comprising amino acids 22 to 47 of pUL99 was attempted for co-expression with pUL94. Co-expression of this peptide fused to mNeon Green fluorescent protein successfully mitigated aggregation, resulting in a monomeric form of pUL94 in solution.

Dr. Giacomo Castoro extensively attempted to crystallize the pUL94/pUL99 complex, but without success. This prompted collaboration with Dr. Tânia Custodio and Dr. Christian Löw to screen for sybodies binding to the pUL94/pUL99 complex. Eight sybodies were shortlisted and subsequently tested in co-crystallization trials with pUL94/pUL99. Despite these extensive efforts, diffraction-quality crystals were not obtained. However, these attempts resulted in the successful crystallization of sybodies alone, with several sybody structures being resolved.

Predictive analysis using IUPRED2A suggests that amino acids 20 to 190 of pUL99 are intrinsically disordered, while MobiDB predicts that this disordered region spans amino acids 27 to 190 (refer to Figure 4.3.1). Regions with IUPRED2 scores greater than 0.5 indicate intrinsic disorder, implying a lack of fixed or ordered structure in these areas of the protein. This intrinsic disorder explains why co-expression of the wild-type pUL99 protein was never attempted.

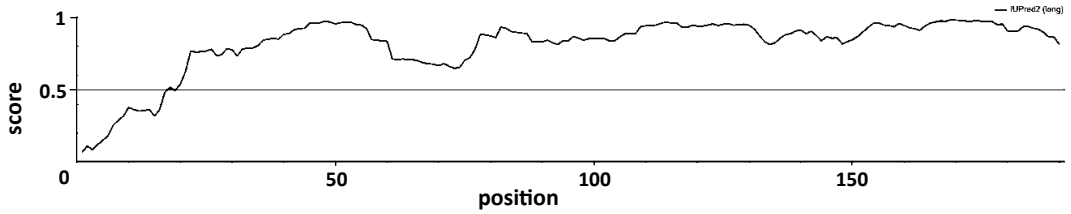


Figure 4.3.1: Bokeh plot of amino acid residues of pUL99 versus IUPred2 prediction score

Table 4.8 pUL94 and pUL99 constructs		
Gene of interest	**Vector**	**Description**
pUL94 (39kDa)	pMT	N-TwinStrep-EK-pUL94(aa1-aa345)-C
pUL99 (30kDa)	pMT	N-10xHis-TEV-pUL99(aa22-aa47)-TEV-mNeonGreen-TwinStrep-C

4.3.2 Expression and purification of sybodies

Sybodies were expressed in Rosetta-Gami2(DE3) E. coli strain, as described in Section 3.12, using TB media. The harvested cell pellets were resuspended in sybody_lysis buffer and lysed using a sonicator, following the procedure outlined in Section 3.18. The lysates were purified through Ni-affinity chromatography, as detailed in Section 3.20, using sybody_wash and sybody_elution buffers. Elution from the HisTrap FF column was performed as a gradient, increasing imidazole concentrations from 30 mM to 500 mM.

The eluted sybodies were subsequently injected into a Superdex 75 26/600 SEC column, as described in Section 3.22. The yield of sybodies varied, ranging from 1 mg/L to 7 mg/L of culture.

4.3.3 Cross-competition assay

When I began my work, I reattempted co-crystallization of the pUL94/pUL99 complex with different sybodies. However, these efforts were not fruitful. To enhance the chances of success, a cross-competition assay was performed to identify and group sybodies binding to non-overlapping antigenic regions on pUL94/pUL99, enabling co-crystallization with multiple sybodies. The cross-competition assay was conducted using bio-layer interferometry (BLI), as outlined in Section 3.28.

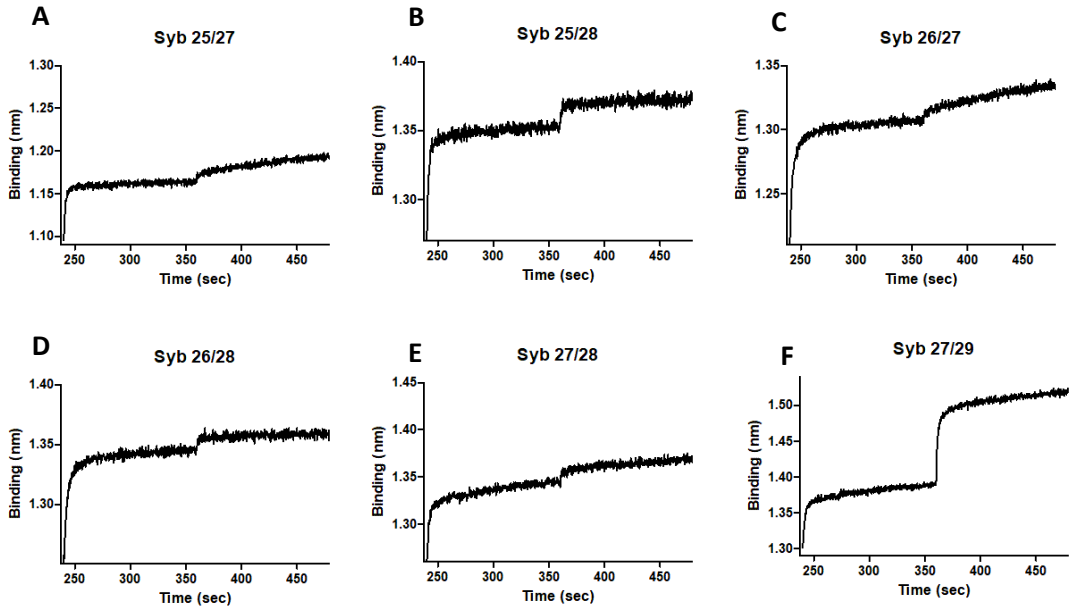


Figure 4.3.2: Sensograms from BLI showing sybody pairs that does not interact together with pUL94/pUL99 complex

Panel (A) shows that sybodies 25 and 27 do not bind simultaneously. Panel (B) reveals that sybodies 25 and 28 do not bind simultaneously, suggesting they recognize overlapping regions. Panel (C) demonstrates that sybodies 26 and 27 do not bind at the same time, Panel (D) indicates that sybodies 26 and 28 do not bind simultaneously. Panel (E) shows that sybodies 27 and 28 do not bind at the same time, Panel (F) shows that sybodies 27 and 29 can bind simultaneously and is a control to analyse other sensograms.

In this assay, sybodies capable of binding simultaneously were considered to recognize distinct, non-overlapping antigenic regions. Based on sybody pairs that did not bind together, sybodies binding to unique non-overlapping antigenic regions were reverse-calculated. Figures 4.3.2 (A–E) display the sensograms for sybody pairs that bind simultaneously, while Figure 4.3.2 (F) represents a control in which sybody pairs bind together. In these sensograms, the association of the first sybody is observed from 240 to 360 seconds, followed by the association of the second sybody from 360 to 480 seconds.

During the assay, one of the eight sybodies aggregated and was subsequently excluded from further testing. The assay identified three distinct antigenic regions to which the sybodies bind, as summarized in Table 4.9.

Table 4.9 Sybodies grouped based on binding to unique non-overlapping antigenic regions of pUL94/pUL99		
Group A	**Group B**	**Group C**
24	25	29
	26	
	27	
	28	
	31	

4.3.4 Expression and purification of pUL94/pUL99 in S2 cells

Expression of pUL94/pUL99 was carried out in S2 cells, as described in Section 3.15. The resulting cell pellet was resuspended in 9499_lysis buffer and lysed by sonication, following the procedure outlined in Section 3.19. The lysate was purified using streptactin purification with streptactin_wash and streptactin_elution buffers. Batch elution was directly performed using 100% elution buffer.

The eluted protein underwent TEV protease digestion, as detailed in Section 3.34, to cleave the mNeonGreen fusion protein. It was then injected into a Superdex 75 26/600 SEC column, as described in Section 3.22, using 9499_sec buffer. The protein eluted from SEC was treated with EK-Max (enterokinase enzyme) in a 1:20 weight-to-volume dilution, using 100 µg of pUL94/pUL99 mixed with 0.16 µL of enterokinase enzyme to remove the Twin Strep tag from the N-terminal end of pUL94. The solution was incubated at 27°C for 12 hours without shaking and subsequently reinjected into the Superdex 75 26/600 SEC column with 9499_sec buffer. Figure 4.3.3 shows streptactin chromatogram, SEC chromatogram and SDS-PAGE gel for pUL94/pUL99 loaded with protein from SEC.

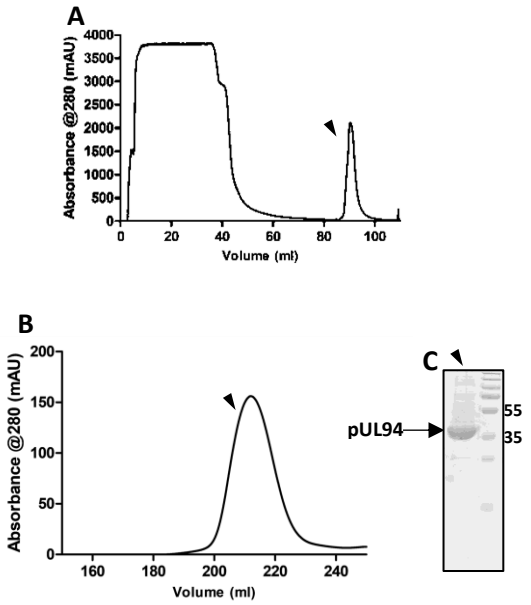


Figure 4.3.3: Streptactin chromatogram, SEC chromatogram and SDS-PAGE Gel for pUL94/pUL99 purification

Panel (A) shows Streptactin chromatogram for pUL94/pUL99. Panel (B) shows SEC chromatogram for affinity tag cleaved pUL94/pUL99 protein. Panel (C) 15% SDS-PAGE Gel loaded with protein from a fraction from SEC.

4.3.5 Expression enrichment of pUL94/pUL99 in S2 cells using fluorescence-activated cell sorting (FACS)

One of the bottlenecks in this work is the relatively low yield of protein. My colleague, Giacomo, utilized the conventional subculturing method to identify high-yielding clones for pUL94. However, after the transfection of an S2 stable cell line harboring the gene for pUL94, identifying high-yielding clones became imperative.

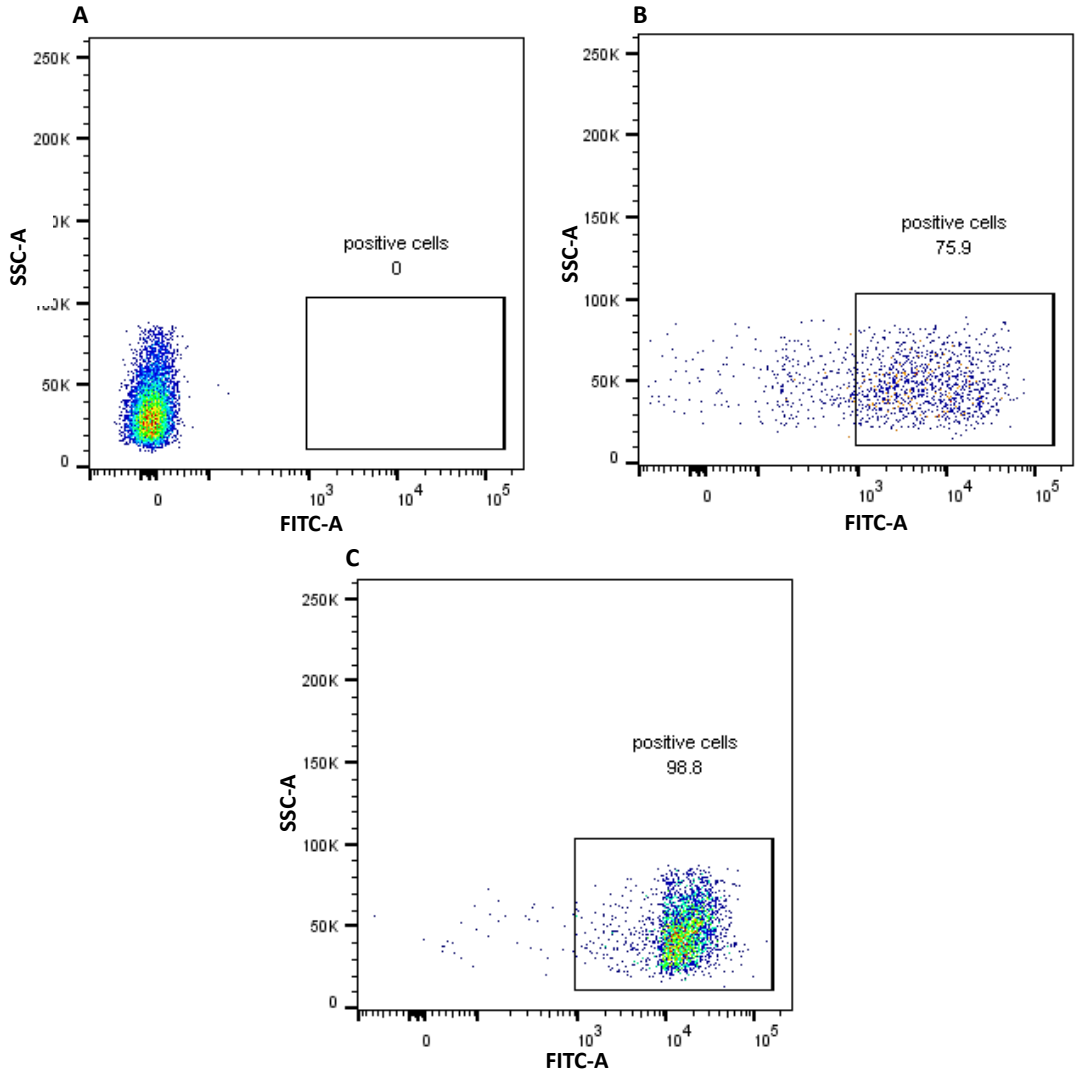


Figure 4.3.4: FACS sorting of S2 cells expressing pUL99-mNeonGreen with pUL94
Panel (A) depicts un-induced cells serving as the negative control, showing no significant fluorescence in forward scattering. Panel (B) illustrates the sorting process of cells with a mean fluorescence intensity (MFU) higher than 10^3. (C) shows the final quality check of sorted cells, confirming robust fluorescence in the enriched population.

To address this, the pUL99 peptide was fused to mNeonGreen at the carboxy terminus. Instead of employing the conventional subculturing method, fluorescence-activated cell sorting (FACS) was used to sort high-expression clones based on the fluorescence emitted by mNeonGreen. After three days of expression, the cells were

subjected to FACS for enrichment. A blue laser with an excitation wavelength of 498 nm and an emission filter bandpass at 517 nm (Fluorescein isothiocyanate – FITC – setting) was employed to sort cells with a mean fluorescence intensity (MFI) exceeding 103 (refer to Figure 4.3.4).

Figure 4.3.4-A depicts the uninduced negative control, while Figure 4.3.4-B highlights the positive cells inside the rectangular box that were sorted, excluding the remaining cells. Figure 4.3.4-C displays the quality check plot of the sorted cells, showing the distribution of fluorescence intensities measured with the FITC-A forward scattering detector and SSC-A (side scatter area) detector. The sorted cells were subsequently revived, cultured, and used for further expression. FACS sorting resulted in a two-fold increase in the yield of pUL94/pUL99.

4.3.6 Co-crystallization of pUL94/pUL99 with multiple sybody pairs

The pUL94/pUL99 protein was incubated with all combinations of sybody pairs at a 1:2 molar ratio of pUL94/pUL99 to each sybody at 4°C for 12 hours. After incubation, the mixtures were injected into a Superdex 200 26/600 SEC column, as detailed in Section 3.22. The pUL94/pUL99-sybody complexes eluted from SEC were concentrated to final concentrations of 4 mg/mL, 7 mg/mL, and 12 mg/mL, and used to set up sitting-drop vapor diffusion crystallization plates. Despite these efforts, no diffraction-quality crystals were obtained.

4.3.7 Cloning of pUL94 gene constructs for expression in *E.coli*

Following the unsuccessful attempts at crystallizing the pUL94/pUL99 complex thus far, two truncated gene constructs of pUL94 (constructs 8 and 6) were cloned into the pET vector for expression in E. coli. Construct 8 was cloned using the primer 94_382 through RF cloning, while construct 6 was cloned using the primer 94_137 through insertion mutagenesis cloning, as described in Sections 3.8 and 3.9, respectively. Details of the constructs are provided in Table 4.10.

4.3.8 Expression and purification of pUL94 cons.6 and cons.8

Constructs 6 and 8 of pUL94 were used to transform the BL21DE3 E. coli strain, as described in Section 3.6, and expressed in 3-liter cultures following the protocol outlined in Section 3.12.

Table 4.10 pUL94 truncated constructs

Constructs	Vector	Description
Cons.6 (22.7kDa)	pET	N-pUL94(aa146-aa328)-TEV-6xHis-C
Cons.8 (24.8kDa)	pET	N-pUL94(aa137-aa328)-TEV-TwinStrep-C

The harvested cell pellet from construct 6 was lysed using sybody_lysis buffer by sonication, as described in Section 3.18. The lysate was purified via Ni-affinity chromatography (outlined in Section 3.20) using sybody_wash and sybody_elution buffers. The eluted protein was then injected into a Superdex 75 26/600 SEC column, following the method detailed in Section 3.22, using 94_sec buffer. The protein eluted from SEC exhibited impurities, as shown in the SDS-PAGE gel in Figure 4.3.5–A. Attempts to remove these impurities were unsuccessful. Furthermore, the yield of construct 6 was 300 µg from the 3-liter culture. Despite this, one set of sitting-drop vapor diffusion

crystallization screens was set up with construct 6 protein. However, this attempt did not result in the formation of crystals.

The harvested cell pellet from construct 8 was lysed using 9499_lysis buffer by sonication, as described in Section 3.18. The lysate was purified through streptactin purification (as described in Section 3.21) using streptactin_wash and streptactin_elution buffers. The protein eluted from the streptactin column also displayed impurities, as evident from the SDS-PAGE gel in Figure 4.3.5–B. Due to the high level of impurities, with pUL94 construct 8 making up only a small proportion of the total protein, this construct was not pursued further.

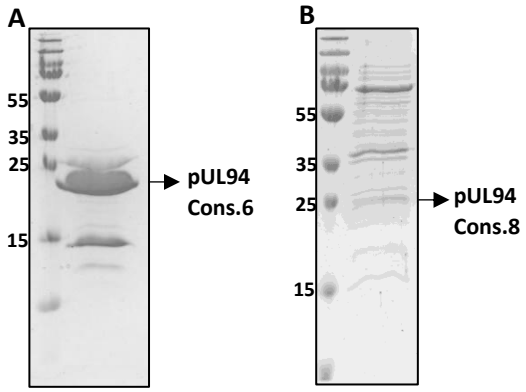


Figure 4.3.5: SDS-PAGE Gels with pUL94 cons.6 (A) and cons.8 (B) protein

4.3.9 Legobody scaffold to study pUL94/pUL99 through cryoEM SPA

Since pUL94/pUL99 proved intractable to crystallization in my hands, we decided to employ cryoEM single-particle analysis (SPA) for structural characterization of the complex. However, the size of pUL94/pUL99 (~41 kDa) posed a significant challenge for cryoEM at that time. To address this, it became imperative to use an interaction partner to increase the size of the complex. Since no interaction partners were known, we explored alternative scaffolds that could be engineered to bind pUL94/pUL99.

In 2021, Wu and Rapoport introduced nanobody-binding scaffolds termed Legobodies, designed to facilitate the structural characterization of small proteins (e.g., SARS-CoV-2 RBD and KDEL receptor) smaller than ~25 kDa via SPA. Legobodies are complex structures comprising maltose-binding protein (MBP) fused to domains C and D of protein A and protein G. These scaffolds incorporate Fab fragments that bind nanobodies, which, in turn, bind the target protein. The architecture of the Legobody complex—characterized by its rigid structure and distinct shape—not only increases the overall size of the target protein complex by approximately 120 kDa but also enhances particle alignment during image processing, making it suitable for cryoEM analysis.

For this work, plasmids were procured from Addgene: Fab_8D3_2 light chain (ID: 176076), Fab_8D3_2 heavy chain (ID: 176075), and MBP_PrAG (ID: 176077).

4.3.10 Cloning, expression and purification of Fab_8D3_2 and MBP_PrAG

Genes for the Fab_8D3_2 heavy and light chains were cloned into the pMT vector for secretory expression in S2 cells using the RF cloning method [Bond et al., 2022], as described in Section 3.8. The primers fabHC and fabLC, listed in Table 3.3, were utilized, with a Twin Strep tag added to the N-terminus of the heavy chain.

A stable S2 cell line containing plasmids for the secretory expression of Fab_8D3_2 was generated, following the protocol detailed in Section 3.14. Expression of Fab_8D3_2 was conducted as outlined in Section 3.16. The culture supernatant containing Fab_8D3_2 was concentrated and prepared for purification, as described in Section 3.17. The concentrated supernatant was purified using streptactin purification (described in Section 3.21) with streptactin_wash and streptactin_elution buffers. The eluted protein was then concentrated to a volume of 10 mL and injected into a Superdex 75 26/600 SEC column, as described in Section 3.22, using 9499_sec buffer.

The plasmid for MBP_PrAG was directly used to transform the BL21(DE3) E. coli strain, as detailed in Section 3.6. Expression and purification of MBP_PrAG were performed by my colleague, Dr. Kumar Nagarathinam, following the protocols outlined in Sections 3.12, 3.20, and 3.22, using sybody_lysis, sybody_wash, sybody_elution, and 9499_sec buffers.

4.3.11 Assembly of a legobody

Proteins pUL94/pUL99, sybody, MBP_PrAG, and Fab_8D3_2, essential for preparing the legobody_pUL94/pUL99 complex, were expressed and purified individually (refer to Sections 4.3.3, 4.3.5, and 4.3.11). Optimization of the complex assembly process began with sybody 24, focusing on component concentrations, the order of addition, and incubation times to resolve challenges like asymmetric peaks during preparation and aggregation of the legobody complex at higher sybody concentrations (>2 mg/mL). Following optimization, legobody_pUL94/pUL99 complexes were successfully assembled with sybodies 28, 29, and 31. The finalized protocol involved incubating 2 mg of Fab_8D3_2 with 3 mg of MBP_PrAG in 10 mM D-maltose for 6 hours at 4°C on a rotating wheel, followed by the addition of 1 mg of sybody and further incubation for 12 hours. The solution was then loaded onto a 1 mL streptactin resin in a gravity-flow column equilibrated with strep_lego_wash buffer, which captured the legobody complex via the Twin Strep tag on Fab_8D3_2. After washing the resin with 10 mL of strep_lego_wash buffer, the legobody was eluted with 5 mL of strep_lego_elution buffer, concentrated, and quantified using a UV nano-spectrophotometer. To differentiate legobodies containing specific sybodies, the naming convention legobody_sybXX.

4.3.12 Preparation of pUL94/pUL99 in complex with legobody

500 µg of legobody_syb24 added with 330 µg of pUL94/pUL99 to attain 1:2 molar ratio in the total volume of 700 µL and incubated for 12 hours at 4^{o}C in rotating wheel. This solution then injected into superdex 200 10/300 GL SEC column with 9499_sec buffer. Figure 4.3.6 shows SEC chromatogram and SDS-PAGE gel with legobody_pUL94/pUL99 protein loaded from SEC.

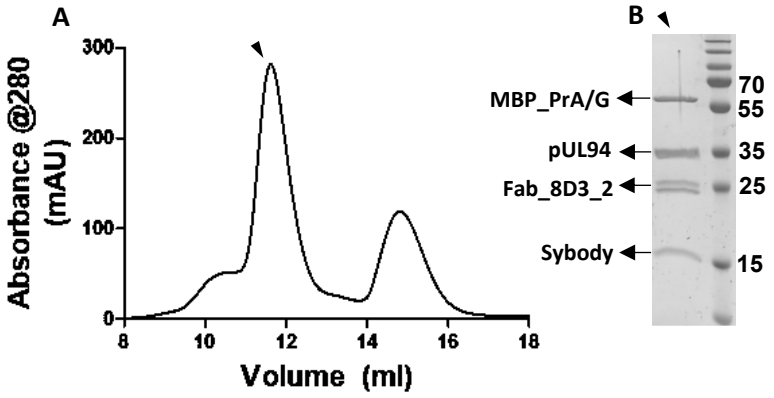


Figure 4.3.6: SEC chromatogram and SDS-PAGE Gel for legobody_pUL94/pUL99 complex

Panel (A) shows SEC chromatogram for legobody_pUL94/pUL99. Panel (B) 15% SDS-PAGE Gel loaded with protein from a fraction from SEC.

4.3.13 Screening of legobody_pUL94/pUL99 complexes with different sybodies

As outlined in Section 4.3.13, legobody assemblies were prepared with sybodies 24, 28, 29, and 31. Legobody_pUL94/pUL99 complexes incorporating all four sybodies were vitrified and screened to identify the assemblies with the least aggregation and dissociation. Based on these screening results, as well as the grand average hydropathy index of the sybodies, the legobody_syb31_pUL94/pUL99 complex was selected for subsequent experiments. Dr. Guendalina Marini from Prof. Maya Topf's lab conducted the vitrification, screening, and cryo-EM data collection.

To summarize the vitrification and data collection process: sample vitrification for cryo-EM involves applying the sample to a specialized grid, which consists of a carbon film with circular holes supported by a metal mesh. Excess liquid is blotted to leave a thin sample film, and the grid is rapidly frozen in liquid ethane, forming vitreous ice that preserves the native structure of the sample. During data collection, the vitrified grids are loaded into a cryo-electron microscope, where regions of interest are identified, and high-resolution movies are captured. The individual frames of these movies are computationally corrected for motion and merged into single micrographs for further analysis.

4.3.14 PEGylation to reduce aggregation of pUL94/pUL99 on grids

Despite improvements in aggregation and dissociation on cryoEM grids with the legobody_syb31_pUL94/pUL99 complex, the sample remained suboptimal. To address this issue, and based on methods for enhancing sample quality in cryoEM SPA (Wu et al., 2018; Zhang et al., 2021), pegylation was applied to the legobody_syb24_pUL94/pUL99 complex. Pegylation involves attaching polyethylene glycol (PEG) polymers of varying lengths to the primary amine group of lysine side chains on the protein surface. This modification is known to reduce sample aggregation during vitrification. Following the protocol outlined by Wu et al., 2018, legobody_syb24_pUL94/pUL99 was pegylated with a 40-fold molar excess of Methyl-PEG12-NHS and incubated for 12 minutes on ice. The reaction was stopped by adding 40 mM final concentration of tris buffer at pH 8.0. The

pegylated sample was then injected into a Superdex 200 10/300 GL SEC column, as described in Section 3.23, using lego1_9499 SEC buffer, and the eluted protein was collected in 200 μL fractions.

4.3.15 Optimization of sample buffer for cryoEM via dynamic light scattering (DLS) experiment

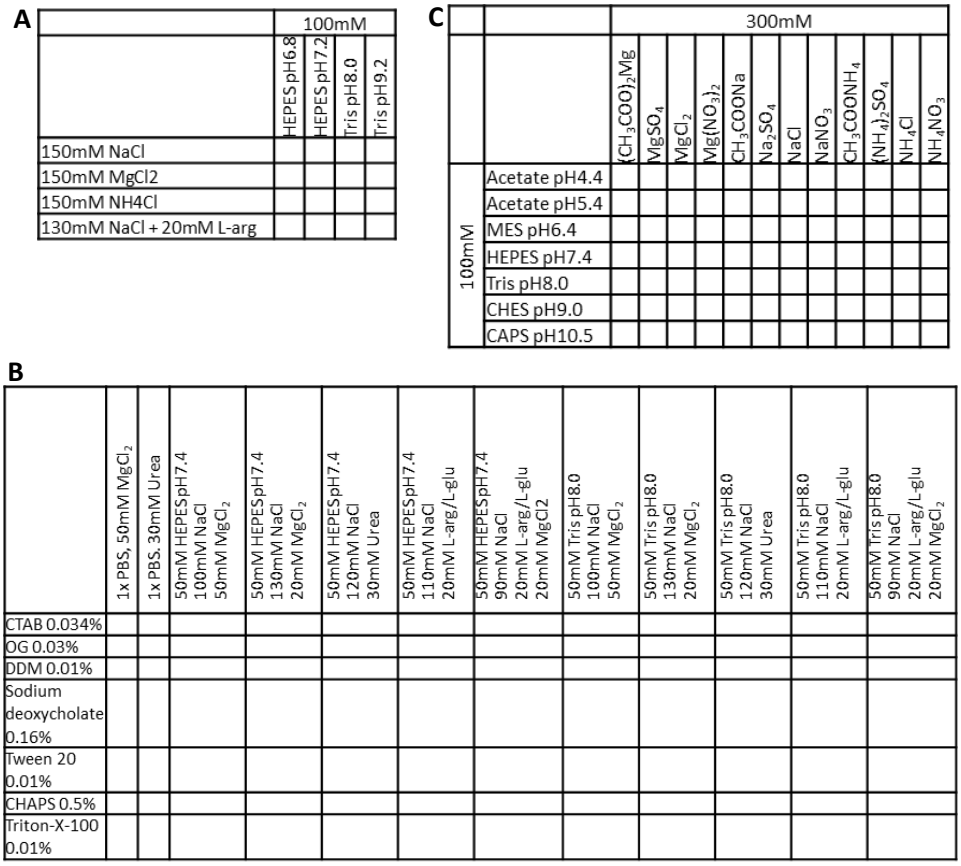
Although PEGylation significantly improved sample aggregation on cryoEM grids, further optimization was required to minimize aggregation and prevent dissociation of the legobody_syb31_pUL94/pUL99 complex. To achieve this, dynamic light scattering (DLS) was employed to screen for buffer and salt conditions that would stabilize the complex for at least four days. At a concentration of 0.8 mg/mL, the legobody_syb31_pUL94/pUL99 sample began aggregating after two days, as illustrated in the radius distribution plot in Figure 4.3.8–A.

A radius distribution plot generated by DLS provides valuable insights into the size and aggregation state of a protein. The x-axis represents the particle size as hydrodynamic radius in nanometers (nm), while the y-axis reflects time for measurements taken across multiple days. A narrow, sharp distribution indicates a monodisperse sample, while broader distributions signal the presence of multiple protein populations. By combining data from different time points, the plot facilitates the assessment of sample quality over time, with the color representing the proportion of the sample contributing to the distribution.

High-throughput DLS experiments in a 96-well format (as described in Section 3.29) were conducted over three rounds with systematically formulated buffer and salt screens for the legobody_syb31_pUL94/pUL99 complex. The first screen included four salts in combination with four buffers (refer to Figure 4.3.7–A). On the first day, the protein complex remained stable in all salts except ammonium chloride (result not shown). By the second day, samples in buffers containing L-arginine began aggregating (Figure 4.3.8–C), while samples with NaCl aggregated on the third day (Figure 4.3.8–A). Samples in buffers containing $MgCl_2$ showed stability over three days (Figure 4.3.8–B), but complex dissociation was also observed in this condition.

The second screen explored different salt and buffer combinations along with seven detergents at concentrations below their critical micellar concentration (CMC) (Figure 4.3.7–B). Unfortunately, the legobody_syb24_pUL94/pUL99 complex was unstable under all tested conditions, with samples either aggregating or dissociating (results not shown).

A third screen was designed based on the Hofmeister series, varying pH (4.4 to 10.5) and buffer compositions (Figure 4.3.7–C). This screen identified six conditions where legobody_syb24_pUL94/pUL99 complexes remained relatively stable for five days (Figures 4.3.8–D to I). Notably, five of these conditions were at pH ≤ 5.4, while the sixth condition comprised CHES buffer at pH 9.0, with 75 mM NaCl and 135 mM $MgCl_2$. The final buffer composition for this condition was proportionally adjusted to reduce salt concentrations, yielding the following optimized condition: 10 mM CHES at pH 9.0, 50 mM NaCl, 100 mM $MgCl_2$, and 2 mM D-maltose (labeled as *lego2_9499*). D-maltose was included in all buffers during preparation to stabilize MBP_PrAG, as it serves as a substrate for MBP. Throughout all DLS experiments, the sample concentration was consistently maintained at 0.8 mg/mL.



A

	100mM			
	HEPES pH6.8	HEPES pH7.2	Tris pH8.0	Tris pH9.2
150mM NaCl				
150mM MgCl2				
150mM NH4Cl				
130mM NaCl + 20mM L-arg				

C

		300mM											
		$(CH_3COO)_2Mg$	$MgSO_4$	$MgCl_2$	$Mg(NO_3)_2$	CH_3COONa	Na_2SO_4	NaCl	$NaNO_3$	CH_3COONH_4	$(NH_4)_2SO_4$	NH_4Cl	NH_4NO_3
100mM	Acetate pH4.4												
	Acetate pH5.4												
	MES pH6.4												
	HEPES pH7.4												
	Tris pH8.0												
	CHES pH9.0												
	CAPS pH10.5												

B

	1x PBS, 50mM $MgCl_2$	1x PBS. 30mM Urea	50mM HEPES pH7.4 100mM NaCl 50mM $MgCl_2$	50mM HEPES pH7.4 130mM NaCl 20mM $MgCl_2$	50mM HEPES pH7.4 120mM NaCl 30mM Urea	50mM HEPES pH7.4 110mM NaCl 20mM L-arg/L-glu	50mM HEPES pH7.4 90mM NaCl 20mM L-arg/L-glu 20mM MgCl2	50mM Tris pH8.0 100mM NaCl 50mM $MgCl_2$	50mM Tris pH8.0 130mM NaCl 20mM $MgCl_2$	50mM Tris pH8.0 120mM NaCl 30mM Urea	50mM Tris pH8.0 110mM NaCl 20mM L-arg/L-glu	50mM Tris pH8.0 90mM NaCl 20mM L-arg/L-glu 20mM $MgCl_2$
CTAB 0.034%												
OG 0.03%												
DDM 0.01%												
Sodium deoxycholate 0.16%												
Tween 20 0.01%												
CHAPS 0.5%												
Triton-X-100 0.01%												

Figure 4.3.7: Buffer and salt screens for optimizing legobody_pUL94/pUL99 using DLS
Panels (A), (B), and (C) display the buffer and salt screens conducted to identify optimal conditions for the legobody_syb31_pUL94/pUL99 complex using Dynamic Light Scattering (DLS). Panel (A) shows various pH, buffer and salt composition screened initially. Panel (B) shows different salts and detergents screened (A). Panel (C) shows the salts, based on Hoffmeister series, and buffers at various pH screened.

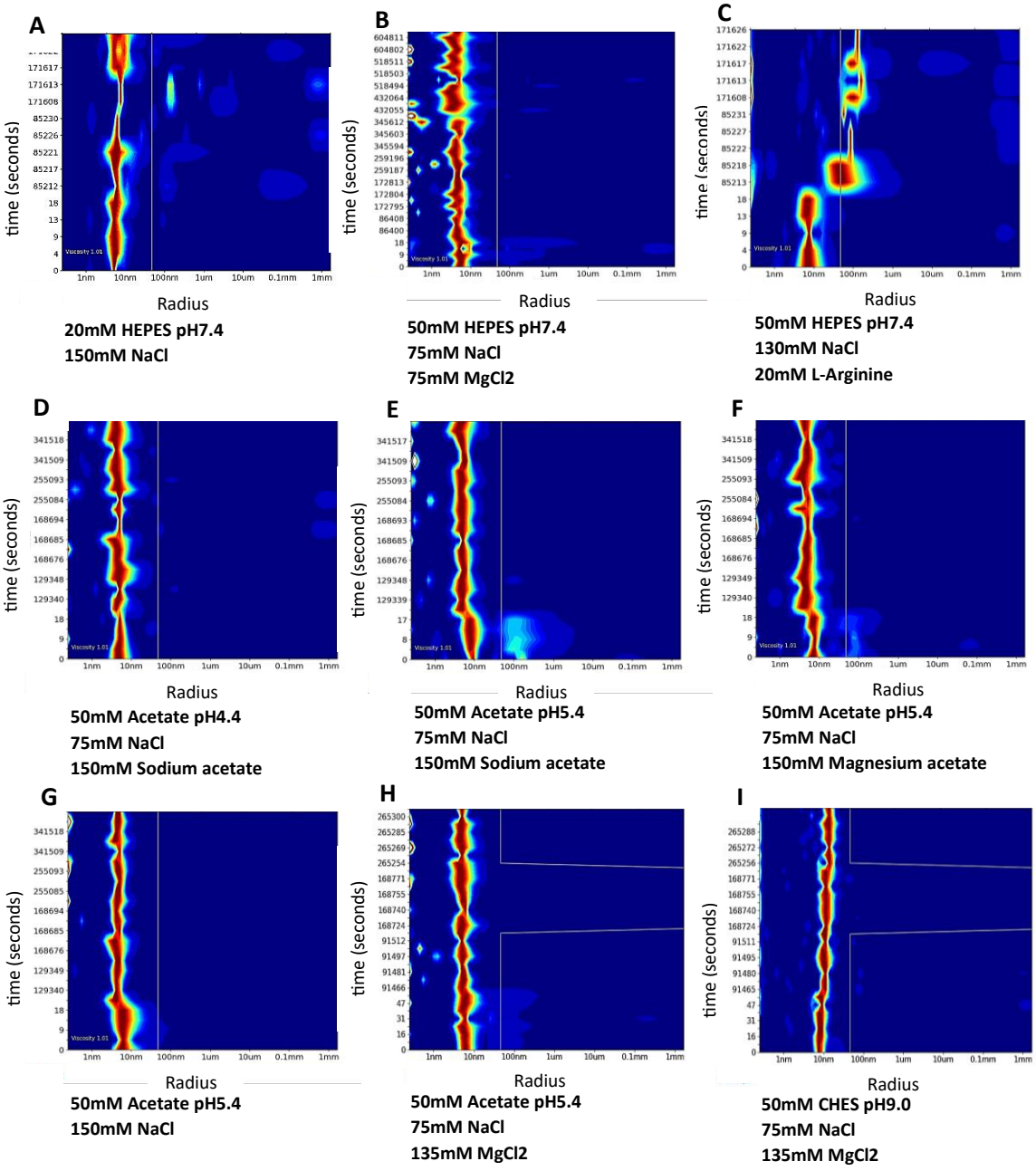


Figure 4.3.8: DLS radial distribution plots for legobody_pUL94/pUL99 complex under various conditions

In all the panels, x-axis represent hydrodynamic radius and y-axis represent time in seconds. Panels (A) to (I) display the Dynamic Light Scattering (DLS) radial distribution plots for legobody_syb31_pUL94/pUL99 complex under different conditions of buffer, pH and salt.

4.3.16 Optimization of vitrification

After optimizing the legobody_syb31_pUL94/pUL99 complex in solution, Dr. Guandalina Marini further refined the freezing parameters to improve grid preparation. This optimization process included evaluating sample concentration, the effect of detergent presence or absence during vitrification, chemical coating with graphene oxide, grid types (such as Quantifoil or C-Flats, varying in copper/gold support, mesh, and hole sizes), glow discharge settings (time and power), vitrification devices (Vitrobot Mark IV and Leica GP2), sample volumes, and blotting times. These efforts revealed that the sample performed optimally at a concentration of 0.05 mg/mL, marking a significant advancement in the preparation of grids for cryoEM analysis.

4.3.17 Image processing of cryoEM dataset

As previously noted, cryoEM data collection was conducted by Dr. Guendalina Marini. During the optimization process, a total of 12 datasets were collected and processed. The typical workflow for cryoEM data processing involves motion correction of particles in individual frames of the recorded movies, CTF (contrast transfer function) estimation and correction, particle picking, 2D classification, reconstruction, and refinement. All data processing was performed using cryoSPARC.

Among the datasets collected, the best one so far comprised 13,153 movies captured at a magnification of 105,000x using a Titan Krios G3 equipped with a 300 keV electron source. This dataset had a total electron exposure of 50.83 $e^{-}/Å^{2}$ and a pixel size of 0.832 Å, making it a significant step forward in the structural analysis of the legobody_pUL94/pUL99 complex.

4.3.17.1 Pre-processing to particle picking

Data must be pre-processed and corrected for motion blur and errors in contrast transfer function (CTF). These are achieved using patch motion correction job and patch CTF estimation job in cryoSPARC suite. The CTF corrected micrographs are curated using manual curation job in cryoSPARC to remove micrographs with CTF fit less than 5.0Å resolution, resulting in 5,756 good micrographs.

It was challenging to pick particles using blob picking tool of cryoSPARC. To pick particles, a map was generated in ChimeraX from the Alphafold2 multimer model of legobody_syb31_pUL94/pUL99. Using this map as template, initial particles are picked using template picking job. 2D classification, Ab-initio reconstruction follows template picking. These three jobs were performed iteratively three times until a good set of particles were identified which were then used to train RESNET8 topaz neural network model. The trained model was then used to pick particles which resulted in 5,921,799 particles. This workflow from pre-processing to particle picking was illustrated in figure 4.3.9.

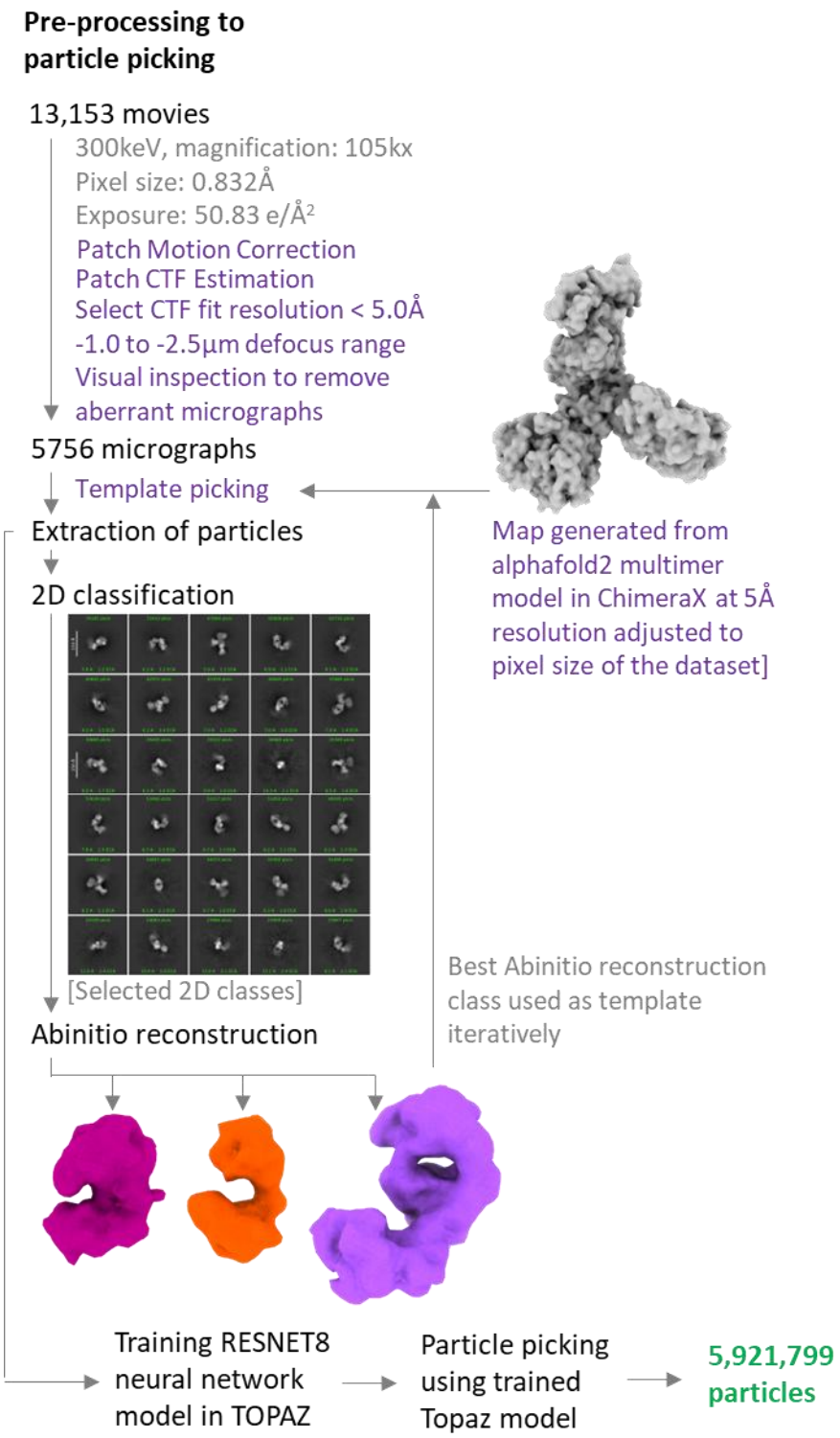


Figure 4.3.9: Particle picking workflow

4.3.17.2 Sorting particles using heterogenous refinement

Approximately 5.9 million particles, picked using the trained Topaz model, were utilized to reconstruct three volumes ab-initio using the ab-initio reconstruction job. These reconstructed volumes were then sorted by inputting three maps into the heterogenous refinement job, with the forced hard classification feature enabled. From the heterogenous refinement results, the best class among the three was selected, along with its associated particles, for further processing. Refer to Figure 4.3.10 for an illustrated overview of this workflow.

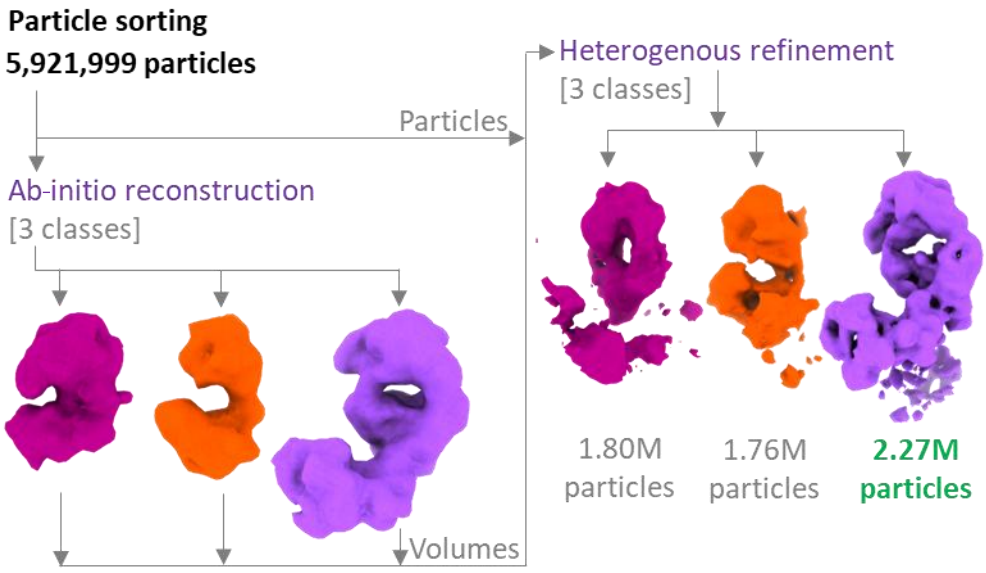


Figure 4.3.10: Particle sorting through heterogenous refinement

4.3.17.3 Map refinement

A total of 2.26 million particles from heterogenous refinement were further refined using homogenous refinement, resulting in a map with a Fourier shell correlation (FSC) 0.143 resolution of 4.32 Å. This consensus map was used for 3D classification, both with and without a mask. The mask was generated to cover the volume of pUL94/pUL99. The workflow without the mask produced a map with an FSC 0.143 of 4.06 Å (refer to Figures 4.3.11 and 4.3.12), while the workflow with the mask resulted in a map with an FSC 0.143 of 5.42 Å.

The best map (out of 20 classes) from the unmasked workflow underwent further homogenous refinement with 276,715 particles. This refined map exhibited relatively well-resolved features for Fab_8D3_2, MBP_PrAG, and the sybody, allowing a structure derived from the cryoEM map of the legobody without the target protein (PDB: 7RXC) to fit appropriately (see Figures 4.3.12–C and D). However, the volume corresponding to pUL94/pUL99 was only partially visible (highlighted in yellow in Figures 4.3.12–A and B) and insufficient for modeling the structure. The conical Fourier area ratio (cFAR) of 0.39 (see Figure 4.3.12–F) indicated some orientation bias in the particle set. A cFAR value of 0.5 is generally considered the borderline threshold for atomic resolution reconstruction.

Map refinement

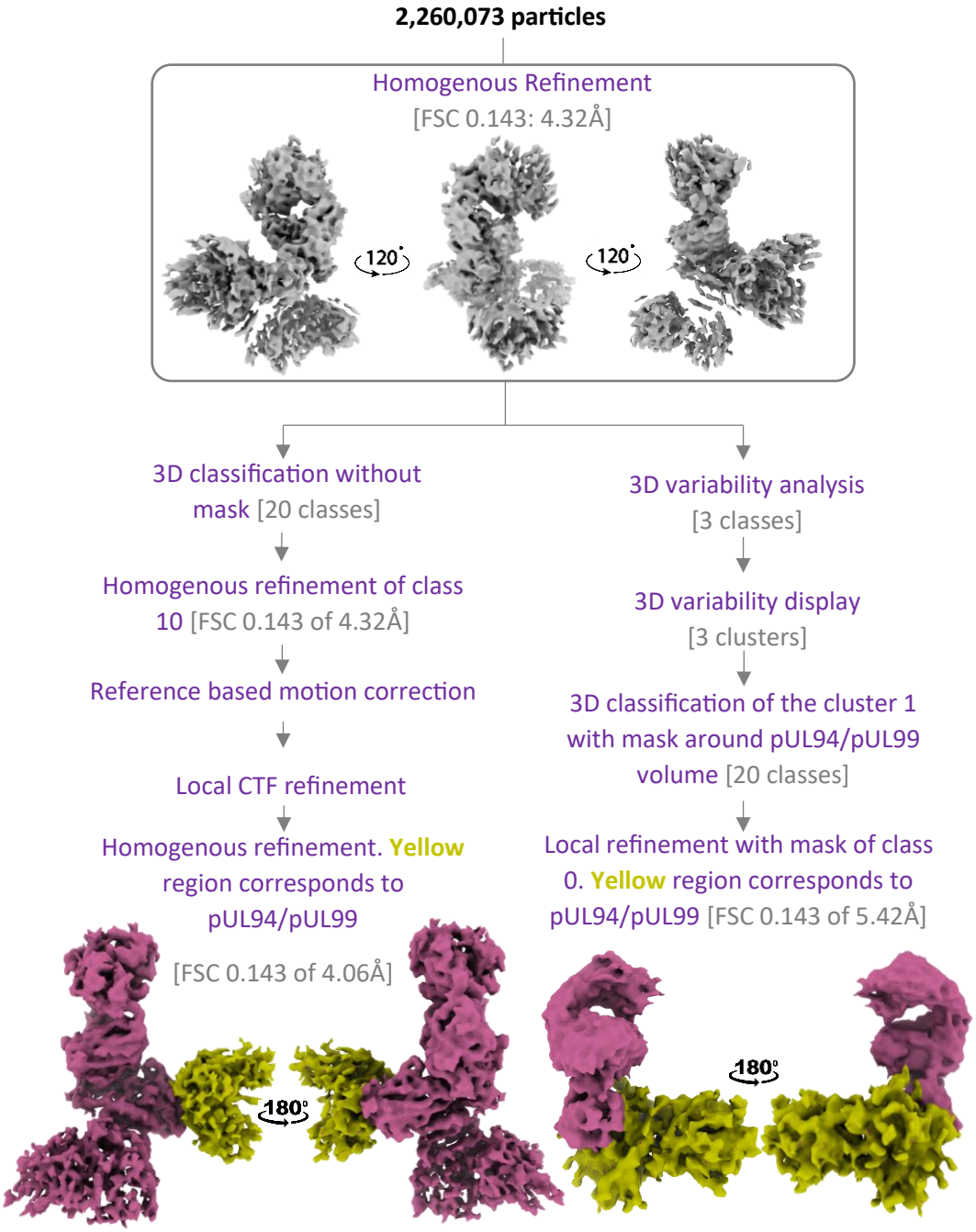


Figure 4.3.11: Map refinement

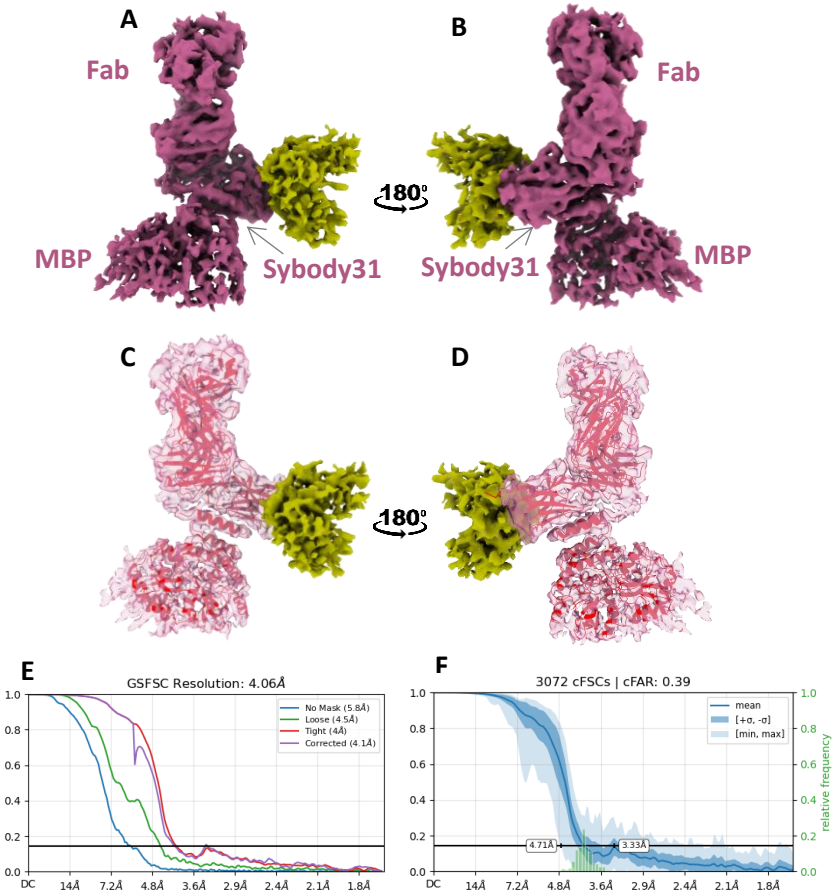


Figure 4.3.12: Cryo-EM maps and analysis of legobody_syb31_pUL94/pUL99 complex
Panels (A) and (B) display homogenous reconstruction cryo-EM maps of the Legobody_syb31_pUL94/pUL99 complex with 180° horizontal rotation, highlighting the pUL94/pUL99 complex in yellow. Panels (C) and (D) show the same cryo-EM maps overlaid with the AlphaFold2 multimer model at 180° rotations, Panel (E) presents the FSC plot estimating an FSC 0.143 resolution of ~4.1Å. Panel (F) displays the conical Fourier shell correlation plot with a ratio of 0.39

The best map from the masked workflow was refined using a local refinement job with 29,880 particles (refer to Figure 4.3.11 for the workflow). The refined map showed an increase in the total volume of pUL94/pUL99; however, the volumes for Fab_8D3_2 appeared more diffuse, and MBP_PrAG was not visible (see Figures 4.3.13–A and B). The FSC 0.143 of this map was 5.42 Å (see Figure 4.3.13–E), which is insufficient for structural modeling. The cFAR ratio of 0.41 (see Figure 4.3.13–F) also suggested orientation bias in the particle set.

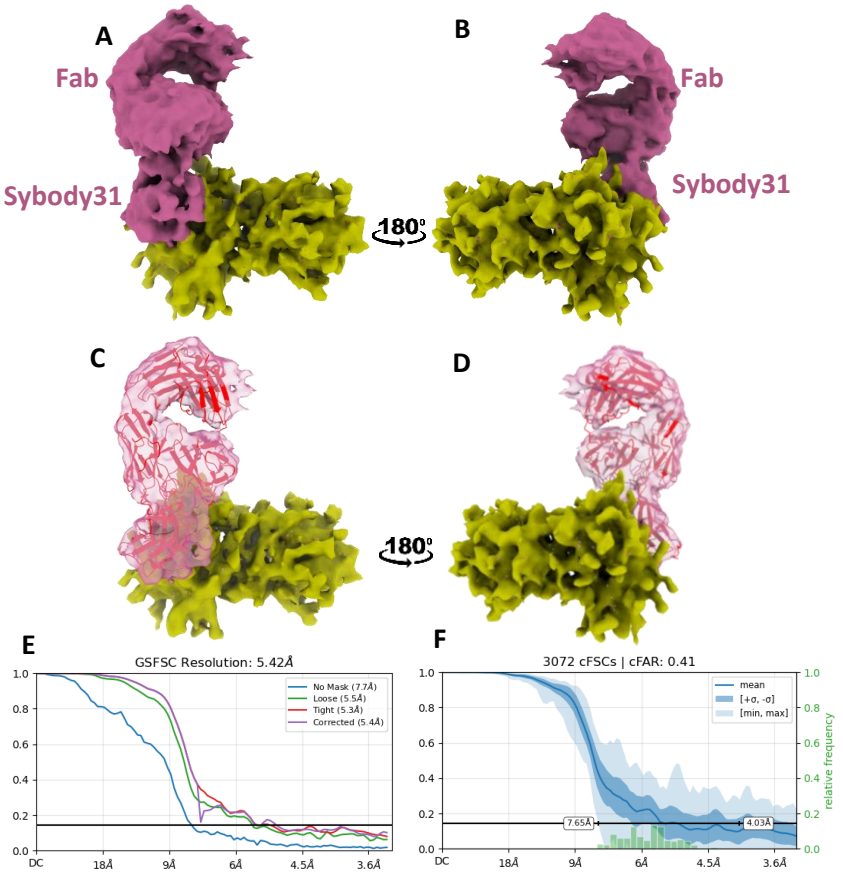


Figure 4.3.13: Cryo-EM analysis of legobody_syb31_pUL94/pUL99 complex with mask
Panels (A) and (B) display homogenous reconstruction cryo-EM maps of the Legobody(s31)_pUL94/pUL99 complex focused on the pUL94/pUL99 region using a mask, with 180° horizontal rotations, and the partial volume colored in yellow. Panels (C) and (D) show these maps overlaid with the AlphaFold2 multimer model of Legobody_syb31 at 180° rotations, again highlighting the pUL94/pUL99 volume in yellow. Panel (E) provides the gold standard Fourier shell correlation plot. Panel (F) illustrates the conical Fourier shell correlation plot with a conical Fourier area ratio of 0.41.

Analysis of volume maps from 3D classification jobs, as well as 3D variability analysis, revealed significant heterogeneity in the binding angles of the legobody components (Fab_8D3_2, MBP_PrAG, and pUL94/pUL99) and within each individual component. This heterogeneity underscores the need for a larger number of high-quality particles. Plans are in place to collect additional data, as a larger particle set is expected to mitigate orientation bias and improve map resolution.

5. Discussion

Human Cytomegalovirus (HCMV) infection is widespread and can cause severe disease, particularly in organ transplant recipients and newborns. Most of the currently available antivirals are associated with significant toxicity, and the emergence of drug-resistant HCMV strains is an increasing concern. These challenges highlight the need for developing new antivirals targeting additional HCMV proteins.

Not all HCMV antivirals are approved for all three therapeutic regimens—prophylaxis, pre-emptive treatment, and treatement of HCMV disease. For example, Letermovir is approved solely for prophylaxis, whereas Ganciclovir is approved for all three regimens; however, Ganciclovir has a higher toxicity profile compared to Letermovir. One strategy to address both toxicity concerns and the emergence of antiviral-resistant HCMV strains is the development of multidrug combination therapy. Studies on combination therapy for HCMV infection and disease have demonstrated significant improvements over monotherapy, particularly in patients infected with drug-resistant HCMV strains [Mylonakis et al., 2002; Cai et al., 2014; Vora et al., 2018]. However, the toxic effects of multidrug combination therapies have not been extensively studied. Drug-drug interactions can exacerbate toxicity depending on the specific combination used, although synergistic effects have also been observed [Lehar et al., 2009].

Developing antivirals with lower individual toxicity profiles, enhanced synergistic effects, and the ability to combat drug-resistant HCMV strains is at the core of this thesis. While such research is a long-term endeavor, this study specifically focuses on three HCMV proteins—pUL94/pUL99 complex, pUL98, and pUL77—to identify potential inhibitors for combination therapy. These targets play critical roles at different stages of the viral lifecycle: pUL98 is suspected to be involved in DNA replication, pUL77 is essential for the stability of DNA-packed capsids and contributes to DNA packaging, and pUL94/pUL99 is crucial for secondary envelopment. In vitro knockout cell culture studies have demonstrated the essential nature of these proteins for HCMV replication. Their involvement in distinct viral lifecycle stages presents a theoretical advantage for drug synergism and combating drug-resistant HCMV strains.

This study employs a structure-based drug discovery approach to identify potential inhibitors of the pUL94/pUL99 complex, pUL98, and pUL77 through virtual screening of a small molecule library. Structural characterization of target proteins is fundamental to structure-based drug discovery. While the structure of pUL77 is available (PDB: 7NXP, 8TEP) in the RCSB database, the structures of pUL94/pUL99 complex and pUL98 are not publicly available. Thus, this study focuses on structural characterization to allow for virtual screening and identification of small molecule hits against all three target proteins.

Virtual screening of one million small molecules from the ZINC22 library against the published pUL77 structure was initiated using AutoDock Vina. However, more than 60,000 molecules exhibited an estimated binding energy below -5 kcal/mol. Given the limited resources available, it was impractical to validate all 60,000 compounds in vitro, as a significant proportion were likely false positives. To improve the true positive rate of virtual screening hits and rationally select candidates for further validation through in vitro assays, a semi-automated consensus virtual screening pipeline was developed. This pipeline integrates AutoDock Vina, DiffDock, and AlphaFold3 using Python.

Consensus docking programs have been shown to enhance true positive rates compared to stand-alone docking tools [Morris et al., 2022]. Stand-alone docking programs—such as AutoDock Vina, AutoDock 4, Smina, Gnina, Glide, and Ledock—focus on improving sampling and scoring of docking poses [Eberhardt et al., 2021; Morris et al., 2009; Koes et al., 2013]. Advanced consensus docking pipelines, including 'dockECR' and 'MILCDock' [Ochoa et al., 2021; Morris et al., 2022], utilize multiple docking tools and employ complex scoring models to rank small molecules based on docking results. However, existing tools do not provide context-specific scores. Accurate predictions of docking poses and binding affinities do not necessarily translate into desired biological activity. While manual inspection and analysis of docked poses can offer insights into the probability of achieving the desired activity, this process has not been automated in published pipelines.

Although it remains impossible to fully predict a small molecule's mode of action solely from docking, this newly developed virtual screening pipeline takes a step in this direction. It sorts, clusters, and ranks compounds based on their interactions with specific amino acid residues in the target protein, in addition to analyzing docking poses using three different tools. The pipeline evaluates docking results, eliminates compounds with improbable binding poses, clusters hits based on the similarity of chemical groups interacting with target protein residues, and scores candidates based on the alignment and consistency of poses across the three tools.

Docking scores alone are meaningless if the pose is improbable to achieve the biological effect needed. For instance, if a small molecule does not bind to the chosen active site residues, a good docking score becomes irrelevant as the target protein cannot be inhibited. When millions of molecules are virtually screened, initial shortlisting relies on docking scores, as manually visualizing millions of poses is impractical. The presented pipeline automates the analysis, removing poses that fail to bind desired residues.

Clustering molecules based on their chemical signatures allows for shortlisting small molecules with diverse chemical properties. The pipeline includes functionality to cluster small molecules based on their overall structure or the substructures that interact with the chosen residues, ignoring parts of the molecule that do not contribute to the interaction with the target protein.

A scoring model was developed to evaluate small molecules based on the average atom-to-atom root mean squared distance and the coverage of active site/hotspot interfaces by the small molecules across poses from the three tools, providing a ranking within each cluster.

The goal of developing this pipeline, as part of this thesis, was to use it for screening small molecule inhibitors targeting HCMV proteins pUL77, pUL94/pUL99, and pUL98. The pipeline was utilized to screen one million small molecules to inhibit pUL77 multimerization, using the monomeric crystal structure of pUL77 available in public databases. This screening resulted in 48 potential protein-protein interaction inhibitors against pUL77, and 16 hits were ordered for activity assessment in in-vitro cell-based assays through collaboration. However, the structures of pUL94/pUL99 and pUL98 were unavailable, and their structural characterization formed the remaining part of the thesis's aim.

Extensive attempts to characterize pUL98 through X-ray crystallography, both alone and in complexes with oligos, crystallization chaperones such as sybodies and megabodies, were unsuccessful. Although megabody scaffolds were engineered for use as

crystallization chaperones, failed crystallization attempts prompted us to start the characterization of pUL98 with megabodies through cryoEM single-particle analysis. Initial feasibility studies based on one cryoEM dataset yielded a map at ~6 Å resolution (FSC 0.143). While this resolution was insufficient for atomic modeling, it justified further efforts in this direction. Due to limitations with microscope access, proposals were written to cryoEM facilities to collaborate for data collection.

While the main goal was structural characterization, new findings emerged regarding pUL98. Several nucleases from the DEK superfamily of exonucleases are known to dimerize to function. Although pUL98 dimerization was previously unreported, this study revealed that pUL98 dimerizes in solution in vitro. Additionally, deleting the intrinsically disordered region does not affect dimerization. The dimerization of pUL98 was concentration-dependent, with over 10% of the population forming dimers at slightly above 1 mg/mL, and over 90% at 10 mg/mL. A stretch of 20 amino acids near the N-terminal labelled as NCR, which is not conserved in alkaline nucleases from other human herpesviruses, was shown to play an essential role in protein folding but did not abolish enzymatic activity upon deletion. Double-stranded DNA was found to stabilize pUL98 more effectively than single-stranded DNA, with a difference of ~7°C suggesting a preference for double-stranded DNA.

This study also screened sybodies that bind to pUL98 and grouped them based on non-overlapping antigenic regions using bio-layer interferometry (BLI). These results pave the way for cell biology studies using sybodies in formats such as chromobodies, aimed at elucidating pUL98's function in the viral lifecycle, which remains unknown.

Efforts to characterize pUL94/pUL99 via X-ray crystallography began before this study but did not yield diffraction-quality crystals. Sybodies were engineered to assemble legobody scaffolds to improve cryoEM sample preparation. Strategies such as PEGylation and buffer screening using DLS were employed to reduce aggregation and dissociation of legobody_pUL94/pUL99 complexes during vitrification. These efforts resulted in properly dispersed samples on grids with negligible aggregation. Optimized sample preparation revealed that legobody_pUL94/pUL99 exhibited extreme heterogeneity, with flexible binding angles and components presenting continuous conformational landscapes. This posed challenges for cryoEM data processing. After extensive efforts with support from experienced cryoEM researchers (Dr. Guido Hansen, Dr. Guendalina Marini, Dr. Gonzalo Obal, Dr. Mauro Maiorca, and Dr. Siavash Mostafavi), resolutions of ~5.8 Å for pUL94/pUL99 and ~4.1 Å for the full legobody_pUL94/pUL99 complex were achieved. However, only partial volumes of pUL94/pUL99 were resolved due to its inherent flexibility and the map quality of pUL94/pUL99 is worse, restricting structural modelling. In the course of this thesis, a proposal for cryoEM access at the SOLEIL Polaris cryoEM beamline was approved, and further data collection is planned to achieve atomic resolution.

The finding that legobody scaffolds are inherently flexible has not been previously reported, providing valuable insights for future users. Structural characterization of the pUL94/pUL99 complex is approaching atomic resolution, while cryoEM characterization of pUL98 is still in its early stages but has been demonstrated to be feasible. Despite extensive efforts, both pUL94/pUL99 and pUL98 have been recalcitrant to crystallization. Obtaining diffraction-quality crystals of target proteins has always been a major bottleneck in X-ray crystallography. However, in cryoEM, it is possible to resolve the structures of relatively rigid proteins at atomic resolution without the need for

crystallization. Continuous flexibility within a target protein, as seen in the pUL94/pUL99 complex, significantly complicates structure determination.

The virtual screening pipeline I developed was intended for identifying small molecule hits against the pUL94/pUL99 complex and pUL98. However, the complexities associated with structural characterization of these proteins have delayed this goal. Once their structures are fully resolved, screening will be pursued in the future.

Functional studies for hits identified against pUL77 will help advance lead selection and optimization. Additionally, more small molecules will be screened using the pipeline in the future. Currently, the pipeline lacks functionality to assess stereochemical and ADME properties, but these features will be added to improve its utility. Computational efficiency will also be enhanced. The pipeline has already been applied to screen small molecules against another target from KSHV, identifying 430 potential hits, which are being validated in in-vitro assays. Continued structural characterization of pUL94/pUL99 and pUL98 will reveal more about their functions and facilitate structure-based drug discovery against HCMV.

6. References

Abdalla AE, Mahjoob MO, Abosalif KOA, Ejaz H, Alameen AAM, Elsaman T. Human cytomegalovirus-encoded MicroRNAs: A master regulator of latent infection. Infect Genet Evol. 2020 Mar; 78:104119. doi: 10.1016/j.meegid.2019.104119.

Adam BL, Jervey TY, Kohler CP, Wright GL Jr, Nelson JA, Stenberg RM. The human cytomegalovirus UL98 gene transcription unit overlaps with the pp28 true late gene (UL99) and encodes a 58-kilodalton early protein. J Virol. 1995 Sep;69(9):5304-10. doi: 10.1128/JVI.69.9.5304-5310.1995.

Adamson CS, Nevels MM. Bright and early: Inhibiting human cytomegalovirus by targeting major immediate-early gene expression or protein function. Viruses. 2020;12(1). doi:10.3390/v12010110

Agrawal G, Bader F, Günthner J, Wurzer S. Fast to First-in-Human: Getting New Medicines to Patients More Quickly Preclinical Development Can Benefit from Advances to Accelerate Innovation, Unlock Value, and Improve Quality.; 2023.

Ahlqvist J, Mocarski E. Cytomegalovirus UL103 Controls Virion and Dense Body Egress. Journal of Virology. 2011;85(10):5125-5135. doi:10.1128/jvi.01682-10

Alam Z, Al-Mahdi Z, Zhu Y, McKee Z, Parris DS, Parikh HI, Kellogg GE, Kuchta A, McVoy MA. Anti-cytomegalovirus activity of the anthraquinone atanyl blue PRL. Antiviral Res. 2015 Feb;114:86-95. doi: 10.1016/j.antiviral.2014.12.003.

Albert A. Relations between molecular structure 6501 and biological activity: stages in the evolution of current concepts. Annu Rev Pharmacol. 1971;11:13-36. doi: 10.1146/annurev.pa.11.040171.000305.

Alford RF, Leaver-Fay A, Jeliazkov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RL Jr, Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. J Chem Theory Comput. 2017 Jun 13;13(6):3031-3048. doi: 10.1021/acs.jctc.7b00125. Epub 2017 May 12. Erratum in: J Chem Theory Comput. 2022 Jul 12;18(7):4594. doi: 10.1021/acs.jctc.2c00500.

Alkhashrom S, Kicuntod J, Stillger K, et al. A Peptide Inhibitor of the Human Cytomegalovirus Core Nuclear Egress Complex. Pharmaceuticals. 2022;15(9). doi:10.3390/ph15091040

Aponte J, Vera JD, Chaar A, Martin D, Yency F. Cytomegalovirus Infection (CMV) and Human Immunodeficiency Virus: Diagnostic and Therapeutic Challenges in Virology. Archives of Microbiology & Immunology. 2018;02(02):18-35. doi:10.26502/ami.93650019

Archer MA, Brechtel TM, Davis LE, Parmar RC, Hasan MH, Tandon R. Inhibition of endocytic pathways impacts cytomegalovirus maturation. Scientific Reports. 2017;7. doi:10.1038/srep46069

Arend KC, Lenarcic EM, Vincent HA, Rashid N, Lazear E, McDonald IM, Gilbert TS, East MP, Herring LE, Johnson GL, Graves LM, Moorman NJ. Kinome Profiling Identifies Druggable Targets for Novel Human Cytomegalovirus (HCMV) Antivirals. Mol Cell Proteomics. 2017 Apr;16(4 suppl 1):S263-S276. doi: 10.1074/mcp.M116.065375.

Arkin MMR, Wells JA. Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. Nature Reviews Drug Discovery. 2004;3(4):301-317. doi:10.1038/nrd1343

Armour D, de Groot MJ, Edwards M, et al. The discovery of CCR5 receptor antagonists for the treatment of HIV infection: Hit-to-lead studies. ChemMedChem. 2006;1(7):706-709. doi:10.1002/cmdc.200600031

Arvin A, Campadelli-Fiume G, Mocarski E, et al., editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007. Available from: https://www.ncbi.nlm.nih.gov/books/NBK47376/

AuCoin DP, Smith GB, Meiering CD, Mocarski ES. Betaherpesvirus-Conserved Cytomegalovirus Tegument Protein ppUL32 (pp150) Controls Cytoplasmic Events during Virion Maturation. Journal of Virology. 2006;80(16):8199-8210. doi:10.1128/jvi.00457-06

Azevedo LS, Pierrotti LC, Abdala E, et al. Cytomegalovirus infection in transplant recipients. Clinics. 2015;70(7):515-523. doi:10.6061/clinics/2015(07)09

Azzeh M, Honigman A, Taraboulos A, Rouvinski A, Wolf DG. Structural changes in human cytomegalovirus cytoplasmic assembly sites in the absence of UL97 kinase activity. Virology. 2006 Oct 10;354(1):69-79. doi: 10.1016/j.virol.2006.05.037.

Bagnéris C, Briggs LC, Savva R, Ebrahimi B, Barrett TE. Crystal structure of a KSHV-SOX-DNA complex: Insights into the molecular mechanisms underlying DNase activity and host shutoff. Nucleic Acids Research. 2011;39(13):5744-5756. doi:10.1093/nar/gkr111

Baldanti F, Gerna G. Human cytomegalovirus resistance to antiviral drugs: Diagnosis, monitoring and clinical impact. Journal of Antimicrobial Chemotherapy. 2003;52(3):324-330. doi:10.1093/jac/dkg354

Baldick CJ Jr, Shenk T. Proteins associated with purified human cytomegalovirus particles. J Virol. 1996 Sep;70(9):6097-105. doi: 10.1128/JVI.70.9.6097-6105.1996.

Bale JF Jr, Petheram SJ, Souza IE, Murph JR. Cytomegalovirus reinfection in young children. J Pediatr. 1996 Mar;128(3):347-52. doi: 10.1016/s0022-3476(96)70279-2.

Ban C, Yang W. Structural Basis for MutH Activation in *E.coli* Mismatch Repair and Relationship of MutH to Restriction Endonucleases. Vol 17.; 1998.

Barnard DL, Fairbairn DW, O'Neill KL, Gage TL, Sidwell RW. Anti-human cytomegalovirus activity and toxicity of sulfonated anthraquinones and anthraquinone derivatives. Antiviral Res. 1995 Dec;28(4):317-29. doi: 10.1016/0166-3542(95)00057-7.

Barnard DL, Huffman JH, Morris JL, Wood SG, Hughes BG, Sidwell RW. Evaluation of the antiviral activity of anthraquinones, anthrones and anthraquinone derivatives against human cytomegalovirus. Antiviral Res. 1992 Jan;17(1):63-77. doi: 10.1016/0166-3542(92)90091-i.

Bateman A, Martin MJ, Orchard S, et al. UniProt: the Universal Protein Knowledgebase in 2025. Nucleic Acids Research. Published online November 18, 2024. doi:10.1093/nar/gkae1010

Beier H, Grimm M. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Res. 2001 Dec 1;29(23):4767-82. doi: 10.1093/nar/29.23.4767. PMID: 11726686;

Belabbas T, Yamada T, Tsuchiya Y, Suetsugu K, Egashira N, Ieiri I. Development and Full Validation of a Bioanalytical Method for Quantifying Letermovir in Human Plasma Using Ultra-Performance Liquid Chromatography Coupled with Mass Spectrometry. Vol 69.; 2021.

Benjamin Rauwel, Suk Min Jang, Marco Cassano, Adamandia Kapopoulou, Isabelle Barde, Didier Trono (2015) Release of human cytomegalovirus from latency by a KAP1/TRIM28 phosphorylation switch eLife 4:e06068. https://doi.org/10.7554/eLife.06068

Bergner T, Cortez Rayas L, Freimann G, Read C, von Einem J. Secondary Envelopment of Human Cytomegalovirus Is a Fast Process Utilizing the Endocytic Compartment as a Major Membrane Source. Biomolecules. 2024;14(9):1149. doi:10.3390/biom14091149

Berry R, Watson GM, Jonjic S, Degli-Esposti MA, Rossjohn J. Modulation of innate and adaptive immunity by cytomegaloviruses. Nature Reviews Immunology. 2020;20(2):113-127. doi:10.1038/s41577-019-0225-5

Bigalke JM, Heuser T, Nicastro D, Heldwein EE. Membrane deformation and scission by the HSV-1 nuclear egress complex. Nature Communications. 2014;5. doi:10.1038/ncomms5131

Biron KK. Antiviral drugs for cytomegalovirus diseases. Antiviral Res. 2006 Sep;71(2-3):154-63. doi: 10.1016/j.antiviral.2006.05.002

Blum M, Andreeva A, Florentino LC, et al. InterPro: the protein sequence classification resource in 2025. Nucleic Acids Research. Published online November 20, 2024. doi:10.1093/nar/gkae1082

Bodro M, Sabé N, Lladõ L, et al. Prophylaxis versus preemptive therapy for cytomegalovirus disease in high-risk liver transplant recipients. Liver Transplantation. 2012;18(9):1093-1099. doi:10.1002/lt.23460

Boeckh M, Geballe AP. Cytomegalovirus: Pathogen, paradigm, and puzzle. Journal of Clinical Investigation. 2011;121(5):1673-1680. doi:10.1172/JCI45449

Bogdanow B, Gruska I, Mühlberg L, et al. Spatially resolved protein map of intact human cytomegalovirus virions. Nature Microbiology. 2023;8(9):1732-1747. doi:10.1038/s41564-023-01433-8

Bogner E, Radsak K, Stinski MF. The gene product of human cytomegalovirus open reading frame UL56 binds the pac motif and has specific nuclease activity. J Virol. 1998 Mar;72(3):2259-64. doi: 10.1128/JVI.72.3.2259-2264.1998

Bonavita CM, Cardin RD. Don’t go breaking my heart: Mcmv as a model for hcmv-associated cardiovascular diseases. Pathogens. 2021;10(5). doi:10.3390/pathogens10050619

Bond SR, Naus CC. RF-Cloning.org: an online tool for the design of restriction-free cloning projects. Nucleic Acids Res. 2012 Jul;40(Web Server issue):W209-13. doi: 10.1093/nar/gks396

Borst EM, Bauerfeind R, Binz A, et al. The Essential Human Cytomegalovirus Proteins pUL77 and pUL93 Are Structural Components Necessary for Viral Genome Encapsidation. Journal of Virology. 2016;90(13):5860-5875. doi:10.1128/jvi.00384-16

Borst EM, Hahn G, Koszinowski UH, Messerle M. Cloning of the Human Cytomegalovirus (HCMV) Genome as an Infectious Bacterial Artificial Chromosome in Escherichia Coli: A New Approach for Construction of HCMV Mutants. Vol 73.; 1999. https://journals.asm.org/journal/jvi

Borst EM, Harmening S, Sanders S, et al. A Unique Role of the Human Cytomegalovirus Small Capsid Protein in Capsid Assembly. mBio. 2022;13(5). doi:10.1128/mbio.01007-22

Borst EM, Kleine-Albers J, Gabaev I, et al. The Human Cytomegalovirus UL51 Protein Is Essential for Viral Genome Cleavage-Packaging and Interacts with the Terminase Subunits pUL56 and pUL89. Journal of Virology. 2013;87(3):1720-1732. doi:10.1128/jvi.01955-12

Bosworth, Andrew; Atabani, Sowsan F.; Theodosiou, Anastasia; Shahi, Avneet; Peate, Tatiana; Wilson, Steven; Pelosi, Emanuela; Rosser, Andrew . (2020). Letermovir salvage therapy in the management of a case of cytomegalovirus ventriculitis complicated by drug resistance. Clinical Infection in Practice, 7-8(), 100039. doi:10.1016/j.clinpr.2020.100039

Bottino P, Pastrone L, Curtoni A, et al. Antiviral Approach to Cytomegalovirus Infection: An Overview of Conventional and Novel Strategies. Microorganisms. 2023;11(10). doi:10.3390/microorganisms11102372

Bou JV, Taguwa S, Matsuura Y. Trick-or-Trap: Extracellular Vesicles and Viral Transmission. Vaccines. 2023;11(10). doi:10.3390/vaccines11101532

Britt WJ, Jarvis M, Seo JY, Drummond D, Nelson J. Rapid Genetic Engineering of Human Cytomegalovirus by Using a Lambda Phage Linear Recombination System: Demonstration that pp28 (UL99) Is Essential for Production of Infectious Virus. Journal of Virology. 2004;78(1):539-543. doi:10.1128/jvi.78.1.539-543.2004

Britt WJ, Prichard MN. New therapies for human cytomegalovirus infections. Antiviral Res. 2018 Nov;159:153-174. doi: 10.1016/j.antiviral.2018.09.003.

Britt WJ. Congenital Human Cytomegalovirus Infection and the Enigma of Maternal Immunity. Journal of Virology. 2017;91(15). doi:10.1128/jvi.02392-16

Brock I, Krüger M, Mertens T, von Einem J. Nuclear Targeting of Human Cytomegalovirus Large Tegument Protein pUL48 Is Essential for Viral Growth. Journal of Virology. 2013;87(10):6005-6019. doi:10.1128/jvi.03558-12

Buchwald P. Small-molecule protein-protein interaction inhibitors: Therapeutic potential in light of molecular size, chemical space, and ligand binding efficiency considerations. IUBMB Life. 2010;62(10):724-731. doi:10.1002/iub.383

Burgen AS. Conformational changes and drug action. Federation Proceedings. 1981 Nov;40(13):2723-2728.

Butt BG, Fischer D, Rep AR, et al. Human cytomegalovirus deploys molecular mimicry to recruit VPS4A to sites of virus assembly. PLoS Pathogens. 2024;20(6). doi:10.1371/journal.ppat.1012300

Butt BG, Owen DJ, Jeffries CM, et al. Insights into herpesvirus assembly from the structure of the pUL7:PUL51 complex. eLife. 2020;9:1-40. doi:10.7554/eLife.53789

Buxmann H, Hamprecht K, Meyer-Wittkopf M, Friese K. Primary human cytomegalovirus (HCMV) infection in pregnancy. Deutsches Arzteblatt International. 2017;114(4):45-52. doi:10.3238/arztebl.2017.0045

Cai H, Kapoor A, He R, et al. In vitro combination of anti-cytomegalovirus compounds acting through different targets: Role of the slope parameter and insights into mechanisms of action. *Antimicrobial Agents and Chemotherapy*. 2014;58(2):986-994. doi:10.1128/AAC.01972-13

Caló S, Cortese M, Ciferri C, et al. The Human Cytomegalovirus UL116 Gene Encodes an Envelope Glycoprotein Forming a Complex with gH Independently from gL . Journal of Virology. 2016;90(10):4926-4938. doi:10.1128/jvi.02517-15

Campos AB, Ribeiro J, Boutolleau D, Sousa H. Human cytomegalovirus antiviral drug resistance in hematopoietic stem cell transplantation: Current state of the art. Reviews in Medical Virology. 2016;26(3):161-182. doi:10.1002/rmv.1873

Cannon MJ, Schmid DS, Hyde TB. Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. Reviews in Medical Virology. 2010;20(4):202-213. doi:10.1002/rmv.655

Cao K, Zhang Y, Yao Q, Peng Y, Pan Q, Jiao Q, Ren K, Sun F, Zhang Q, Guo R, Zhang J, Chen T. Hypericin blocks the function of HSV-1 alkaline nuclease and suppresses viral replication. J Ethnopharmacol. 2022 Oct 5;296:115524. doi: 10.1016/j.jep.2022.115524.

Caragliano E, Bonazza S, Frascaroli G, et al. Human cytomegalovirus forms phase-separated compartments at viral genomes to facilitate viral replication. Cell Reports. 2022;38(10). doi:10.1016/j.celrep.2022.110469

Cepeda V, Fraile-Ramos A. A role for the SNARE protein syntaxin 3 in human cytomegalovirus morphogenesis. Cellular Microbiology. 2011;13(6):846-858. doi:10.1111/j.1462-5822.2011.01583.x

Chadha P, Han J, Starkey JL, Wills JW. Regulated Interaction of Tegument Proteins UL16 and UL11 from Herpes Simplex Virus. Journal of Virology. 2012;86(21):11886-11898. doi:10.1128/jvi.01879-12

Chang Y, Hawkins BA, Du JJ, Groundwater PW, Hibbs DE, Lai F. A Guide to In Silico Drug Design. Pharmaceutics. 2023;15(1). doi:10.3390/pharmaceutics15010049

Chaudhury S, Lyskov S, Gray JJ. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. Bioinformatics. 2010;26(5):689-691. doi:10.1093/bioinformatics/btq007

Chayen NE, Saridakis E, Sear RP. Experiment and theory for heterogeneous nucleation of protein crystals in a porous media. Proc Natl Acad Sci U S A. 2006 Jan 17;103(3):597-601. doi: 10.1073/pnas.0504860102.

Chee MS, Bankier AT, Beck S, Bohni R, Brown CM, Cerny R, Horsnell T, Hutchison CA 3rd, Kouzarides T, Martignetti JA, et al. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr Top Microbiol Immunol. 1990;154:125-69. doi: 10.1007/978-3-642-74980-3_6.

Chen DH, Jiang H, Lee M, Liu F, Zhou ZH. Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. Virology. 1999 Jul 20;260(1):10-6. doi: 10.1006/viro.1999.9791.

Chen SJ, Wang SC, Chen YC. Challenges, Recent Advances and Perspectives in the Treatment of Human Cytomegalovirus Infections. Tropical Medicine and Infectious Disease. 2022;7(12). doi:10.3390/tropicalmed7120439

Cheng AC, Coleman RG, Smyth KT, et al. Structure-based maximal affinity model predicts small-molecule druggability. Nature Biotechnology. 2007;25(1):71-75. doi:10.1038/nbt1273

Chevillotte M, Landwehr S, Linta L, et al. Major Tegument Protein pp65 of Human Cytomegalovirus Is Required for the Incorporation of pUL69 and pUL97 into the Virus Particle and for Viral Growth in Macrophages. Journal of Virology. 2009;83(6):2480-2490. doi:10.1128/jvi.01818-08

Chiopris G, Veronese P, Cusenza F, et al. Congenital cytomegalovirus infection: Update on diagnosis and treatment. Microorganisms. 2020;8(10):1-17. doi:10.3390/microorganisms8101516

Chou S, Satterwhite LE, Ercolani RJ. New Locus of Drug Resistance in the Human Cytomegalovirus UL56 Gene Revealed by In Vitro Exposure to Letermovir and Ganciclovir. Published online 2018. doi:10.1128/AAC

Close WL, Glassbrook JE, Gurczynski SJ, Pellett PE. Infection-induced changes within the endocytic recycling compartment suggest a roadmap of human cytomegalovirus egress. Frontiers in Microbiology. 2018;9(AUG). doi:10.3389/fmicb.2018.01888

Congenital Cytomegalovirus Infection: Update on Treatment: Scientific Impact Paper No. 56. BJOG: An International Journal of Obstetrics and Gynaecology. 2018;125(1):e1-e11. doi:10.1111/1471-0528.14836

Cook KC, Tsopurashvili E, Needham JM, Thompson SR, Cristea IM. Restructured membrane contacts rewire organelles for human cytomegalovirus infection. Nature Communications. 2022;13(1). doi:10.1038/s41467-022-32488-6

Corso G, Stärk H, Jing B, Barzilay R, Jaakkola T. DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. Published online October 4, 2022. http://arxiv.org/abs/2210.01776

Craighead JE, Kanich RE, Almeida JD. Nonviral microbodies with viral antigenicity produced in cytomegalovirus-infected cells. J Virol. 1972 Oct;10(4):766-75. doi: 10.1128/JVI.10.4.766-775.1972.

Crawford LB, Diggins NL, Caposio P, Hancock MH. Advances in Model Systems for Human Cytomegalovirus Latency and Reactivation. mBio. 2022 Feb 22;13(1):e0172421. doi: 10.1128/mbio.01724-21.

Crough T, Khanna R. Immimobiology of human cytomegalovirus: From bench to bedside. Clinical Microbiology Reviews. 2009;22(1):76-98. doi:10.1128/CMR.00034-08

Crump CM, Hung CH, Thomas L, Wan L, Thomas G. Role of PACS-1 in Trafficking of Human Cytomegalovirus Glycoprotein B and Virus Production. Journal of Virology. 2003;77(20):11105-11113. doi:10.1128/jvi.77.20.11105-11113.2003

Cullen PJ, Steinberg F. To degrade or not to degrade: mechanisms and significance of endocytic recycling. Nature Reviews Molecular Cell Biology. 2018;19(11):679-696. doi:10.1038/s41580-018-0053-7

Dahlroth SL, Gurmu D, Haas J, Erlandsen H, Nordlund P. Crystal structure of the shutoff and exonuclease protein from the oncogenic Kaposi's sarcoma-associated herpesvirus. FEBS Journal. 2009;276(22):6636-6645. doi:10.1111/j.1742-4658.2009.07374.x

Dai X, Yu X, Gong H, et al. The Smallest Capsid Protein Mediates Binding of the Essential Tegument Protein pp150 to Stabilize DNA-Containing Capsids in Human Cytomegalovirus. PLoS Pathogens. 2013;9(8). doi:10.1371/journal.ppat.1003525

Das S, Ortiz DA, Gurczynski SJ, Khan F, Pellett PE. Identification of Human Cytomegalovirus Genes Important for Biogenesis of the Cytoplasmic Virion Assembly Complex. Journal of Virology. 2014;88(16):9086-9099. doi:10.1128/jvi.01141-14

Das S, Pellett PE. Spatial Relationships between Markers for Secretory and Endosomal Machinery in Human Cytomegalovirus-Infected Cells versus Those in Uninfected Cells. Journal of Virology. 2011;85(12):5864-5879. doi:10.1128/jvi.00155-11

Das S, Vasanji A, Pellett PE. Three-Dimensional Structure of the Human Cytomegalovirus Cytoplasmic Virion Assembly Complex Includes a Reoriented Secretory Apparatus. Journal of Virology. 2007;81(21):11861-11869. doi:10.1128/jvi.01077-07

de Clercq E, Li G. Approved antiviral drugs over the past 50 years. Clinical Microbiology Reviews. 2016;29(3):695-747. doi:10.1128/CMR.00102-15

de Lepper M, Stephan AJ, Wölle R, et al. Burden of sequelae and healthcare resource utilization in the first year of life in infants born with congenital cytomegalovirus (cCMV) infection in Germany: A retrospective statutory health insurance claims database analysis. PLoS ONE. 2023;18(11 November). doi:10.1371/journal.pone.0293869

de Vries JJC, van Zwet EW, Dekker FW, Kroes ACM, Verkerk PH, Vossen ACTM. The apparent paradox of maternal seropositivity as a risk factor for congenital cytomegalovirus infection: A population-based prediction model. Reviews in Medical Virology. 2013;23(4):241-249. doi:10.1002/rmv.1744

DeFilippis VR, Sali T, Alvarado D, White L, Bresnahan W, Früh KJ. Activation of the Interferon Response by Human Cytomegalovirus Occurs via Cytoplasmic Double-Stranded DNA but Not Glycoprotein B. Journal of Virology. 2010;84(17):8913-8925. doi:10.1128/jvi.00169-10

Deller MC, Kong L, Rupp B. Protein stability: A crystallographer's perspective. Acta Crystallographica Section:F Structural Biology Communications. 2016;72:72-95. doi:10.1107/S2053230X15024619

Deray KGV, Danziger-Isakov LA, Downes KJ. Current and Emerging Antiviral Agents in the Prevention and Treatment of Cytomegalovirus in Pediatric Transplant Recipients. Journal of the Pediatric Infectious Diseases Society. 2024;13:S14-S21. doi:10.1093/jpids/piad059

DeRussy BM, Boland MT, Tandon R. Human Cytomegalovirus pUL93 Links Nucleocapsid Maturation and Nuclear Egress. Journal of Virology. 2016;90(16):7109-7117. doi:10.1128/jvi.00728-16

DeRussy BM, Tandon R. Human Cytomegalovirus pUL93 Is Required for Viral Genome Cleavage and Packaging. Journal of Virology. 2015;89(23):12221-12225. doi:10.1128/jvi.02382-15

Dessau MA, Modis Y. Protein crystallization for X-ray crystallography. Journal of Visualized Experiments. 2010;(47). doi:10.3791/2285

Dietz AN, Villinger C, Becker S, Frick M, von Einem J. A Tyrosine-Based Trafficking Motif of the Tegument Protein pUL71 Is Crucial for Human Cytomegalovirus Secondary Envelopment. Journal of Virology. 2018;92(1). doi:10.1128/jvi.00907-17

Dittmer A, Bogner E. Analysis of the quaternary structure of the putative HCMV portal protein PUL104. Biochemistry. 2005;44(2):759-765. doi:10.1021/bi047911w

Dittmer A, Drach JC, Townsend LB, Fischer A, Bogner E. Interaction of the Putative Human Cytomegalovirus Portal Protein pUL104 with the Large Terminase Subunit pUL56 and Its Inhibition by Benzimidazole- d -Ribonucleosides . Journal of Virology. 2005;79(23):14660-14667. doi:10.1128/jvi.79.23.14660-14667.2005

Dittmer D, Mocarski ES. Human cytomegalovirus infection inhibits G1/S transition. J Virol. 1997 Feb;71(2):1629-34. doi: 10.1128/JVI.71.2.1629-1634.1997.

do Carmo AM, Santos FM, Ortiz-Agostinho CL, et al. Cytomegalovirus infection in inflammatory bowel disease is not associated with worsening of intestinal inflammatory activity. PLoS ONE. 2014;9(11). doi:10.1371/journal.pone.0111574

Döhner K, Cornelius A, Serrero MC, Sodeik B. The journey of herpesvirus capsids and genomes to the host cell nucleus. Curr Opin Virol. 2021 Oct;50:147-158. doi: 10.1016/j.coviro.2021.08.005.

Döhner K, Serrero MC, Sodeik B. The role of nuclear pores and importins for herpes simplex virus infection. Curr Opin Virol. 2023 Oct;62:101361. doi: 10.1016/j.coviro.2023.101361.

Dolan A, Cunningham C, Hector RD, et al. Genetic content of wild-type human cytomegalovirus. Journal of General Virology. 2004;85(5):1301-1312. doi:10.1099/vir.0.79888-0

Domingo W, Nguyen IT, Johnsrud JJ, Brown JW. Continuous-Infusion Foscarnet Facilitates Administration in Hematopoietic Stem Cell Transplantation Patients. Transplant Cell Ther. 2021 Jul;27(7):622.e1-622.e5. doi: 10.1016/j.jtct.2021.03.018.

Dong A, Xu X, Edwards AM, et al. In situ proteolysis for protein crystallization and structure determination. *Nature Methods*. 2007;4(12):1019-1021. doi:10.1038/nmeth1118

Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H, Liu F. Functional profiling of a human cytomegalovirus genome. Proc Natl Acad Sci U S A. 2003 Nov 25;100(24):14223-8. doi: 10.1073/pnas.2334032100.

Dupont L, Reeves MB. Cytomegalovirus latency and reactivation: recent insights into an age old problem. Reviews in Medical Virology. Published online March 1, 2016:75-89. doi:10.1002/rmv.1862

Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. Journal of Chemical Information and Modeling. 2021;61(8):3891-3898. doi:10.1021/acs.jcim.1c00203

Eckle T, Lang P, Prix L, et al. Rapid development of ganciclovir-resistant cytomegalovirus infection in children after allogeneic stem cell transplantation in the early phase of immune cell recovery. Bone Marrow Transplantation. 2002;30(7):433-439. doi:10.1038/sj.bmt.1703666

Ecsédi P, Gógl G, Hóf H, Kiss B, Harmat V, Nyitray L. Structure Determination of the Transactivation Domain of p53 in Complex with S100A4 Using Annexin A2 as a Crystallization Chaperone. Structure. 2020;28(8):943-953.e4. doi:10.1016/j.str.2020.05.001

Ekins S, Mestres J, Testa B. In silico pharmacology for drug discovery: Methods for virtual ligand screening and profiling. British Journal of Pharmacology. 2007;152(1):9-20. doi:10.1038/sj.bjp.0707305

el Helou G, Razonable RR. Safety considerations with current and emerging antiviral therapies for cytomegalovirus infection in transplantation. Expert Opinion on Drug Safety. 2019;18(11):1017-1030. doi:10.1080/14740338.2019.1662787

Emmerich CH, Gamboa LM, Hofmann MCJ, et al. Improving target assessment in biomedical research: the GOT-IT recommendations. Nature Reviews Drug Discovery. 2021;20(1):64-81. doi:10.1038/s41573-020-0087-3

Erdős G, Dosztányi Z. Analyzing Protein Disorder with IUPred2A. Current Protocols in Bioinformatics. 2020;70(1). doi:10.1002/cpbi.99

Faulds D, Heel RC. Ganciclovir. A review of its antiviral activity, pharmacokinetic properties and therapeutic efficacy in cytomegalovirus infections. Drugs. 1990 Apr;39(4):597-638. doi: 10.2165/00003495-199039040-00008.

Feng B, Li C, Zhang Z, et al. A shark-derived broadly neutralizing nanobody targeting a highly conserved epitope on the S2 domain of sarbecoviruses. *Journal of Nanobiotechnology* . 2025;23(1). doi:10.1186/s12951-025-03150-2

Feng X, Schröer J, Yu D, Shenk T. Human Cytomegalovirus pUS24 Is a Virion Protein That Functions Very Early in the Replication Cycle. Journal of Virology. 2006;80(17):8371-8378. doi:10.1128/jvi.00399-06

Fernández-Moreno R, Torre-Cisneros J, Cantisán S. Human cytomegalovirus (HCMV)-encoded microRNAs: potential biomarkers and clinical applications. RNA Biology. 2021;18(12):2194-2202. doi:10.1080/15476286.2021.1930757

Filikov AV, James TL. Structure-based design of ligands for protein basic domains: application to the HIV-1 Tat protein. J Comput Aided Mol Des. 1998 May;12(3):229-40. doi: 10.1023/a:1007949625270.

Fletcher-Etherington A, Nobre L, Nightingale K, Antrobus R, Nichols J, Davison AJ, Stanton RJ, Weekes MP. Human cytomegalovirus protein pUL36: A dual cell death pathway inhibitor. Proc Natl Acad Sci U S A. 2020 Aug 4;117(31):18771-18779. doi: 10.1073/pnas.2001887117.

Flomm FJ, Soh TK, Schneider C, et al. Intermittent bulk release of human cytomegalovirus. PLoS Pathogens. 2022;18(8). doi:10.1371/journal.ppat.1010575

Foglierini M, Marcandalli J, Perez L. HCMV envelope glycoprotein diversity demystified. Frontiers in Microbiology. 2019;10(MAY). doi:10.3389/fmicb.2019.01005

Fowler K, Mucha J, Neumann M, et al. A systematic literature review of the global seroprevalence of cytomegalovirus: possible implications for treatment, screening, and vaccine development. BMC Public Health. 2022;22(1). doi:10.1186/s12889-022-13971-7

Fu YZ, Guo Y, Zou HM, et al. Human cytomegalovirus protein UL42 antagonizes cGAS/MITA-mediated innate antiviral response. PLoS Pathogens. 2019;15(5). doi:10.1371/journal.ppat.1007691

Fu YZ, Su S, Gao YQ, Wang PP, Huang ZF, Hu MM, Luo WW, Li S, Luo MH, Wang YY, Shu HB. Human Cytomegalovirus Tegument Protein UL82 Inhibits STING-Mediated Signaling to Evade Antiviral Immunity. Cell Host Microbe. 2017 Feb 8;21(2):231-243. doi: 10.1016/j.chom.2017.01.001.

Fung M, DeVoe C, Spottiswoode N, Doernberg SB. Maribavir for Cytomegalovirus Treatment in the Real World—Not a Silver Bullet. Open Forum Infectious Diseases. 2023;10(1). doi:10.1093/ofid/ofac686

Furukawa H, Inaba H, Sasaki Y, Akiyoshi K, Matsuura K. Embedding a membrane protein into an enveloped artificial viral replica. RSC Chemical Biology. 2022;3(2):231-241. doi:10.1039/d1cb00166c

Gangrade, D., Sawant, G., & Mehta, A. (2016). Re-thinking drug discovery: in silico method. J Chem Pharmaceutical Res, 8(8), 1092-1099.

Gao J, Hay TJM, Banfield BW. The Product of the Herpes Simplex Virus 2 UL16 Gene Is Critical for the Egress of Capsids from the Nuclei of Infected Cells. Journal of Virology. 2017;91(10). doi:10.1128/jvi.00350-17

Gao M, Robertson BJ, McCann PJ, O'Boyle DR, Weller SK, Newcomb WW, Brown JC, Weinheimer SP. Functional conservations of the alkaline nuclease of herpes simplex type 1 and human cytomegalovirus. Virology. 1998 Sep 30;249(2):460-70. doi: 10.1006/viro.1998.9344. PMID: 9791036.

Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opinion on Drug Discovery. 2015;10(5):449-461. doi:10.1517/17460441.2015.1032936

Gentry BG, Bogner E, Drach JC. Targeting the terminase: An important step forward in the treatment and prophylaxis of human cytomegalovirus infections. Antiviral Res. 2019 Jan;161:116-124. doi: 10.1016/j.antiviral.2018.11.005.

Geppert T, Hoy B, Wessler S, Schneider G. Context-based identification of protein-protein interfaces and "hot-spot" residues. Chemistry and Biology. 2011;18(3):344-353. doi:10.1016/j.chembiol.2011.01.005

Gerna G, Fornara C, Furione M, Lilleri D. Congenital human cytomegalovirus infection: A narrative review of maternal immune response and diagnosis in view of the development of a vaccine and prevention of primary and non-primary infections in pregnancy. Microorganisms. 2021;9(8). doi:10.3390/microorganisms9081749

Glotzbecker B, Duncan C, Alyea E 3rd, Campbell B, Soiffer R. Important drug interactions in hematopoietic stem cell transplantation: what every physician should know. Biol Blood Marrow Transplant. 2012 Jul;18(7):989-1006. doi: 10.1016/j.bbmt.2011.11.029.

Goldstein JN, Weller SK. In vitro processing of herpes simplex virus type 1 DNA replication intermediates by the viral alkaline nuclease, UL12. J Virol. 1998 Nov;72(11):8772-81. doi: 10.1128/JVI.72.11.8772-8781.1998.

Goodrum F, Caviness K, Zagallo P. Human cytomegalovirus persistence. Cellular Microbiology. 2012;14(5):644-655. doi:10.1111/j.1462-5822.2012.01774.x

Gorrec F, Bellini D. The FUSION protein crystallization screen. Journal of Applied Crystallography. 2022;55:310-319. doi:10.1107/S1600576722001765

Gourin C, Alain S, Hantz S. Anti-CMV therapy, what next? A systematic review. Frontiers in Microbiology. 2023;14. doi:10.3389/fmicb.2023.1321116

Grady LM, Szczepaniak R, Murelli RP, Masaoka T, Le Grice SFJ, Wright DL, Weller SK. The Exonuclease Activity of Herpes Simplex Virus 1 UL12 Is Required for Production of Viral DNA That Can Be Packaged To Produce Infectious Virus. J Virol. 2017 Nov 14;91(23):e01380-17. doi: 10.1128/JVI.01380-17.

Green DR. A BH3 Mimetic for Killing Cancer Cells. Cell. 2016;165(7):1560. doi:10.1016/j.cell.2016.05.080

Grey F, Meyers H, White EA, Spector DH, Nelson J. A human cytomegalovirus-encoded microRNA regulates expression of multiple viral genes involved in replication. PLoS Pathogens. 2007;3(11):1593-1602. doi:10.1371/journal.ppat.0030163

Grey F. Role of microRNAs in herpesvirus latency and persistence. Journal of General Virology. 2015;96(4):739-751. doi:10.1099/vir.0.070862-0

Grgic I, Gorenec L. Human Cytomegalovirus (HCMV) Genetic Diversity, Drug Resistance Testing and Prevalence of the Resistance Mutations: A Literature Review. Tropical Medicine and Infectious Disease. 2024;9(2). doi:10.3390/tropicalmed9020049

Griffante G, Gugliesi F, Pasquero S, et al. Human cytomegalovirus-induced host protein citrullination is crucial for viral replication. Nature Communications. 2021;12(1). doi:10.1038/s41467-021-24178-6

Griffiths P, Baraniak I, Reeves M. The pathogenesis of human cytomegalovirus. Journal of Pathology. 2015;235(2):288-297. doi:10.1002/path.4437

Griffiths P, Reeves M. Pathogenesis of human cytomegalovirus in the immunocompromised host. Nature Reviews Microbiology. 2021;19(12):759-773. doi:10.1038/s41579-021-00582-z

Griffiths P. The direct and indirect consequences of cytomegalovirus infection and potential benefits of vaccination. Antiviral Research. 2020;176. doi:10.1016/j.antiviral.2020.104732

Griffiths PD. Burden of disease associated with human cytomegalovirus and prospects for elimination by universal immunisation. The Lancet Infectious Diseases. 2012;12(10):790-798. doi:10.1016/S1473-3099(12)70197-4

Guo H, Shen S, Wang L, Deng H. Role of tegument proteins in herpesvirus assembly and egress. Protein and Cell. 2010;1(11):987-998. doi:10.1007/s13238-010-0120-0

Hagen C, Dent KC, Zeev-Ben-Mordehai T, et al. Structural Basis of Vesicle Formation at the Inner Nuclear Membrane. Cell. 2015;163(7):1692-1701. doi:10.1016/j.cell.2015.11.029

Halpern-Cohen V, Blumberg EA. New Perspectives on Antimicrobial Agents: Maribavir. Antimicrobial Agents and Chemotherapy. 2022;66(9). doi:10.1128/aac.02405-21

Hamilton ST, Milbradt J, Marschall M, Rawlinson WD. Human cytomegalovirus replication is strictly inhibited by siRNAs targeting UL54, UL97 or UL122/123 gene transcripts. PLoS ONE. 2014;9(6). doi:10.1371/journal.pone.0097231

Hansch C, Fujita Vol T, Roberts RB, et al. Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. Vol 39.; 1953.

Hansch C. Quantitative relationships between lipophilic character and drug metabolism. Drug Metabolism Reviews. 1972;1(1):1-13. doi:10.3109/03602537208993906

Harihar B, Saravanan KM, Gromiha MM, Selvaraj S. Importance of Inter-residue Contacts for Understanding Protein Folding and Unfolding Rates, Remote Homology, and Drug Design. Molecular Biotechnology. Published online 2024. doi:10.1007/s12033-024-01119-4

Harper AL, Meckes DG, Marsh JA, et al. Interaction Domains of the UL16 and UL21 Tegument Proteins of Herpes Simplex Virus. Journal of Virology. 2010;84(6):2963-2971. doi:10.1128/jvi.02015-09

Harris D Riley JR, 1997, history of cytomegalovirus.

Hartenian E, Mendez AS, Didychuk AL, Khosla S, Glaunsinger BA. DNA processing by the Kaposi's sarcoma-associated herpesvirus alkaline exonuclease SOX contributes to viral gene expression and infectious virion production. Nucleic Acids Research. 2023;51(1):182-197. doi:10.1093/nar/gkac1190

Hasegawa J, Hatakeyama S, Wakai S, et al. Preemptive anti-cytomegalovirus therapy in high-risk (donor-positive, recipient-negative cytomegalovirus serostatus) kidney transplant recipients. International Journal of Infectious Diseases. 2017;65:50-56. doi:10.1016/j.ijid.2017.09.023

Hashimoto Y, Sheng X, Murray-Nerger LA, Cristea IM. Temporal dynamics of protein complex formation and dissociation during human cytomegalovirus infection. Nature Communications. 2020;11(1). doi:10.1038/s41467-020-14586-5

Hassan Z, Kumar ND, Reggiori F, Khan G. How viruses hijack and modify the secretory transport pathway. Cells. 2021;10(10). doi:10.3390/cells10102535

He HP, Gao S. Structure Overview of Herpesvirus Tegument Proteins. Published online November 2, 2023. doi:10.20944/preprints202311.0147.v1

He HP, Luo M, Cao YL, et al. Structure of Epstein-Barr virus tegument protein complex BBRF2-BSRF1 reveals its potential role in viral envelopment. Nature Communications. 2020;11(1). doi:10.1038/s41467-020-19259-x

Hein MY, Weissman JS. Functional single-cell genomics of human cytomegalovirus infection. Nature Biotechnology. 2022;40(3):391-401. doi:10.1038/s41587-021-01059-3

Heming JD, Conway JF, Homa FL. Herpesvirus capsid assembly and DNA packaging. In: Advances in Anatomy Embryology and Cell Biology. Vol 223. Springer Verlag; 2017:119-142. doi:10.1007/978-3-319-53168-7_6

Heng HH. The Rationale and Challenges of Molecular Medicine. In: Genome Chaos. Elsevier; 2019:427-479. doi:10.1016/b978-0-12-813635-5.00008-2

Hennig T, O'Hare P. Viruses and the nuclear envelope. Curr Opin Cell Biol. 2015 Jun;34:113-21. doi: 10.1016/j.ceb.2015.06.002.

Hensel G, Meyer H, Gärtner S, Brand G, Kern HF. Nuclear localization of the human cytomegalovirus tegument protein pp150 (ppUL32). J Gen Virol. 1995 Jul;76 (Pt 7):1591-601. doi: 10.1099/0022-1317-76-7-1591.

Ho JSY, Zhu Z, Marazzi I. Unconventional viral gene expression mechanisms as therapeutic targets. Nature. 2021;593(7859):362-371. doi:10.1038/s41586-021-03511-5

Ho M, Suwansirikul S, Dowling JN, Youngblood LA, Armstrong JA. The transplanted kidney as a source of cytomegalovirus infection. N Engl J Med. 1975 Nov 27;293(22):1109-12. doi: 10.1056/NEJM197511272932201

Ho M. The history of cytomegalovirus and its diseases. Medical Microbiology and Immunology. 2008;197(2):65-73. doi:10.1007/s00430-007-0066-x

Hoehl S, Berger A, Ciesek S, Rabenau HF. Thirty years of CMV seroprevalence—a longitudinal analysis in a German university hospital. European Journal of Clinical Microbiology and Infectious Diseases. 2020;39(6):1095-1102. doi:10.1007/s10096-020-03814-x

Hogue IB. Tegument assembly, secondary envelopment and exocytosis. Current Issues in Molecular Biology. 2021;42:551-604. doi:10.21775/cimb.042.551

Homman-Loudiyi M, Hultenby K, Britt W, Söderberg-Nauclér C. Envelopment of Human Cytomegalovirus Occurs by Budding into Golgi-Derived Vacuole Compartments Positive for gB, Rab 3, Trans-Golgi Network 46, and Mannosidase II. Journal of Virology. 2003;77(14):8179-8179. doi:10.1128/jvi.77.14.8179.2003

Hong, J.C., Kahan, B.D. (2001). The History of Immunosuppression for Organ Transplantation. In: Sayegh, M.H., Remuzzi, G. (eds) Current and Future Immunosuppressive Therapies Following Transplantation. Springer, Dordrecht. https://doi.org/10.1007/978-94-010-1005-4_1

Hönig SMN, Lemmen C, Rarey M. Small molecule superposition: A comprehensive overview on pose scoring of the latest methods. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2023;13(2). doi:10.1002/wcms.1640

Horst D, Burmeister WP, Boer IGJ, et al. The “Bridge” in the Epstein-Barr Virus Alkaline Exonuclease Protein BGLF5 Contributes to Shutoff Activity during Productive Infection. Journal of Virology. 2012;86(17):9175-9187. doi:10.1128/jvi.00309-12

Hou T, Wang J, Li Y, Wang W. Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. the accuracy of ranking poses generated from docking. Journal of Computational Chemistry. 2011;32(5):866-877. doi:10.1002/jcc.21666

Hsiang CY, Ho TY. Emodin is a novel alkaline nuclease inhibitor that suppresses herpes simplex virus type 1 yields in cell cultures. British Journal of Pharmacology. 2008;155(2):227-235. doi:10.1038/bjp.2008.242

Hu F, Ma Y, Peng X. Does ganciclovir exert retinal toxicity after multiple continuous intravitreal injections? BMC Infectious Diseases. 2021;21(1). doi:10.1186/s12879-021-06365-4

Huang ZF, Zou HM, Liao BW, Zhang HY, Yang Y, Fu YZ, Wang SY, Luo MH, Wang YY. Human Cytomegalovirus Protein UL31 Inhibits DNA Sensing of cGAS to Mediate Immune Evasion. Cell Host Microbe. 2018 Jul 11;24(1):69-80.e4. doi: 10.1016/j.chom.2018.05.007.

Hughes JP, Rees SS, Kalindjian SB, Philpott KL. Principles of early drug discovery. British Journal of Pharmacology. 2011;162(6):1239-1249. doi:10.1111/j.1476-5381.2010.01127.x

Imlay HN, Kaul DR. Letermovir and Maribavir for the Treatment and Prevention of Cytomegalovirus Infection in Solid Organ and Stem Cell Transplant Recipients. Clinical Infectious Diseases. 2021;73(1):156-160. doi:10.1093/cid/ciaa1713

Irmiere A, Gibson W. Isolation and characterization of a noninfectious virion-like particle released from cells infected with human strains of cytomegalovirus. Virology. 1983 Oct 15;130(1):118-33. doi: 10.1016/0042-6822(83)90122-8.

Isomura H, Stinski MF, Murata T, et al. The Human Cytomegalovirus Gene Products Essential for Late Viral Gene Expression Assemble into Prereplication Complexes before Viral DNA Replication. Journal of Virology. 2011;85(13):6629-6644. doi:10.1128/jvi.00384-11

Ivanov AA, Khuri FR, Fu H. Targeting protein-protein interactions as an anticancer strategy. Trends in Pharmacological Sciences. 2013;34(7):393-400. doi:10.1016/j.tips.2013.04.007

Iwan Zimmermann, Pascal Egloff, Cedric AJ Hutter, Fabian M Arnold, Peter Stohler, Nicolas Bocquet, Melanie N Hug, Sylwia Huber, Martin Siegrist, Lisa Hetemann, Jennifer Gera, Samira Gmür, Peter Spies, Daniel Gygax, Eric R Geertsma, Roger JP Dawson, Markus A Seeger (2018) Synthetic single domain antibodies for the conformational trapping of membrane proteins eLife 7:e34317 https://doi.org/10.7554/eLife.34317

James SH, Kimberlin DW. Advances in the prevention and treatment of congenital cytomegalovirus infection. Current Opinion in Pediatrics. 2016;28(1):81-85. doi:10.1097/MOP.0000000000000305

Janin J, Bahadur RP, Chakrabarti P. Protein-protein interaction and quaternary structure. Quarterly Reviews of Biophysics. 2008;41(2):133-180. doi:10.1017/S0033583508004708

Jean Beltran PM, Cristea IM. The life cycle and pathogenesis of human cytomegalovirus infection: Lessons from proteomics. Expert Review of Proteomics. 2014;11(6):697-711. doi:10.1586/14789450.2014.971116

Jean Beltran PM, Mathias RA, Cristea IM. A Portrait of the Human Organelle Proteome In Space and Time during Cytomegalovirus Infection. Cell Systems. 2016;3(4):361-373.e6. doi:10.1016/j.cels.2016.08.012

Jensen NB, Justesen SD, Larsen A, Ernst E, Pedersen LH. A systematic overview of the spermatotoxic and genotoxic effects of methotrexate, ganciclovir and mycophenolate mofetil. Acta Obstetricia et Gynecologica Scandinavica. 2021;100(9):1557-1580. doi:10.1111/aogs.14151

Jih J, Liu YT, Liu W, Zhou ZH. The incredible bulk: Human cytomegalovirus tegument architectures uncovered by AI-empowered cryo-EM. Sci Adv. 2024 Feb 23;10(8):eadj1640. doi: 10.1126/sciadv.adj1640.

Jin BK, Odongo S, Radwanska M, Magez S. NANOBODIES®: A Review of Diagnostic and Therapeutic Applications. *International Journal of Molecular Sciences*. 2023;24(6). doi:10.3390/ijms24065994

Jones CE, Bailey H, Bamford A, et al. Managing challenges in congenital CMV: current thinking. Archives of Disease in Childhood. 2023;108(8):601-607. doi:10.1136/archdischild-2022-323809

Jones TR, Lee SW. An Acidic Cluster of Human Cytomegalovirus UL99 Tegument Protein Is Required for Trafficking and Function. Journal of Virology. 2004;78(3):1488-1502. doi:10.1128/jvi.78.3.1488-1502.2004

Junyao Xiong,Usama Ashraf,Jing Ye,Shengbo Cao. Extracellular Vesicles in Pathogenic Infection, Transmission, and Immunity. Engineering, 2024, 43(12): 228–240 https://doi.org/10.1016/j.eng.2024.06.011

Kalejta RF. Tegument Proteins of Human Cytomegalovirus. Microbiology and Molecular Biology Reviews. 2008;72(2):249-265. doi:10.1128/mmbr.00040-07

Kalluri R. The biology and function of extracellular vesicles in immune response and immunity. Immunity. 2024;57(8):1752-1768. doi:10.1016/j.immuni.2024.07.009

Kamal A, Goldstein LS. Connecting vesicle transport to the cytoskeleton. Curr Opin Cell Biol. 2000 Aug;12(4):503-8. doi: 10.1016/s0955-0674(00)00123-x.

Kaminski H, Fishman JA. The Cell Biology of Cytomegalovirus: Implications for Transplantation. American Journal of Transplantation. 2016;16(8):2254-2269. doi:10.1111/ajt.13791

Kang C. Maribavir: First Approval. Drugs. 2022;82(3):335-340. doi:10.1007/s40265-022-01677-4

Karafin MS, Hillyer CD, Shaz BH. Transfusion of Plasma and Plasma Derivatives. In: Hematology: Basic Principles and Practice. Elsevier; 2017:1744-1758. doi:10.1016/B978-0-323-35762-3.00115-3

Karleuša L, Lučin HM, Zagorac GB, Lučin P. Cytoplasmic virion assembly compartment of betaherpesviruses. Periodicum Biologorum. 2020;121-122(3-4):97-106. doi:10.18054/pb.v121-122i3-4.10757

Kelly BJ, Fraefel C, Cunningham AL, Diefenbach RJ. Functional roles of the tegument proteins of herpes simplex virus type 1. Virus Research. 2009;145(2):173-186. doi:10.1016/j.virusres.2009.07.007

Kenneson A, Cannon MJ. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. Reviews in Medical Virology. 2007;17(4):253-276. doi:10.1002/rmv.535

Kim CU, Lew W, Williams MA, Liu H, Zhang L, Swaminathan S, Bischofberger N, Chen MS, Mendel DB, Tai CY, Laver WG, Stevens RC. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. J Am Chem Soc. 1997 Jan 29;119(4):681-90. doi: 10.1021/ja963036t.

Kleiboeker SB. Prevalence of cytomegalovirus antiviral drug resistance in transplant recipients. Antiviral Res. 2023 Jul;215:105623. doi: 10.1016/j.antiviral.2023.105623.

Klupp BG, Böttcher S, Granzow H, Kopp M, Mettenleiter TC. Complex Formation between the UL16 and UL21 Tegument Proteins of Pseudorabies Virus. Journal of Virology. 2005;79(3):1510-1522. doi:10.1128/jvi.79.3.1510-1522.2005

Kobe B, Ve T, Williams SJ. Fusion-protein-assisted protein crystallization. Acta Crystallographica Section:F Structural Biology Communications. 2015;71:861-869. doi:10.1107/S2053230X15011061

Koes DR, Baumgartner MP, Camacho CJ. Lessons learned in empirical scoring with smina from the CSAR 2011 Benchmarking exercise. J. Chem. Inf. Model. 2013, 53, 8, 1893-1904. https://doi.org/10.1021/ci300604z

Kojima M, Abe S, Furuta T, et al. High-throughput structure determination of an intrinsically disordered protein using cell-free protein crystallization. Proceedings of the National Academy of Sciences of the United States of America. 2024;121(25). doi:10.1073/pnas.2322452121

Kokkonen P, Kokkola T, Suuronen T, Poso A, Jarho E, Lahtela-Kakkonen M. Virtual screening approach of sirtuin inhibitors results in two new scaffolds. European Journal of Pharmaceutical Sciences. 2015;76:27-32. doi:10.1016/j.ejps.2015.04.025

Köppen-Rung P, Dittmer A, Bogner E. Intracellular Distribution of Capsid-Associated pUL77 of Human Cytomegalovirus and Interactions with Packaging Proteins and pUL93. Journal of Virology. 2016;90(13):5876-5885. doi:10.1128/jvi.00351-16

Kosinski J, Feder M, Bujnicki JM. The PD-(D/E)XK superfamily revisited: Identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. BMC Bioinformatics. 2005;6. doi:10.1186/1471-2105-6-172

Krishna BA, Wills MR, Sinclair JH. Advances in the treatment of cytomegalovirus. British Medical Bulletin. 2019;131(1):5-17. doi:10.1093/bmb/ldz031

Krzyzaniak MA, Mach M, Britt WJ. HCMV-encoded glycoprotein M (UL100) interacts with rab11 effector protein FIP4. Traffic. 2009;10(10):1439-1457. doi:10.1111/j.1600-0854.2009.00967.x

Kschonsak M, Rougé L, Arthur CP, Hoangdung H, Patel N, Kim I, Johnson MC, Kraft E, Rohou AL, Gill A, Martinez-Martin N, Payandeh J, Ciferri C. Structures of HCMV Trimer reveal the basis for receptor recognition and cell entry. Cell. 2021 Mar 4;184(5):1232-1244.e16. doi: 10.1016/j.cell.2021.01.036.

Kuchta AL, Parikh H, Zhu Y, Kellogg GE, Parris DS, Mcvoy MA. Structural modelling and mutagenesis of human cytomegalovirus alkaline nuclease UL98. Journal of General Virology. 2012;93(1):130-138. doi:10.1099/vir.0.034876-0

Kumar R, Cruz L, Sandhu PK, Buchkovich NJ. UL88 Mediates the Incorporation of a Subset of Proteins into the Virion Tegument. Journal of Virology. 2020;94(14). doi:10.1128/jvi.00474-20

Lachmann R, Loenenbach A, Waterboer T, et al. Cytomegalovirus (CMV) seroprevalence in the adult population of Germany. PLoS ONE. 2018;13(7). doi:10.1371/journal.pone.0200267

Lahijani RS, Otteson EW, St Jeor SC. A possible role for nonsense suppression in the synthesis of a human cytomegalovirus 58-kDa virion protein. Virology. 1992 Jan;186(1):309-12. doi: 10.1016/0042-6822(92)90087-6.

Landini MP, Severi B, Furlini G, Badiali De Giorgi L. Human cytomegalovirus structural components: intracellular and intraviral localization of p28 and p65-69 by immunoelectron microscopy. Virus Res. 1987 Jul;8(1):15-23. doi: 10.1016/0168-1702(87)90036-0.

Lanzieri TM, Dollard SC, Bialek SR, Grosse SD. Systematic review of the birth prevalence of congenital cytomegalovirus infection in developing countries. Int J Infect Dis. 2014 May;22:44-8. doi: 10.1016/j.ijid.2013.12.010.

Lee H, Patschull AOM, Bagnéris C, et al. KSHV SOX mediated host shutoff: The molecular mechanism underlying mRNA transcript processing. Nucleic Acids Research. 2017;45(8):4756-4767. doi:10.1093/nar/gkw1340

Lee JH, Kwon M, Lim WY, et al. YAP inhibits HCMV replication by impairing STING-mediated nuclear transport of the viral genome. PLoS Pathogens. 2022;18(12). doi:10.1371/journal.ppat.1011007

Lee JH, Lee H, Lee SW, Hwang SD, Song JH. Efficacy and Safety According to the Dose of Valganciclovir for Cytomegalovirus Prophylaxis in Transplantation: Network Meta-analysis Using Recent Data. Transplantation Proceedings. 2021;53(6):1945-1950. doi:10.1016/j.transproceed.2021.05.006

Lee SH, Albright ER, Lee JH, Jacobs D, Kalejta RF. Virology: Cellular defense against latent colonization foiled by human cytomegalovirus UL138 protein. Science Advances. 2015;1(10). doi:10.1126/sciadv.1501164

Lehár J, Krueger AS, Avery W, et al. Synergistic drug combinations tend to improve therapeutically relevant selectivity. *Nature Biotechnology*. 2009;27(7):659-666. doi:10.1038/nbt.1549

Lemnitzer F, Raschbichler V, Kolodziejczak D, et al. Mouse cytomegalovirus egress protein pM50 interacts with cellular endophilin-A2. Cellular Microbiology. 2013;15(2):335-351. doi:10.1111/cmi.12080

Li M, Ball CB, Collins G, et al. Human cytomegalovirus IE2 drives transcription initiation from a select subset of late infection viral promoters by host RNA polymerase II. PLoS Pathogens. 2020;16(4). doi:10.1371/journal.ppat.1008402

Li S, Xie Y, Yu C, Zheng C, Xu Z. The battle between host antiviral innate immunity and immune evasion by cytomegalovirus. Cellular and Molecular Life Sciences. 2024;81(1). doi:10.1007/s00018-024-05369-y

Li T, Chen ZJ. The cGAS-cGAMP-STI NG pathway connects DNA damage to inflammation, senescence, and cancer. Journal of Experimental Medicine. 2018;215(5):1287-1299. doi:10.1084/jem.20180139

Li Z, Pang J, Dong L, Yu X. Structural basis for genome packaging, retention, and ejection in human cytomegalovirus. Nature Communications. 2021;12(1). doi:10.1038/s41467-021-24820-3

Li Z, Zhang X, Dong L, et al. CryoEM structure of the tegumented capsid of Epstein-Barr virus. Cell Research. 2020;30(10):873-884. doi:10.1038/s41422-020-0363-0

Liang J, Luo Y, Zhao H. Synthetic biology: Putting synthesis into biology. Wiley Interdisciplinary Reviews: Systems Biology and Medicine. 2011;3(1):7-20. doi:10.1002/wsbm.104

Lieberman-Blum SS, Fung HB, Bandres JC. Maraviroc: a CCR5-receptor antagonist for the treatment of HIV-1 infection. Clin Ther. 2008 Jul;30(7):1228-50. doi: 10.1016/s0149-2918(08)80048-3.

Ligat G, Cazal R, Hantz S, Alain S. The human cytomegalovirus terminase complex as an antiviral target: A close-up view. FEMS Microbiology Reviews. 2018;42(2):137-145. doi:10.1093/femsre/fuy004

Lim, W. Y., Lee, J. H., Choi, Y., & Yoon, K. (2024). Verteporfin is an effective inhibitor of HCMV replication. Virus Research, 350, 199475. https://doi.org/10.1016/j.virusres.2024.199475

Limaye AP, Corey L, Koelle DM, Davis CL, Boeckh M. Emergence of ganciclovir-resistant cytomegalovirus disease among recipients of solid-organ transplants. Lancet. 2000 Aug 19;356(9230):645-9. doi: 10.1016/S0140-6736(00)02607-6.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev. 2001 Mar 1;46(1-3):3-26. doi: 10.1016/s0169-409x(00)00129-0. Lischka P, Zimmermann H. Antiviral strategies to combat cytomegalovirus infections in transplant recipients. Curr Opin Pharmacol. 2008 Oct;8(5):541-8. doi: 10.1016/j.coph.2008.07.002.

Liu MT, Hsu TY, Chen JY, Yang CS. Epstein-Barr virus DNase contains two nuclear localization signals, which are different in sensitivity to the hydrophobic regions. Virology. 1998 Jul 20;247(1):62-73. doi: 10.1006/viro.1998.9228.

Liu Y, Biegalke BJ. The human cytomegalovirus UL35 gene encodes two proteins with different functions. J Virol. 2002 Mar;76(5):2460-8. doi: 10.1128/jvi.76.5.2460-2468.2002.

Liu Y, Cui Z, Zhang Z, et al. The tegument protein UL94 of human cytomegalovirus as a binding partner for tegument protein pp28 identified by intracellular imaging. Virology. 2009;388(1):68-77. doi:10.1016/j.virol.2009.03.007

Liu Y, Zhang Z, Zhao X, Wei H, Deng J, Cui Z, Zhang XE. Human cytomegalovirus UL94 is a nucleocytoplasmic shuttling protein containing two NLSs and one NES. Virus Res. 2012 Jun;166(1-2):31-42. doi: 10.1016/j.virusres.2012.02.023.

Ljungman P, Boeckh M, Hirsch HH, et al. Definitions of cytomegalovirus infection and disease in transplant patients for use in clinical trials. Clinical Infectious Diseases. 2017;64(1):87-91. doi:10.1093/cid/ciw668

Ljungman P, de la Camara R, Robin C, et al. Guidelines for the management of cytomegalovirus infection in patients with haematological malignancies and after stem cell transplantation from the 2017 European Conference on Infections in Leukaemia (ECIL 7). The Lancet Infectious Diseases. 2019;19(8):e260-e272. doi:10.1016/S1473-3099(19)30107-0

Long X, Qiu Y, Zhang Z, Wu M. Insight for immunotherapy of hcmv infection. International Journal of Biological Sciences. 2021;17(11):2899-2911. doi:10.7150/ijbs.58127

Lu H, Zhou Q, He J, et al. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. Signal Transduction and Targeted Therapy. 2020;5(1). doi:10.1038/s41392-020-00315-3

Lu J, Wu T, Zhang B, et al. Types of nuclear localization signals and mechanisms of protein import into the nucleus. Cell Communication and Signaling. 2021;19(1). doi:10.1186/s12964-021-00741-y

Lu S, Shen Q, Zhang J. Allosteric Methods and Their Applications: Facilitating the Discovery of Allosteric Drugs and the Investigation of Allosteric Mechanisms. Accounts of Chemical Research. 2019;52(2):492-500. doi:10.1021/acs.accounts.8b00570

Lučin P, Mahmutefendić Lučin H, Blagojević Zagorac G. Cytomegaloviruses reorganize endomembrane system to intersect endosomal and amphisome-like egress pathway. Frontiers in Cell and Developmental Biology. 2023;11. doi:10.3389/fcell.2023.1328751

Lurain NS, Chou S. Antiviral drug resistance of human cytomegalovirus. Clinical Microbiology Reviews. 2010;23(4):689-712. doi:10.1128/CMR.00009-10

Lyman MG, Enquist LW. Herpesvirus Interactions with the Host Cytoskeleton. Journal of Virology. 2009;83(5):2058-2066. doi:10.1128/jvi.01718-08

Lyon SM, Kalejta RF. HCMV Assembly Is Totally Tubular. Dev Cell. 2018 Apr 9;45(1):1-2. doi: 10.1016/j.devcel.2018.03.014.

Lyon SM, Yetming KD, Paulus C, Nevels M, Kalejta RF. Human Cytomegalovirus Genomes Survive Mitosis via the IE19 Chromatin-Tethering Domain. Published online 2020. doi:10.1128/mBio

Macalino SJY, Basith S, Clavio NAB, Chang H, Kang S, Choi S. Evolution of in silico strategies for protein-protein interaction drug discovery. Molecules. 2018;23(8). doi:10.3390/molecules23081963

Mahmutefendić Lučin H, Blagojević Zagorac G, Marcelić M, Lučin P. Host Cell Signatures of the Envelopment Site within Beta-Herpes Virions. International Journal of Molecular Sciences. 2022;23(17). doi:10.3390/ijms23179994

Maita N. Crystal Structure Determination of Ubiquitin by Fusion to a Protein That Forms a Highly Porous Crystal Lattice. Journal of the American Chemical Society. 2018;140(42):13546-13549. doi:10.1021/jacs.8b07512

Manandhar T, Hò GGT, Pump WC, Blasczyk R, Bade-Doeding C. Battle between host immune cellular responses and hcmv immune evasion. International Journal of Molecular Sciences. 2019;20(15). doi:10.3390/ijms20153626

Manicklal S, Emery VC, Lazzarotto T, Boppana SB, Gupta RK. The “Silent” global burden of congenital cytomegalovirus. Clinical Microbiology Reviews. 2013;26(1):86-102. doi:10.1128/CMR.00062-12

Manska S, Rossetto CC. Characteristics of Immediate-Early 2 (IE2) and UL84 Proteins in UL84-Independent Strains of Human Cytomegalovirus (HCMV). Microbiology Spectrum. 2021;9(2). doi:10.1128/spectrum.00539-21

Manska S, Rossetto CC. Identification of cellular proteins associated with human cytomegalovirus (HCMV) DNA replication suggests novel cellular and viral interactions. Virology. 2022;566:26-41. doi:10.1016/j.virol.2021.11.004

Marimani M, Ahmad A, Duse A. Combination therapy as an effective tool for treatment of drug-resistant viral infections. In: *Combination Therapy against Multidrug Resistance*. Elsevier; 2020:157-182. doi:10.1016/B978-0-12-820576-1.00009-6

Marques M, Ferreira AR, Ribeiro D. The interplay between human cytomegalovirus and pathogen recognition receptor signaling. Viruses. 2018;10(10). doi:10.3390/v10100514

Marschall M, Häge S, Conrad M, et al. Nuclear Egress Complexes of HCMV and Other Herpesviruses: Solving the Puzzle of Sequence Coevolution, Conserved Structures and Subfamily-Spanning Binding Properties. Viruses. 2020;12(6). doi:10.3390/v12060683

Marschall M, Marzi A, Aus Dem Siepen P, et al. Cellular p32 recruits cytomegalovirus kinase pUL97 to redistribute the nuclear lamina. Journal of Biological Chemistry. 2005;280(39):33357-33367. doi:10.1074/jbc.M502672200

Marschall M, Schütz M, Wild M, et al. Understanding the Cytomegalovirus Cyclin-Dependent Kinase Ortholog pUL97 as a Multifaceted Regulator and an Antiviral Drug Target. Cells. 2024;13(16). doi:10.3390/cells13161338

Martí-Carreras J, Maes P. Human cytomegalovirus genomics and transcriptomics through the lens of next-generation sequencing: revision and future challenges. Virus Genes. 2019;55(2):138-164. doi:10.1007/s11262-018-1627-3

Martinez R, Sarisky RT, Weber PC, Weller SK. Herpes simplex virus type 1 alkaline nuclease is required for efficient processing of viral DNA replication intermediates. J Virol. 1996 Apr;70(4):2075-85. doi: 10.1128/JVI.70.4.2075-2085.1996.

Martinez R, Shao L, Bronstein JC, Weber PC, Weller SK. The product of a 1.9-kb mRNA which overlaps the HSV-1 alkaline nuclease gene (UL12) cannot relieve the growth defects of a null mutant. Virology. 1996 Jan 15;215(2):152-64. doi: 10.1006/viro.1996.0018.

Martínez-Menárguez JA. Intra-Golgi Transport: Roles for Vesicles, Tubules, and Cisternae. ISRN Cell Biology. 2013;2013:1-15. doi:10.1155/2013/126731

Maschkowitz G, Gärtner S, Hofmann-Winkler H, Fickenscher H, Winkler M. Interaction of Human Cytomegalovirus Tegument Proteins ppUL35 and ppUL35A with Sorting Nexin 5 Regulates Glycoprotein B (gpUL55) Localization. Journal of Virology. 2018;92(9). doi:10.1128/jvi.00013-18

Matthews SM, Groves IJ, O'Connor CM. Chromatin control of human cytomegalovirus infection. mBio. 2023;14(4):e0032623. doi:10.1128/mbio.00326-23

McPherson A, Gavira JA. Introduction to protein crystallization. Acta Crystallographica Section F:Structural Biology Communications. 2014;70(1):2-20. doi:10.1107/S2053230X13033141

Mehravaran H, Makvandi M, Zade AS, et al. Association of human cytomegalovirus with Hodgkin's disease and non-Hodgkin's lymphomas. Asian Pacific Journal of Cancer Prevention. 2017;18(3):593-597. doi:10.22034/APJCP.2017.18.3.593

Meissner CS, Köppen-Rung P, Dittmer A, Lapp S, Bogner E. A "coiled-coil" motif is important for oligomerization and DNA binding properties of human cytomegalovirus protein UL77. PLoS ONE. 2011;6(10). doi:10.1371/journal.pone.0025115

Meng B, Lever AML. The interplay between escrt and viral factors in the enveloped virus life cycle. Viruses. 2021;13(2). doi:10.3390/v13020324

Mercorelli B, Lembo D, Palù G, Loregian A. Early inhibitors of human cytomegalovirus: State-of-art and therapeutic perspectives. Pharmacology and Therapeutics. 2011;131(3):309-329. doi:10.1016/j.pharmthera.2011.04.007

Mészáros B, Erdös G, Dosztányi Z. IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Research. 2018;46(W1):W329-W337. doi:10.1093/nar/gky384

Metrick CM, Chadha P, Heldwein EE. The Unusual Fold of Herpes Simplex Virus 1 UL21, a Multifunctional Tegument Protein. Journal of Virology. 2015;89(5):2979-2984. doi:10.1128/jvi.03516-14

Metrick CM, Koenigsberg AL, Heldwein EE. Conserved Outer Tegument Component UL11 from Herpes Simplex Virus 1 Is an Intrinsically Disordered, RNA-Binding Protein. mBio. 2020 May 5;11(3):e00810-20. doi: 10.1128/mBio.00810-20.

Mettenleiter TC, Klupp BG, Granzow H. Herpesvirus assembly: An update. Virus Research. 2009;143(2):222-234. doi:10.1016/j.virusres.2009.03.018

Michelson S, Turowski P, Picard L, Goris J, Landini MP, Topilko A, Hemmings B, Bessia C, Garcia A, Virelizier JL. Human cytomegalovirus carries serine/threonine protein phosphatases PP1 and a host-cell derived PP2A. J Virol. 1996 Mar;70(3):1415-23. doi: 10.1128/JVI.70.3.1415-1423.1996.

Milbradt J, Auerochs S, Sevvana M, Muller YA, Sticht H, Marschall M. Specific residues of a conserved domain in the N terminus of the human cytomegalovirus pUL50 protein determine its intranuclear interaction with pUL53. Journal of Biological Chemistry. 2012;287(28):24004-24016. doi:10.1074/jbc.M111.331207

Milbradt J, Sonntag E, Wagner S, et al. Human cytomegalovirus nuclear capsids associate with the core nuclear egress complex and the viral protein kinase pUL97. Viruses. 2018;10(1). doi:10.3390/v10010035

Modrzejewska M, Połubiński P, Zdanowska O. Ophthalmic Complications, Diagnosis, and Treatment of Congenital Human Cytomegalovirus Infection. Journal of Clinical Medicine. 2024;13(12). doi:10.3390/jcm13123379

Moffat JG, Vincent F, Lee JA, Eder J, Prunotto M. Opportunities and challenges in phenotypic drug discovery: An industry perspective. Nature Reviews Drug Discovery. 2017;16(8):531-543. doi:10.1038/nrd.2017.111

Monti CE, Mokry RL, Schumacher ML, Dash RK, Terhune SS, by Thomas Shenk E. Computational modeling of protracted HCMV replication using genome substrates and protein temporal profiles. Published online 2024. doi:10.1073/pnas

Moreira IS, Fernandes PA, Ramos MJ. Hot spots - A review of the protein-protein interface determinant amino-acid residues. Proteins: Structure, Function and Genetics. 2007;68(4):803-812. doi:10.1002/prot.21396

Morris CJ, Stern JA, Stark B, Christopherson M, Corte DD. MILCDock: Machine Learning Enhanced Consensus Docking for Virtual Screening in Drug Discovery. J. chem. Inf. Model. 2022, 62, 22, 5342-5350. https://doi.org/10.1021/acs.jcim.2c00705

Morris GM, Ruth H, Lindstrom W, et al. Software news and updates AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*. 2009;30(16):2785-2791. doi:10.1002/jcc.21256

Mosher BS, Kowalik TF, Yurochko AD. Overview of how HCMV manipulation of host cell intracellular trafficking networks can promote productive infection. Frontiers in Virology. 2022;2. doi:10.3389/fviro.2022.1026452

Moumné L, Marie AC, Crouvezier N. Oligonucleotide Therapeutics: From Discovery and Development to Patentability. Pharmaceutics. 2022;14(2). doi:10.3390/pharmaceutics14020260

Mozzi A, Biolatti M, Cagliani R, et al. Past and ongoing adaptation of human cytomegalovirus to its host. PLoS Pathogens. 2020;16(5). doi:10.1371/journal.ppat.1008476

Mueller NH, Gilden D, Cohrs RJ. Varicella-Zoster Virus Open Reading Frame 48 Encodes an Active Nuclease. Journal of Virology. 2013;87(21):11936-11938. doi:10.1128/jvi.01879-13

Muller C, Alain S, Baumert TF, Ligat G, Hantz S. Structures and divergent mechanisms in capsid maturation and stabilization following genome packaging of human cytomegalovirus and herpesviruses. Life. 2021;11(2):1-15. doi:10.3390/life11020150

Muller C, Alain S, Gourin C, Baumert TF, Ligat G, Hantz S. New insights into human cytomegalovirus pul52 structure. Viruses. 2021;13(8). doi:10.3390/v13081638

Munger J, Yu D, Shenk T. UL26-Deficient Human Cytomegalovirus Produces Virions with Hypophosphorylated pp28 Tegument Protein That Is Unstable within Newly Infected Cells. Journal of Virology. 2006;80(7):3541-3548. doi:10.1128/jvi.80.7.3541-3548.2006

Murrell I, Wilkie GS, Davison AJ, et al. Genetic Stability of Bacterial Artificial Chromosome-Derived Human Cytomegalovirus during Culture In Vitro . Journal of Virology. 2016;90(8):3929-3943. doi:10.1128/jvi.02858-15

Mylonakis E, Kallas WM, Fishman JA. *Combination Antiviral Therapy for Ganciclovir-Resistant Cytomegalovirus Infection in Solid-Organ Transplant Recipients*. Vol 34.; 2002. https://academic.oup.com/cid/article/34/10/1337/327283

Naniima P, Naimo E, Koch S, et al. Assembly of infectious Kaposi's sarcoma-associated herpesvirus progeny requires formation of a pORF19 pentamer. PLoS Biology. 2021;19(11). doi:10.1371/journal.pbio.3001423

Necci M, Piovesan D, Hoque MT, et al. Critical assessment of protein intrinsic disorder prediction. Nature Methods. 2021;18(5):472-481. doi:10.1038/s41592-021-01117-3

Neuber S, Wagner K, Goldner T, et al. Mutual Interplay between the Human Cytomegalovirus Terminase Subunits pUL51, pUL56, and pUL89 Promotes Terminase Complex Formation. Journal of Virology. 2017;91(12). doi:10.1128/jvi.02384-16

Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. Journal of Natural Products. 2012;75(3):311-335. doi:10.1021/np200906s

Nguyen CC, Kamil JP. Pathogen at the gates: Human cytomegalovirus entry and cell tropism. Viruses. 2018;10(12). doi:10.3390/v10120704

Nosaki S, Terada T, Nakamura A, et al. Highlighting the potential utility of MBP crystallization chaperone for Arabidopsis BIL1/BZR1 transcription factor-DNA complex. Scientific Reports. 2021;11(1). doi:10.1038/s41598-021-83532-2

O'connor CM, Sen GC. Innate immune responses to herpesvirus infection. Cells. 2021;10(8). doi:10.3390/cells10082122

Ochoa R, Palacio-Rodriguez K, Clemente CM, Adler NS. dockECR: Open consensus docking and ranking protocol for virtual screening of small molecules. *Journal of Molecular Graphics and Modelling*. 2021;109. doi:10.1016/j.jmgm.2021.108023

Ogawa-Goto K, Tanaka K, Gibson W, et al. Microtubule Network Facilitates Nuclear Targeting of Human Cytomegalovirus Capsid. Journal of Virology. 2003;77(15):8541-8547. doi:10.1128/jvi.77.15.8541-8547.2003

Oka T, Yoshimura N. Immunosuppression in organ transplantation. Jpn J Pharmacol. 1996 Jun;71(2):89-100. doi: 10.1254/jjp.71.89.

Ondetti MA, Rubin B, Cushman DW. Design of specific inhibitors of angiotensin-converting enzyme: new class of orally active antihypertensive agents. Science. 1977 Apr 22;196(4288):441-4. doi: 10.1126/science.191908.

Orlowski J, Bujnicki JM. Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. Nucleic Acids Research. 2008;36(11):3552-3569. doi:10.1093/nar/gkn175

Ortiz DA, Glassbrook JE, Pellett PE. Protein-Protein Interactions Suggest Novel Activities of Human Cytomegalovirus Tegument Protein pUL103. Journal of Virology. 2016;90(17):7798-7810. doi:10.1128/jvi.00097-16

Ota R, Hirata A. Relationship between renal dysfunction and electrolyte abnormalities in hematopoietic stem cell transplant patients treated with foscarnet. Journal of Chemotherapy. 2021;33(8):539-546. doi:10.1080/1120009X.2021.1915074

Owen DJ, Crump CM, Graham SC. Tegument assembly and secondary envelopment of alphaherpesviruses. Viruses. 2015;7(9):5084-5114. doi:10.3390/v7092861

Panesso M, Uría ML, Renedo B, et al. CMV hyperimmune globulin as salvage therapy for recurrent or refractory CMV infection in children undergoing hematopoietic stem cell transplantation. Frontiers in Pediatrics. 2023;11. doi:10.3389/fped.2023.1197828

Parascandola J, Jasensky R. Origins of the receptor theory of drug action. Bull Hist Med. 1974 Summer;48(2):199-220.

Pardamean CI, Wu TT. Inhibition of Host Gene Expression by KSHV: Sabotaging mRNA Stability and Nuclear Export. Frontiers in Cellular and Infection Microbiology. 2021;11. doi:10.3389/fcimb.2021.648055

Pardon E, Laeremans T, Triest S, et al. A general protocol for the generation of Nanobodies for structural biology. Nature Protocols. 2014;9(3):674-693. doi:10.1038/nprot.2014.039

Pastano R, Dell'agnola C, Bason C, et al. Antibodies against human cytomegalovirus late protein UL94 in the pathogenesis of scleroderma-like skin lesions in chronic graft-versus-host disease. International Immunology. 2012;24(9):583-591. doi:10.1093/intimm/dxs061

Peng SM, Zhou Y, Huang N. Improving the accuracy of pose prediction in molecular docking via structural filtering and conformational clustering. Chinese Chemical Letters. 2013;24(11):1001-1004. doi:10.1016/j.cclet.2013.06.016

Perchetti GA, Biernacki MA, Xie H, et al. Cytomegalovirus breakthrough and resistance during letermovir prophylaxis. Bone Marrow Transplantation. 2023;58(4):430-436. doi:10.1038/s41409-023-01920-w

Perera MR, Wills MR, Sinclair JH. Hcmv antivirals and strategies to target the latent reservoir. Viruses. 2021;13(5). doi:10.3390/v13050817

Perruccio K, Menconi M, Galaverna F, et al. Safety and efficacy of brincidofovir for Adenovirus infection in children receiving allogeneic stem cell transplantation: an AIEOP retrospective analyses. Bone Marrow Transplantation. 2021;56(12):3104-3107. doi:10.1038/s41409-021-01285-y

Pesch MH, Schleiss MR. Emerging Concepts in Congenital Cytomegalovirus. Pediatrics. 2022;150(2). doi:10.1542/peds.2021-055896

Phillips SL, Bresnahan WA. Identification of Binary Interactions between Human Cytomegalovirus Virion Proteins. Journal of Virology. 2011;85(1):440-447. doi:10.1128/jvi.01551-10

Phillips SL, Bresnahan WA. The Human Cytomegalovirus (HCMV) Tegument Protein UL94 Is Essential for Secondary Envelopment of HCMV Virions. Journal of Virology. 2012;86(5):2523-2532. doi:10.1128/jvi.06548-11

Phillips SL, Cygnar D, Thomas A, Bresnahan WA. Interaction between the Human Cytomegalovirus Tegument Proteins UL94 and UL99 Is Essential for Virus Replication. Journal of Virology. 2012;86(18):9995-10005. doi:10.1128/jvi.01078-12

Piccirilli G, Gabrielli L, Bonasoni MP, et al. Fetal Brain Damage in Human Fetuses with Congenital Cytomegalovirus Infection: Histological Features and Viral Tropism. Cellular and Molecular Neurobiology. 2023;43(3):1385-1399. doi:10.1007/s10571-022-01258-9

Pinninti S, Boppana S. Antiviral Treatment of Maternal and Congenital Cytomegalovirus (CMV) Infections. Viruses. 2023;15(10). doi:10.3390/v15102116

Piovesan D, del Conte A, Mehdiabadi M, et al. MOBIDB in 2025: integrating ensemble properties and function annotations for intrinsically disordered proteins. Nucleic acids research. Published online October 29, 2024. doi:10.1093/nar/gkae969

Pontes KFM, Nardozza LMM, Peixoto AB, et al. Cytomegalovirus and Pregnancy: A Narrative Review. Journal of Clinical Medicine. 2024;13(2). doi:10.3390/jcm13020640

Pracher L, Zeitlinger M. Preclinical and clinical studies in the drug development process of European Medicines Agency-approved non-HIV antiviral agents: a narrative review. Clin Microbiol Infect. 2024 Oct 9:S1198-743X(24)00480-4. doi: 10.1016/j.cmi.2024.10.001.

Procter DJ, Banerjee A, Nukui M, et al. The HCMV Assembly Compartment Is a Dynamic Golgi-Derived MTOC that Controls Nuclear Rotation and Virus Spread. Developmental Cell. 2018;45(1):83-100.e7. doi:10.1016/j.devcel.2018.03.010

Purvinsh L, Gorshkov A, Brodskaia A, Vasin A. Extracellular vesicles in viral pathogenesis: A case of Dr. Jekyll and Mr. Hyde. Life. 2021;11(1):1-19. doi:10.3390/life11010045

Radtke K, Kieneke D, Wolfstein A, et al. Plus- and minus-end directed microtubule motors bind simultaneously to herpes simplex virus capsids using different inner tegument structures. PLoS Pathogens. 2010;6(7):1-20. doi:10.1371/journal.ppat.1000991

Rahbari M, Rahlfs S, Jortzik E, Bogeski I, Becker K. H2O2 dynamics in the malaria parasite Plasmodium falciparum. PLoS ONE. 2017;12(4). doi:10.1371/journal

Razonable RR, Humar A. Cytomegalovirus in solid organ transplant recipients-Guidelines of the American Society of Transplantation Infectious Diseases Community of Practice. Clin Transplant. 2019 Sep;33(9):e13512. doi: 10.1111/ctr.13512.

Razonable RR, van Cruijsen H, Brown RA, et al. Dynamics of Cytomegalovirus Replication during Preemptive Therapy with Oral Ganciclovir. Vol 187.; 2003. https://academic.oup.com/jid/article/187/11/1801/886693

RDKit: Open-source cheminformatics. https://www.rdkit.org

Read C, Schauflinger M, Nikolaenko D, Walther P, von Einem J. Regulation of Human Cytomegalovirus Secondary Envelopment by a C-Terminal Tetralysine Motif in pUL71. Journal of Virology. 2019;93(13). doi:10.1128/jvi.02244-18

Reiss-Gindi N, Hoffman T, Ruderman T, Atamna A, Margalit I, Yahav D. Prophylactic vs preemptive strategy for the prevention of CMV disease in solid organ transplant recipients: systematic review and meta-analysis of randomized controlled trials. Infection. Published online 2024. doi:10.1007/s15010-024-02441-4

Ren Y, Wang A, Zhang B, Ji W, Zhu XX, Lou J, Huang M, Qiu Y, Zhou X. Human cytomegalovirus UL36 inhibits IRF3-dependent immune signaling to counterbalance its immunoenhancement as apoptotic inhibitor. Sci Adv. 2023 Oct 6;9(40):eadi6586. doi: 10.1126/sciadv.adi6586.

Riley HD Jr. History of the cytomegalovirus. South Med J. 1997 Feb;90(2):184-90. doi: 10.1097/00007611-199702000-00004.

Rondeau JM, Schreuder H. Protein Crystallography and Drug Discovery. In: The Practice of Medicinal Chemistry: Fourth Edition. Elsevier Inc.; 2015:511-537. doi:10.1016/B978-0-12-417205-0.00022-5

Rozman B, Nachshon A, Levi Samia R, Lavi M, Schwartz M, Stern-Ginossar N. Temporal dynamics of HCMV gene expression in lytic and latent infections. Cell Reports. 2022;39(2). doi:10.1016/j.celrep.2022.110653

Rustandi RR, Loughney JW, Shang L, Wang S, Pauley CJ, Christanti S, Kristopeit A, Culp TD. Characterization of gH/gL/pUL128-131 pentameric complex, gH/gL/gO trimeric complex, gB and gM/gN glycoproteins in a human cytomegalovirus using automated capillary western blots. Vaccine. 2021 Jul 30;39(33):4705-4715. doi: 10.1016/j.vaccine.2021.06.033.

Rütten H, Rissmann A, Brett B, et al. Congenital cytomegalovirus infection in Central Germany: an underestimated risk. Archives of Gynecology and Obstetrics. 2017;296(2):231-240. doi:10.1007/s00404-017-4435-4

Salvant BS, Fortunato EA, Spector DH. Cell cycle dysregulation by human cytomegalovirus: influence of the cell cycle phase at the time of infection and effects on cyclin transcription. J Virol. 1998 May;72(5):3729-41. doi: 10.1128/JVI.72.5.3729-3741.1998.

Sampaio KL, Cavignac Y, Stierhof YD, Sinzger C. Human Cytomegalovirus Labeled with Green Fluorescent Protein for Live Analysis of Intracellular Particle Movements. Journal of Virology. 2005;79(5):2754-2767. doi:10.1128/jvi.79.5.2754-2767.2005

Samuel E, McNaught KA, Mulbah JL, HajiAlilou H, Mody V, Cates DW. Antiviral drugs. Side Effects of Drugs Annual. 2022;44:291–301. doi: 10.1016/bs.seda.2022.09.006.

Sanchez V, Britt W. Human cytomegalovirus egress: Overcoming barriers and co-opting cellular functions. Viruses. 2022;14(1). doi:10.3390/v14010015

Sanchez V, Greis KD, Sztul E, Britt WJ. Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. J Virol. 2000 Jan;74(2):975-86. doi: 10.1128/jvi.74.2.975-986.2000.

Sanchez V, Sztul E, Britt WJ. Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment. J Virol. 2000 Apr;74(8):3842-51. doi: 10.1128/jvi.74.8.3842-3851.2000.

Sausen DG, Reed KM, Bhutta MS, Gallo ES, Borenstein R. Evasion of the host immune response by betaherpesviruses. International Journal of Molecular Sciences. 2021;22(14). doi:10.3390/ijms22147503

Schauflinger M, Villinger C, Mertens T, Walther P, von Einem J. Analysis of human cytomegalovirus secondary envelopment by advanced electron microscopy. Cellular Microbiology. 2013;15(2):305-314. doi:10.1111/cmi.12077

Schilling EM, Scherer M, Stamminger T. Intrinsic immune mechanisms restricting human cytomegalovirus replication. Viruses. 2021;13(2). doi:10.3390/v13020179

Schmolke S, Kern HF, Drescher P, Jahn G, Plachter B. The dominant phosphoprotein pp65 (UL83) of human cytomegalovirus is dispensable for growth in cell culture. J Virol. 1995 Oct;69(10):5959-68. doi: 10.1128/JVI.69.10.5959-5968.1995.

Scott GM, Barrell BG, Oram J, Rawlinson WD. Characterisation of transcripts from the human cytomegalovirus genes TRL7, UL20a, UL36, UL65, UL94, US3 and US34. Virus Genes. 2002;24(1):39-48. doi: 10.1023/a:1014033920070.

Seitz R. Human Cytomegalovirus (HCMV)-Revised. Transfusion Medicine and Hemotherapy. 2010;37(6):365-375. doi:10.1159/000322141

Seo JY, Britt WJ. Cytoplasmic Envelopment of Human Cytomegalovirus Requires the Postlocalization Function of Tegument Protein pp28 within the Assembly Compartment. Journal of Virology. 2007;81(12):6536-6547. doi:10.1128/jvi.02852-06

Seo JY, Britt WJ. Sequence Requirements for Localization of Human Cytomegalovirus Tegument Protein pp28 to the Virus Assembly Compartment and for Assembly of Infectious Virus. Journal of Virology. 2006;80(11):5611-5626. doi:10.1128/jvi.02630-05

Seo JY, Heo JA, Britt WJ. Phosphorylation of tegument protein pp28 contributes to trafficking to the assembly compartment in human cytomegalovirus infection. Journal of Microbiology. 2020;58(7):624-631. doi:10.1007/s12275-020-0263-5

Severi B, Landini MP, Cenacchi G, Zini N, Maraldi NM. Human cytomegalovirus nuclear and cytoplasmic dense bodies. Arch Virol. 1992;123(1-2):193-207. doi: 10.1007/BF01317149.

Shahar-Nissan K, Pardo J, Peled O, et al. Valaciclovir to prevent vertical transmission of cytomegalovirus after maternal primary infection during pregnancy: a randomised, double-blind, placebo-controlled trial. The Lancet. 2020;396(10253):779-785. doi:10.1016/S0140-6736(20)31868-7

Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: A novel approach for cancer therapy. Annual Review of Pharmacology and Toxicology. 2009;49:223-241. doi:10.1146/annurev.pharmtox.48.113006.094723

Shao Q, Liu T, Wang W, Liu T, Jin X, Chen Z. Promising Role of Emodin as Therapeutics to Against Viral Infections. Frontiers in Pharmacology. 2022;13. doi:10.3389/fphar.2022.902626

Sharma M, Bender BJ, Kamil JP, et al. Human Cytomegalovirus UL97 Phosphorylates the Viral Nuclear Egress Complex. Journal of Virology. 2015;89(1):523-534. doi:10.1128/jvi.02426-14

Sharma M, Kamil JP, Coen DM. Preparation of the Human Cytomegalovirus Nuclear Egress Complex and Associated Proteins. In: Methods in Enzymology. Vol 569. Academic Press Inc.; 2016:517-526. doi:10.1016/bs.mie.2015.08.020

Sharma V, Mobeen F, Prakash T. Comparative Genomics of Herpesviridae Family to Look for Potential Signatures of Human Infecting Strains. Int J Genomics. 2016:9543274. doi: 10.1155/2016/9543274.

Sheaffer AK, Weinheimer SP, Tenney DJ. The human cytomegalovirus UL98 gene encodes the conserved herpesvirus alkaline nuclease. J Gen Virol. 1997 Nov;78 (Pt 11):2953-61. doi: 10.1099/0022-1317-78-11-2953.

Shehata SI, Watkins JM, Burke JM, Parker R. Mechanisms and consequences of mRNA destabilization during viral infections. Virology Journal. 2024;21(1). doi:10.1186/s12985-024-02305-1

Shenk T, Alwine JC. Human cytomegalovirus: Coordinating cellular stress, signaling, and metabolic pathways. Annual Review of Virology. 2014;1(1):355-374. doi:10.1146/annurev-virology-031413-085425

Sherry MR, Hay TJM, Gulak MA, Nassiri A, Finnen RL, Banfield BW. The Herpesvirus Nuclear Egress Complex Component, UL31, Can Be Recruited to Sites of DNA Damage Through Poly-ADP Ribose Binding. Scientific Reports. 2017;7(1). doi:10.1038/s41598-017-02109-0

Shim GH. Treatment of congenital cytomegalovirus infection. Clinical and Experimental Pediatrics. 2023;66(9):384-394. doi:10.3345/cep.2022.01032

Shin WH, Christoffer CW, Kihara D. In silico structure-based approaches to discover protein-protein interaction-targeting drugs. Methods. 2017;131:22-32. doi:10.1016/j.ymeth.2017.08.006

Shintani M, Takahashi G, Hamada M, Okunaga S, Iwai S, Yura Y. Effect of ultrasound on herpes simplex virus infection in cell culture. Virology Journal. 2011;8. doi:10.1186/1743-422X-8-446

Sievers F, Higgins DG. Clustal Omega for making accurate alignments of many protein sequences. Protein Science. 2018;27(1):135-145. doi:10.1002/pro.3290

Silva MC, Yu QC, Enquist L, Shenk T. Human Cytomegalovirus UL99-Encoded pp28 Is Required for the Cytoplasmic Envelopment of Tegument-Associated Capsids. Journal of Virology. 2003;77(19):10594-10605. doi:10.1128/jvi.77.19.10594-10605.2003

Singh G, Gaidhane A. A Review of Sensorineural Hearing Loss in Congenital Cytomegalovirus Infection. Cureus. Published online October 26, 2022. doi:10.7759/cureus.30703

Singh N, Winston DJ, Razonable RR, et al. Effect of Preemptive Therapy vs Antiviral Prophylaxis on Cytomegalovirus Disease in Seronegative Liver Transplant Recipients with Seropositive Donors: A Randomized Clinical Trial. JAMA - Journal of the American Medical Association. 2020;323(14):1378-1387. doi:10.1001/jama.2020.3138

Sledz P, Zheng H, Murzyn K, et al. New surface contacts formed upon reductive lysine methylation: Improving the probability of protein crystallization. *Protein Science*. 2010;19(7):1395-1404. doi:10.1002/pro.420

Smith MC, Gestwicki JE. Features of protein-protein interactions that translate into potent inhibitors: topology, surface area and affinity. Expert reviews in molecular medicine. 2012;14. doi:10.1017/erm.2012.10

Smith NA, Chan GC, O'Connor CM. Modulation of host cell signaling during cytomegalovirus latency and reactivation. Virology Journal. 2021;18(1). doi:10.1186/s12985-021-01674-1

Smith RM, Kosuri S, Kerry JA. Role of human cytomegalovirus tegument proteins in virion assembly. Viruses. 2014;6(2):582-605. doi:10.3390/v6020582

Song B, Sheng X, Justice JL, Lum KK, Metzger PJ, Cook KC, Kostas JC, Cristea IM. Intercellular communication within the virus microenvironment affects the susceptibility of cells to secondary viral infections. Sci Adv. 2023 May 10;9(19):eadg3433. doi: 10.1126/sciadv.adg3433.

Sonntag E, Hamilton ST, Bahsi H, et al. Cytomegalovirus pUL50 is the multi-interacting determinant of the core nuclear egress complex (NEC) that recruits cellular accessory NEC components. Journal of General Virology. 2016;97(7):1676-1685. doi:10.1099/jgv.0.000495

Spector DJ. UL84-independent replication of human cytomegalovirus strains conferred by a single codon change in UL122. Virology. 2015 Feb;476:345-354. doi: 10.1016/j.virol.2014.12.031.

Stephan AJ, de Lepper M, Wölle R, et al. Healthcare costs of congenital cytomegalovirus (cCMV) disease in infants during the first two years of life: a retrospective German claims database analysis. Cost Effectiveness and Resource Allocation. 2023;21(1). doi:10.1186/s12962-022-00411-x

Strang BL, Sinigalia E, Silva LA, Coen DM, Loregian A. Analysis of the Association of the Human Cytomegalovirus DNA Polymerase Subunit UL44 with the Viral DNA Replication Factor UL84. Journal of Virology. 2009;83(15):7581-7589. doi:10.1128/jvi.00663-09

Strang BL, Stow ND. Circularization of the Herpes Simplex Virus Type 1 Genome upon Lytic Infection. Journal of Virology. 2005;79(19):12487-12494. doi:10.1128/jvi.79.19.12487-12494.2005

Stratta RJ, Shaeffer MS, Markin RS, Wood RP, Langnas AN, Reed EC, Donovan JP, Woods GL, Bradshaw KA, Pillen TJ, et al. Cytomegalovirus infection and disease after liver transplantation. An overview. Dig Dis Sci. 1992 May;37(5):673-88. doi: 10.1007/BF01296422.

Streck NT, Zhao Y, Sundstrom JM, Buchkovich NJ. Human Cytomegalovirus Utilizes Extracellular Vesicles To Enhance Virus Spread. J Virol. 2020 Jul 30;94(16):e00609-20. doi: 10.1128/JVI.00609-20.

Stumpf MP, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C. Estimating the size of the human interactome. Proc Natl Acad Sci U S A. 2008 May 13;105(19):6959-64. doi: 10.1073/pnas.0708078105.

Styczyński J. Prophylaxis vs preemptive therapy in prevention of CMV infection: New insight on prophylactic strategy after allogeneic hematopoietic cell transplantation. Acta Haematologica Polonica. 2020;51(1):17-23. doi:10.2478/ahp-2020-0005

Su XD, Zhang H, Terwilliger TC, Liljas A, Xiao J, Dong Y. Protein crystallography from the perspective of technology developments. Crystallography Reviews. 2015;21(1-2):122-153. doi:10.1080/0889311X.2014.973868

Sucharita S, Krishnagopal A, van Drunen Littel-van den Hurk S. Comprehensive Analysis of the Tegument Proteins Involved in Capsid Transport and Virion Morphogenesis of Alpha, Beta and Gamma Herpesviruses. Viruses. 2023;15(10). doi:10.3390/v15102058

Sun J, Zhao Z. A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics. 2010;11(SUPPL. 3). doi:10.1186/1471-2164-11-S3-S5

Swinney DC. Phenotypic vs. Target-based drug discovery for first-in-class medicines. Clinical Pharmacology and Therapeutics. 2013;93(4):299-301. doi:10.1038/clpt.2012.236

Symmers WSC. Opportunistic Infections: The Concept of 'Opportunistic Infections'. Proceedings of the Royal Society of Medicine. 1965;58(5):341-346. doi:10.1177/003591576505800521

Tabana Y, Babu D, Fahlman R, Siraki AG, Barakat K. Target identification of small molecules: an overview of the current applications in drug discovery. BMC Biotechnology. 2023;23(1). doi:10.1186/s12896-023-00815-4

Taisne C, Lussignol M, Hernandez E, Moris A, Mouna L, Esclatine A. Human cytomegalovirus hijacks the autophagic machinery and LC3 homologs in order to optimize cytoplasmic envelopment of mature infectious particles. Scientific Reports. 2019;9(1). doi:10.1038/s41598-019-41029-z

Tan CA, Palen L, Su Y, et al. Impact of Primary Letermovir Prophylaxis Versus Preemptive Antiviral Therapy for Cytomegalovirus on Economic and Clinical Outcomes after Hematopoietic Cell Transplantation. Transplantation and Cellular Therapy. 2024;30(8):792.e1-792.e12. doi:10.1016/j.jtct.2024.05.021

Tandon R, AuCoin DP, Mocarski ES. Human Cytomegalovirus Exploits ESCRT Machinery in the Process of Virion Maturation. Journal of Virology. 2009;83(20):10797-10807. doi:10.1128/jvi.01093-09

Tandon R, Mocarski ES, Conway JF. The A, B, Cs of herpesvirus capsids. Viruses. 2015;7(3):899-914. doi:10.3390/v7030899

Tandon R, Mocarski ES. Control of Cytoplasmic Maturation Events by Cytomegalovirus Tegument Protein pp150. Journal of Virology. 2008;82(19):9433-9444. doi:10.1128/jvi.00533-08

Tandon R, Mocarski ES. Viral and host control of cytomegalovirus maturation. Trends in Microbiology. 2012;20(8):392-401. doi:10.1016/j.tim.2012.04.008

Teklemariam TA, Rivera OD, Nelson SW. Kinetic Analysis of the Exonuclease Activity of the Bacteriophage T4 Mre11–Rad50 Complex. In: Methods in Enzymology. Vol 600. Academic Press Inc.; 2018:135-156. doi:10.1016/bs.mie.2017.12.007

Tessier TM, Dodge MJ, Prusinkiewicz MA, Mymryk JS. Viral Appropriation: Laying Claim to host Nuclear Transport Machinery. Cells. 2019;8(6). doi:10.3390/cells8060559

Thomas ECM, Finnen RL, Mewburn JD, Archer SL, Banfield BW. The Herpes Simplex Virus pUL16 and pUL21 Proteins Prevent Capsids from Docking at Nuclear Pore Complexes. PLoS Pathogens. 2023;19(12). doi:10.1371/journal.ppat.1011832

Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics. 2001 Mar;17(3):284-5. doi: 10.1093/bioinformatics/17.3.284.

Timofeev V, Samygina V. Protein Crystallography: Achievements and Challenges. Crystals. 2023;13(1). doi:10.3390/cryst13010071

To A, Bai Y, Shen A, et al. Yeast two hybrid analyses reveal Novel binary interactions between human cytomegalovirus-encoded virion proteins. PLoS ONE. 2011;6(4). doi:10.1371/journal.pone.0017796

Tomtishen JP 3rd. Human cytomegalovirus tegument proteins (pp65, pp71, pp150, pp28). Virol J. 2012 Jan 17;9:22. doi: 10.1186/1743-422X-9-22.

Turner DL, Korneev D v., Purdy JG, de Marco A, Mathias RA. The host exosome pathway underpins biogenesis of the human cytomegalovirus virion. eLife. 2020;9:1-29. doi:10.7554/ELIFE.58288

Turner DL, Mathias RA. The human cytomegalovirus decathlon: Ten critical replication events provide opportunities for restriction. Frontiers in Cell and Developmental Biology. 2022;10. doi:10.3389/fcell.2022.1053139

Tyl MD, Betsinger CN, Cristea IM. Virus-host protein interactions as footprints of human cytomegalovirus replication. Curr Opin Virol. 2022 Feb;52:135-147. doi: 10.1016/j.coviro.2021.11.016.

Uchański T, Masiulis S, Fischer B, et al. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nature Methods. 2021;18(1):60-68. doi:10.1038/s41592-020-01001-6

Upadhyayula S, Michaels MG. Ganciclovir, foscarnet, and cidofovir: Antiviral drugs not just for cytomegalovirus. Journal of the Pediatric Infectious Diseases Society. 2013;2(3):286-290. doi:10.1093/jpids/pit048

Uppal T, Meyer D, Agrawal A, Verma SC.2019.The DNase Activity of Kaposi's Sarcoma-Associated Herpesvirus SOX Protein Serves an Important Role in Viral Genome Processing during Lytic Replication. J Virol93:10.1128/jvi.01983-18.https://doi.org/10.1128/jvi.01983-18

Urban JD, Clarke WP, von Zastrow M, et al. Functional selectivity and classical concepts of quantitative pharmacology. Journal of Pharmacology and Experimental Therapeutics. 2007;320(1):1-13. doi:10.1124/jpet.106.104463

van Son WJ, The TH. Cytomegalovirus infection after organ transplantation: an update with special emphasis on renal transplantation. Transpl Int. 1989 Oct;2(3):147-64. doi: 10.1007/BF02414602.

Vellani TS, Myers RS. Bacteriophage SPP1 Chu is an alkaline exonuclease in the SynExo family of viral two-component recombinases. Journal of Bacteriology. 2003;185(8):2465-2474. doi:10.1128/JB.185.8.2465-2474.2003

Vescovini R, Telera AR, Pedrazzoni M, et al. Impact of persistent cytomegalovirus infection on dynamic changes in human immune system profile. PLoS ONE. 2016;11(3). doi:10.1371/journal.pone.0151965

Vincent F, Nueda A, Lee J, Schenone M, Prunotto M, Mercola M. Phenotypic drug discovery: recent successes, lessons learned and new directions. Nature Reviews Drug Discovery. 2022;21(12):899-914. doi:10.1038/s41573-022-00472-w

Vo M, Aguiar A, McVoy MA, Hertel L. Cytomegalovirus strain TB40/E restrictions and adaptations to growth in ARPE-19 epithelial cells. Microorganisms. 2020;8(4). doi:10.3390/microorganisms8040615

Vora SB, Brothers AW, Waghmare A, Englund JA. Antiviral combination therapy for cytomegalovirus infection in high-risk infants. *Antiviral Therapy*. 2018;23(6):505-511. doi:10.3851/IMP3238

Walter TS, Meier C, Assenberg R, et al. Lysine Methylation as a Routine Rescue Strategy for Protein Crystallization. *Structure*. 2006;14(11):1617-1622. doi:10.1016/j.str.2006.09.005

Wang B, Dai T, Sun W, et al. Protein N-myristoylation: functions and mechanisms in control of innate immunity. Cellular and Molecular Immunology. 2021;18(4):878-888. doi:10.1038/s41423-021-00663-2

Wang D, Shenk T. Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. Proc Natl Acad Sci U S A. 2005 Dec 13;102(50):18153-8. doi: 10.1073/pnas.0509201102.

Wang D, Yu QC, Schröer J, Murphy E, Shenk T. Human cytomegalovirus uses two distinct pathways to enter retinal pigmented epithelial cells. Proc Natl Acad Sci U S A. 2007 Dec 11;104(50):20037-42. doi: 10.1073/pnas.0709704104.

Wang J ben, McVoy MA. A 128-Base-Pair Sequence Containing the pac1 and a Presumed Cryptic pac2 Sequence Includes cis Elements Sufficient To Mediate Efficient Genome Maturation of Human Cytomegalovirus . Journal of Virology. 2011;85(9):4432-4439. doi:10.1128/jvi.02307-10

Wang N, Clark LD, Gao Y, Kozlov MM, Shemesh T, Rapoport TA. Mechanism of membrane-curvature generation by ER-tubule shaping proteins. Nature Communications. 2021;12(1). doi:10.1038/s41467-020-20625-y

Wang WD, Lee GC, Kim YY, Lee CH. A comparison between low- and high-passage strains of human cytomegalovirus. Journal of Microbiology and Biotechnology. 2016;26(10):1800-1807. doi:10.4014/jmb.1604.04045

Wang Y, Mao L, Kankanala J, Wang Z, Geraghty RJ. Inhibition of Human Cytomegalovirus pUL89 Terminase Subunit Blocks Virus Replication and Genome Cleavage. J Virol. 2017 Jan 18;91(3):e02152-16. doi: 10.1128/JVI.02152-16.

Wang YQ, Zhao XY. Human Cytomegalovirus Primary Infection and Reactivation: Insights From Virion-Carried Molecules. Frontiers in Microbiology. 2020;11. doi:10.3389/fmicb.2020.01511

Wang Z, Geraghty RJ. Viral Nucleases. Viruses. 2023;15(3). doi:10.3390/v15030740

Watanabe S, Radman-Livaja M, Rando OJ, Peterson CL. A histone acetylation switch regulates H2A.Z deposition by the SWR-C remodeling enzyme. Science. 2013;340(6129):195-199. doi:10.1126/science.1229758

Wedemann L, Flomm FJ, Bosse JB. The unconventional way out—Egress of HCMV through multiviral bodies. Molecular Microbiology. 2022;117(6):1317-1323. doi:10.1111/mmi.14946

Weekes MP, Tomasec P, Huttlin EL, Fielding CA, Nusinow D, Stanton RJ, Wang ECY, Aicheler R, Murrell I, Wilkinson GWG, Lehner PJ, Gygi SP. Quantitative temporal viromics: an approach to investigate host-pathogen interaction. Cell. 2014 Jun 5;157(6):1460-1472. doi: 10.1016/j.cell.2014.04.028.

Weiler N, Paal C, Adams K, et al. Role of envelope glycoprotein complexes in cell-associated spread of human cytomegalovirus. Viruses. 2021;13(4). doi:10.3390/v13040614

Weller SK, Seghatoleslami MR, Shao L, Rowse D, Carmichael EP. The herpes simplex virus type 1 alkaline nuclease is not essential for viral DNA synthesis: isolation and characterization of a lacZ insertion mutant. J Gen Virol. 1990 Dec;71 (Pt 12):2941-52. doi: 10.1099/0022-1317-71-12-2941.

Wiebusch L, Hagemeier C. The human cytomegalovirus immediate early 2 protein dissociates cellular DNA synthesis from cyclin-dependent kinase activation. EMBO J. 2001 Mar 1;20(5):1086-98. doi: 10.1093/emboj/20.5.1086.

Wild M, Karner D, Eickhoff J, et al. Combined Treatment with Host-Directed and Anticytomegaloviral Kinase Inhibitors: Mechanisms, Synergisms and Drug Resistance Barriers. Pharmaceutics. 2023;15(12). doi:10.3390/pharmaceutics15122680

Wildy P. Classification and Nomenclature of Viruses: 1st Report of the International committee on Nomenclature of Viruses. S.Karger AG. 1971. https://doi.org/10.1159/isbn.978-3-318-04027-2

Wilkinson GWG, Davison AJ, Tomasec P, et al. Human cytomegalovirus: taking the strain. Medical Microbiology and Immunology. 2015;204(3):273-284. doi:10.1007/s00430-015-0411-4

Wille PT, Wisner TW, Ryckman B, Johnson DC. Human cytomegalovirus (HCMV) glycoprotein gB promotes virus entry in Trans acting as the viral fusion protein rather than as a receptor-binding protein. mBio. 2013;4(3). doi:10.1128/mBio.00332-13

Wing BA, Huang ES. Analysis and mapping of a family of 3'-coterminal transcripts containing coding sequences for human cytomegalovirus open reading frames UL93 through UL99. J Virol. 1995 Mar;69(3):1521-31. doi: 10.1128/JVI.69.3.1521-1531.1995.

Wing BA, Lee GC, Huang ES. The human cytomegalovirus UL94 open reading frame encodes a conserved herpesvirus capsid/tegument-associated virion protein that is expressed with true late kinetics. J Virol. 1996 Jun;70(6):3339-45. doi: 10.1128/JVI.70.6.3339-3345.1996.

Wofford AS, McCusker I, Green JC, Vensko TA, Pellett PE. Betaherpesvirus assembly and egress: Recent advances illuminate the path. In: Advances in Virus Research. Vol 108. Academic Press Inc.; 2020:337-392. doi:10.1016/bs.aivir.2020.09.003

Womack A, Shenk T. Human Cytomegalovirus Tegument Protein pUL71 Is Required for Efficient Virion Egress. Published online 2010. doi:10.1128/mBio.00282

Wright LR, Wright DL, Weller SK. Viral Nucleases from Herpesviruses and Coronavirus in Recombination and Proofreading: Potential Targets for Antiviral Drug Discovery. Viruses. 2022;14(7). doi:10.3390/v14071557

Wu H, Kropff B, Mach M, Britt WJ. Human cytomegalovirus envelope protein gpul132 regulates infectious virus production through formation of the viral assembly compartment. mBio. 2020;11(5):1-16. doi:10.1128/mBio.02044-20

Wu X, Cabanos C, Rapoport TA. Structure of the post-translational protein translocation machinery of the ER membrane. Nature. 2019;566(7742):136-139. doi:10.1038/s41586-018-0856-x

Wu Y, Prager A, Boos S, et al. Human cytomegalovirus glycoprotein complex gH/gL/gO uses PDGFR-α as a key for entry. PLoS Pathogens. 2017;13(4). doi:10.1371/journal.ppat.1006281

Wutzler P, Thust R. Genetic risks of antiviral nucleoside analogues--a survey. Antiviral Res. 2001 Feb;49(2):55-74. doi: 10.1016/s0166-3542(00)00139-x.

Xiaofei E, Meraner P, Lu P, et al. OR14I1 is a receptor for the human cytomegalovirus pentameric complex and defines viral epithelial cell tropism. Proceedings of the National Academy of Sciences of the United States of America. 2019;116(14):7043-7052. doi:10.1073/pnas.1814850116

Xu Y, Cei SA, Rodriguez Huete A, Colletti KS, Pari GS. Human Cytomegalovirus DNA Replication Requires Transcriptional Activation via an IE2- and UL84-Responsive Bidirectional Promoter Element within ori Lyt . Journal of Virology. 2004;78(21):11664-11677. doi:10.1128/jvi.78.21.11664-11677.2004

Y. Hong, H. Jeong, K. Park, S. Lee, J.Y. Shim, H. Kim, Y. Song, S. Park, H.Y. Park, V.N. Kim, K. Ahn, STING facilitates nuclear import of herpesvirus genome during infection, Proc. Natl. Acad. Sci. U.S.A. 118 (33) e2108631118, https://doi.org/10.1073/pnas.2108631118 (2021).

Yadav DK, Adhikari VP, Yadav RK, et al. Antiviral prophylaxis or preemptive therapy for cytomegalovirus after liver transplantation?: A systematic review and meta-analysis. Frontiers in Immunology. 2022;13. doi:10.3389/fimmu.2022.953210

Yang B, Yao Y, Wu H, et al. Localization of the WD Repeat-Containing Protein 5 to the Virion Assembly Compartment Facilitates Human Cytomegalovirus Assembly. Published online 2021. doi:10.1128/JVI

Yang L, Wang M, Cheng A, et al. Features and Functions of the Conserved Herpesvirus Tegument Protein UL11 and Its Binding Partners. Frontiers in Microbiology. 2022;13. doi:10.3389/fmicb.2022.829754

Yang W. Nucleases: Diversity of structure, function and mechanism. Quarterly Reviews of Biophysics. 2011;44(1):1-93. doi:10.1017/S0033583510000181

Ye L, Qian Y, Yu W, Guo G, Wang H, Xue X. Functional Profile of Human Cytomegalovirus Genes and Their Associated Diseases: A Review. Frontiers in Microbiology. 2020;11. doi:10.3389/fmicb.2020.02104

Yu D, Silva MC, Shenk T. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. Proc Natl Acad Sci U S A. 2003 Oct 14;100(21):12396-401. doi: 10.1073/pnas.1635160100.

Yu X, Jih J, Jiang J, Zhou H. Atomic structure of the human cytomegalovirus capsid with its securing tegument layer of pp150. doi:10.1126/science

Yu X, Shah S, Lee M, et al. Biochemical and structural characterization of the capsid-bound tegument proteins of human cytomegalovirus. Journal of Structural Biology. 2011;174(3):451-460. doi:10.1016/j.jsb.2011.03.006

Zeev-Ben-Mordehai T, Weberruß M, Lorenz M, et al. Crystal structure of the herpesvirus nuclear egress complex provides insights into inner nuclear membrane remodeling. Cell Reports. 2015;13(12):2645-2652. doi:10.1016/j.celrep.2015.11.008

Zeng J, Cao D, Yang S, et al. Insights into the Transcriptome of Human Cytomegalovirus: A Comprehensive Review. Viruses. 2023;15(8). doi:10.3390/v15081703

Zeng Y, Middeldorp J, Madjar JJ, Ooka T. A major DNA binding protein encoded by BALF2 open reading frame of Epstein-Barr virus (EBV) forms a complex with other EBV DNA-binding proteins: DNAase, EA-D, and DNA polymerase. Virology. 1997 Dec 22;239(2):285-95. doi: 10.1006/viro.1997.8891.

Zhang L, Yu J, Liu Z. MicroRNAs expressed by human cytomegalovirus. Virology Journal. 2020;17(1). doi:10.1186/s12985-020-1296-4

Zhang T, Potgieter TI, Kosche E, Rückert J, Ostermann E, Schulz T, Empting M, Brune W. Thioxothiazolo[3,4-a]quinazoline derivatives inhibit the human cytomegalovirus alkaline nuclease. Antiviral Res. 2023 Sep;217:105696. doi: 10.1016/j.antiviral.2023.105696.

Zhang Z, Shigematsu H, Shimizu T, Ohto U. Improving particle quality in cryo-EM analysis using a PEGylation method. Structure. 2021 Oct 7;29(10):1192-1199.e4. doi: 10.1016/j.str.2021.05.004.

Zhao Y, Aguilar A, Bernard D, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction (MDM2 inhibitors) in clinical trials for cancer treatment. Journal of Medicinal Chemistry. 2015;58(3):1038-1052. doi:10.1021/jm501092z

Zhong B, Yang Y, Li S, Wang YY, Li Y, Diao F, Lei C, He X, Zhang L, Tien P, Shu HB. The adaptor protein MITA links virus-sensing receptors to IRF3 transcription factor activation. Immunity. 2008 Oct 17;29(4):538-50. doi: 10.1016/j.immuni.2008.09.003.

Zhou S, Fu Z, Zhang Z, et al. Liquid–liquid phase separation mediates the formation of herpesvirus assembly compartments. Journal of Cell Biology. 2023;222(1). doi:10.1083/jcb.202201088

Zhu J, Chiang C, Gack MU. Viral evasion of the interferon response at a glance. Journal of Cell Science. 2023;136(12). doi:10.1242/jcs.260682

Zicari S, Arakelyan A, Palomino RAÑ, Fitzgerald W, Vanpouille C, Lebedeva A, Schmitt A, Bomsel M, Britt W, Margolis L. Human cytomegalovirus-infected cells release extracellular vesicles that carry viral surface proteins. Virology. 2018 Nov;524:97-105. doi: 10.1016/j.virol.2018.08.008.

Zimmermann C, Krämer N, Krauter S, et al. Autophagy interferes with human cytomegalovirus genome replication, morphogenesis, and progeny release. Autophagy. 2021;17(3):779-795. doi:10.1080/15548627.2020.1732686

Zimmermann I, Egloff P, Hutter CAJ, et al. Generation of synthetic nanobodies against delicate proteins. Nature Protocols. 2020;15(5):1707-1741. doi:10.1038/s41596-020-0304-x

Zmasek CM, Knipe DM, Pellett PE, Scheuermann RH. Classification of human Herpesviridae proteins using Domain-architecture Aware Inference of Orthologs (DAIO). Virology. 2019;529:29-42. doi:10.1016/j.virol.2019.01.005

Zou HM, Huang ZF, Yang Y, et al. Human Cytomegalovirus Protein UL94 Targets MITA to Evade the Antiviral Immune Response. Journal of Virology. 2020;94(12). doi:10.1128/jvi.00022-20

Zuhair M, Smit GSA, Wallis G, et al. Estimation of the worldwide seroprevalence of cytomegalovirus: A systematic review and meta-analysis. Reviews in Medical Virology. 2019;29(3). doi:10.1002/rmv.2034

7. Appendices

7.1 List of Abbreviations

HCMV	Human cytomegalovirus
USFDA	United States Food and Drug Administration
Cryo-EM	Cryogenic Electron Microscopy
pULXX	Protein of Unique Long gene XX
DNA	Deoxy Ribonucleic Acid
CDXX	Cluster of Differentiation
HIV	Human Immuno-deficiency Virus
SNHL	Sensorineural hearing loss
IgG	Immunoglobulin G
HSCT	Hematopoietic stem cell transplantation
EMA	European Medicines Agency
dNTP	Deoxynucleotide triphosphate
MMOA	Molecular Mechanism of Action
ACE	Angiotensin-converting enzyme
HTS	High-throughput in-vitro screening
SPR	Surface Plasmon Resonance
SAR	Structure Activity Relationship
LLPS	Liquid-liquid phase separation
IE	Immediate early
RC	Replication compartment
DE	Delayed early
L	Late
HFF	Human foreskin fibroblast
Mab	Monoclonal antibody
MCP	Major capsid protein
CVSC	Capsid vertex-specific components
NEC	Nuclear egress complex
MTOC	Microtubule-organising center
NES	Nuclear egress signal
ESCRT	Endosomal sorting complex required for transport
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptor
SLM	Surface lysine methylation
CDR	Complemantarity determining region
SPA	Single particle analysis

7.2 List of tools used

Microsoft Office (Word, Excel and powerpoint)	For writing thesis, for analysing data and for figures preparation respectively.
Mendeley	Reference management
Benchling and Snapgene viewer	For sanger sequencing result and plasmid sequence analysis
Rf-cloning and Nebasechanger	For desining primers for cloning
Microsoft Copilot	For grammatical errors corrections in thesis
Biorender	For figures

7.3 List of Figures

7.4 List of Tables

8. Acknowledgements

First and foremost, I am immensely grateful to Professor **Dr. Thomas Krey** for the opportunity to write this thesis in his working group. His guidance, interesting projects, scientific freedom, and numerous opportunities to learn cutting-edge techniques, especially cryo-EM single particle analysis, have been invaluable. I deeply appreciate the many conversations, discussions, and advice beyond scientific aspects.

I thank **Deutsches Zentrum für Infektionsforschung (DZIF)** for funding my work. I thank University of Luebeck for the acceptance of my application for doctoral studies. I thank **Dr. Pierre Legrand** and **Dr. Heddy Soufari** from SOLEIL synchrotron facility for their support.

Special thanks to **Dr. Mauro Maiorca, Dr. Gonzalo Obal, Dr. Siavash Mostafavi**, and **PD Dr. Guido Hansen** for their support, guidance, and valuable insights on cryo-EM and image processing. I am grateful to Professor **Dr. Maya Topf** and **Dr. Guendalina Marini** for their collaboration on cryo-EM grid preparation, screening, and data collection.

I extend my gratitude to **Dr. Christian Löw** and **Dr. Tânia Custodio** for training me on the Sybody selection platform, and to **Dr. Sven Geisler** from the CAnaCore facility, University of Lübeck, for his support in FACS. I also thank **Dr. Stephan Neibling** and **Angelica Struve** from the EMBL SPC facility for their assistance in establishing cross-competition assays, and **Dr. George Ssebyatika** for his scientific support.

I am thankful to **Hera Fatima, Gabriela Diaz Degaldo, Nachiket Moti**, and **Chandamita Saikia** for the wonderful time in the lab and in Lübeck. My appreciation also goes to **Dr. Janna Bigalke** and **Dr. Jeroen Mesters** for their support with biophysical technologies, and to **Hans-Joachim Kraus** for his extensive support with the computational infrastructure.

I am grateful to my colleagues **Ms. Divya Selvakumar, Ms. Elisabeth Herold, Dr. Manoj Kumar Rajak, Dr. Johannes Cramer, Dr. Giacomo Castoro, Dr. Luisa Ströh, Mr. Jan Marc Beneke, Ms. Barbora Veselkova, Ms. Stefanie Koehler, Mr. Frithjof Besa, Mr. Sairam Dasika, Mr. Niklas Ebersberger, Dr. Mariana Grieben, Ms. Angela Pampel, Ms. Britta Schwarzloh, Mr. Walter Verheyen, Ms. Doris Mutschall, Ms. Susanne Zoske, Ms. Ye Ke**, and **Ms. Miriam Bernauer** for their unwavering support throughout my doctoral journey at the Institute of Biochemistry, University of Lübeck.

Special thanks to **Dr. Kumar Nagarathinam** and **Mrs. Visakalakshmi Kumar** for their scientific and moral support, and to **Dr. Senthil Kumar Devan, Dr. Mohan**, and **Dr. Balamurugan Sundaram** for their encouragement.

I would like to acknowledge my mentor from secondary and higher secondary school, **Mr. Selvakumar**, for instilling and nurturing my scientific interest since my early school days. I also thank my mentors **Dr. Ramachandraiah Gosu, Dr. Manish Kumar Thakur**, and **Mr. Krishnakumar V** for their support at the beginning of my post-college scientific career.

Finally, I thank my family—mom **Mrs. Gomathi**, dad **Mr. Thiyagarajan**, brother **Mr. Jeganath**, sister-in-law **Ms. Aswini**—and my wife **Ms. Sowmiya Devi**, friends **Mrs. Valli Kaviarasan**, **Mr. Kaviarasan**, **Mr. Saravanan Kandan, Late Mr. Inbaraj**, for their continuous support and encouragement and the Almighty, for the blessings throughout this journey. I also extend my gratitude to all other friends, family, and colleagues who have supported me in this journey.