UNIVERSIDAD DE GRANADA

Development of therapeutic proteins and compounds against viruses with class-I fusion machines

Doctoral thesis to obtain the PhD degree in Chemistry of the University of Granada, Spain. Faculty of Sciences. Department of Physical Chemistry.

PhD Program in Chemistry

2022

Mario Cano Muñoz

Director: Francisco Conejero Lara

Editor: Universidad de Granada. Tesis Doctorales
Autor: Mario Cano Muñoz
ISBN: 978-84-1117-400-8
URI: http://hdl.handle.net/10481/75621

"Decía Bernardo de Chartres que somos como enanos a los hombros de gigantes. Podemos ver más, y más lejos que ellos, no por la agudeza de nuestra vista ni por la altura de nuestro cuerpo, sino porque somos levantados por su gran altura."

Juan de Salisbury. Metalogicon, 1159 (III, 4).

Table of contents

Abbreviations

6-HB	Six Helix Bundle structure of Heptad Repeats 1 and 2
AIDS	Acquired Immunodeficiency Syndrome
AZT	Zidovudine
CCR5	Chemokine Receptor 5 (CD195)
CD	Circular Dichroism
CHR	Gp41 C-terminal Heptad Repeat
COVID-19	Coronavirus Disease 2019
CTP	C-terminal Pocket in the NHR region
CXCR4	CX Chemokine Receptor 4 (CD184)
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic acid
DSC	Differential Scanning Calorimetry
DSL	Gp41 Disulfide Loop
EboV	Ebola virus
ELISA	Enzyme-Linked Immunosorbent Assay
Env	HIV-1 Envelope Glycoprotein
FDA	United States Food and Drug Administration
FP	Gp41 Fusion Peptide
FPPR	Gp41 Fusion Peptide Proximal Region
gp120	Glycoprotein 120
gp160	Glycoprotein 160
gp41	Glycoprotein 41
HAART	Highly Active Antiretroviral Therapy
HIV	Human Immunodeficiency Virus
HP	Hydrophobic Pocket in the NHR region
HR1	Gp41 or S2 N-terminal Heptad Repeat
HR2	Gp41 or S2 C-terminal Heptad Repeat
HTLV	Human T cell leukemia virus
IgA	Immunoglobulin A
IgG	Immunoglobulin G
ITC	Isothermal Titration Calorimetry
MD	Molecular Dynamics
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MP	Middle Pocket in the NHR region
MPER	Gp41 Membrane Proximal External Region
mRNA	Messenger RNA
NHR	Gp41 N-terminal Heptad Repeat
NTP	N-terminal Pocket in the NHR region
PBD	gp41 Pocket Binding Domain
PDB	Protein Data Bank

PDB ID	Identifier for 3D structural data as per the Protein Data Bank database
RNA	Rribonucleic Acid
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SU	Envelope surface protein gp120
TM	Transmembrane protein gp41
T_m	Melting Temperature
ZIKV	Zika virus

Abstract

Zoonotic diseases have caused the world's biggest pandemics in recent history. HIV was first identified in 1981 and, since then, 70 million people have been infected with HIV and more than 37 million have died from AIDS-related illnesses. 40 years after, the SARS-CoV-2 outbreak let the world shocked crushing economies and killing more than 6 million people within the ongoing COVID-19 global pandemic.

Preventing those viruses from entering the host's cells constitutes a major goal in the development of effective anti-viral therapies. As enveloped viruses, HIV and SARS-CoV-2 need to fuse their membranes with those of human cells in order to infect them. Their class-I fusion machines, the HIV Envelope and the SARS-CoV-2 Spike respectively are responsible for membrane fusion. In this process the N-terminal Heptad Repeat (NHR or HR1) regions of gp41 and S2 subunits stand out because of their high conservation and because their interaction with the C-terminal Heptad Repeat (CHR or HR2) region is the driving force that ultimately leads to membrane fusion. Consequently, HR1 and HR2 represent vulnerable sites for the development of broad and effective therapies. In previous works, we designed and studied chimeric miniproteins (named covNHR) that mimic the exposed NHR of HIV gp41 in its pre-hairpin conformation as fusion inhibitors and as molecular models to explore the complexity of the interactions governing their function. In this thesis:

We have used the covNHR miniproteins in combination with a small-molecule fluorescent probe to investigate the conformational and dynamic properties of a conserved hydrophobic pocket (HP) located in the NHR-groove, which is a promising target for the development of small-molecule inhibitors. We demonstrated that the HP is conformationally flexible and connected allosterically to other NHR regions, which strongly influence the binding of potential ligands.

We have studied the role of buried polar residues in the interior of the NHR coil by mutating them to a hydrophobic amino acid, Isoleucine. The mutants show an extreme thermostability but an enhanced self-association and unaltered HIV-1 inhibitory activity. The results support a role of buried core polar residues in maintaining structural uniqueness and promoting an energetic coupling between conformational stability and NHR–CHR binding.

We have further stabilized covNHR proteins with the addition of an intra-chain disulfide-bond. The new disulfide-bonded strongly stabilizes the miniprotein, increases binding affinity for the CHR region and strongly improves HIV broad inhibitory activity, without targeting the preserved HP motif of gp41. The results suggest new strategies to inhibit HIV.

Finally, we have translated covNHR's approach to design and characterize chimeric proteins that imitate HR1 helical trimers of SARS-CoV-2 S2. The proteins bind strongly HR2 and have broad inhibitory activity against WT and Omicron viruses. Moreover, they are recognized by antibodies present in sera from COVID-19 patients, revealing previously undetected Spike epitopes that may guide the design of new vaccine and therapies.

Resumen

Las enfermedades zoonóticas han causado las mayores pandemias en la historia reciente. El VIH se identificó por primera vez en 1981 y, desde entonces, 70 millones de personas se han infectado y más de 37 millones han muerto por enfermedades relacionadas con el SIDA. 40 años después, el brote de SARS-CoV-2 ha dejado al mundo conmocionado, aplastando las economías de muchos países y matando a más de 6 millones de personas en el marco de la actual pandemia mundial de COVID-19.

Evitar que esos virus entren en las células del huésped constituye un objetivo importante en el desarrollo de terapias antivirales eficaces. Al tratarse de virus envueltos, el VIH y el SARS-CoV-2 necesitan fusionar sus membranas con las de las células humanas para infectarlas. Sus máquinas de fusión de clase I son responsables de esta fusión. En este proceso, las regiones con hepta-repeticiones N-terminales (NHR o HR1) de gp41 y S2 destacan por su alto nivel de conservación y porque su interacción con la región hepta-repetida C-terminal (CHR o HR2) es la fuerza impulsora que conduce a la fusión de membranas. En consecuencia, HR1 y HR2 representan sitios vulnerables para el desarrollo de terapias de amplio espectro y efectivas. En trabajos anteriores, diseñamos y estudiamos proteínas, llamadas covNHR, que imitan el NHR de gp41 en su conformación previa a la horquilla como inhibidores de fusión y como modelos moleculares para explorar la complejidad de las interacciones que rigen su función. En esta tesis:

Se han utilizado las proteínas covNHR en combinación con una sonda fluorescente de tipo molécula pequeña para investigar las propiedades conformacionales y dinámicas de un bolsillo hidrofóbico (HP) conservado ubicado en el surco NHR, que es un objetivo prometedor para el desarrollo de inhibidores de molécula pequeña. Demostramos que HP es conformacionalmente flexible y está conectado alostéricamente a otras regiones de NHR, lo que influye fuertemente en la unión de ligandos potenciales.

Se ha estudiado el papel de los residuos polares enterrados en el interior de NHR al mutarlos a un aminoácido hidrófobo, la isoleucina. Estos mutantes muestran una termoestabilidad extremadamente alta pero también una tendencia a la autoasociación y una actividad inhibidora del VIH-1 inalterada. Los resultados respaldan el papel de los residuos polares enterrados en el núcleo para el mantenimiento de la singularidad estructural y la existencia de un acoplamiento energético entre estabilidad conformacional y la unión de NHR-CHR.

Se ha estabilizado aún más las proteínas covNHR con la adición de un puente disulfuro intracatenario. El nuevo enlace disulfuro estabiliza fuertemente la proteína, aumenta la afinidad de unión por la región CHR y mejora fuertemente la amplia actividad inhibidora del VIH, sin dirigirse frente al motivo HP conservado de gp41. Estos resultados sugieren nuevas estrategias para lograr inhibir el VIH.

Finalmente, se ha trasladado el enfoque de las proteínas covNHR para diseñar y caracterizar proteínas quiméricas que imitan los trímeros helicoidales HR1 de S2 en el SARS-CoV-2. Las proteínas se unen fuertemente a HR2 y tienen una amplia actividad inhibitoria contra la cepa original y contra Omicron. Además, son reconocidas por los anticuerpos presentes en los sueros de pacientes de COVID-19, lo que revela epítopos de la proteína de la espícula no detectados previamente que pueden guiar el diseño de nuevas vacunas y terapias.

1. Two Global Pandemics.

1.1 Global Pandemics in a globalized world.

As early as the fifth century B.C., Hippocrates defined the foundations of medicine in his treatise on “Airs, Waters, and Places”. According to him, external conditions, including the air, influence human health: disease is then considered "no longer as a consequence of the internal disturbance of the humors, but as the result of interactions between the human body and its environment”[1].

Emerging infectious diseases (EIDs), especially those with zoonotic potential, are a growing threat to global health and economy. Climate change, as one of this century’s greatest threats to human health, is not only increasing food and water insecurity, extreme weather and pollution but also infectious disease threats[23]. Because zoonotic infectious diseases are situated at this nexus between environmental change, ecosystems and health, they are affected by climate change effects, such as global warming and geoclimatic variations which alter host, vector, and pathogen dynamics and their interactions[45]. The maintenance of healthy people requires the preservation of healthy ecosystems as the whole environment must be considered to achieve effective disease prevention[6].

Zoonoses constitute the main cause of emerging diseases affecting the human population[7]. The transmission of zoonotic viruses has led to major epidemics in the past and in recent years. Several outbreaks corroborate this fact, such as HIV in the early 80’s[8,9], SARS-CoV (severe acute respiratory syndrome coronavirus), which appeared in southern China in late 2002[10,11], MERS-CoV (Middle East respiratory syndrome coronavirus), which emerged in 2012 in Saudi Arabia, Avian influenza A H5N1, with the first case identified in May 1997 in Hong Kong, China[12], but has also been the foundation of many outbreaks since then[13], Ebola virus (EboV) was firstly discovered in 1975[14] although its first large epidemic is thought to have begun in late 2013 in Guinea, but was not detected and reported until March 2014[15], Zika virus (ZIKV), whose epidemic started in 2007 and spread over Africa and Asia[16], and SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), which originated in the city of Wuhan, province of Hubei, China at the end of 2019[17,18], among others.

1.2 AIDS and COVID-19 outbreaks: A glance at the past.

The 5th of June 1981 the Centers for Disease Control and Prevention (CDC) of the United States of America called a press conference where they described five cases of *Pneumocystis carinii* pneumonia in Los Angeles, California. Only one year before, the World Health Organization announced the eradication of smallpox in its XXXIII Convention. The summer of 1981 continued however surprising doctors and researchers because of the appearance of Kaposi's sarcoma in several patients. Although doctors knew both *Pneumocystis carinii* pneumonia and Kaposi's sarcoma, the combined appearance of both in several patients caught their attention. Most of these patients were sexually active homosexual men, many of whom also suffered from other chronic illnesses that were later identified as opportunistic infections. Blood tests carried out on these patients showed

that they lacked adequate numbers of $CD4^+$ T cells. Most of these patients died within a few months[19].

A year after, in 1982, the new disease was officially named Acquired Immune Deficiency Syndrome (AIDS), a name that thankfully replaced others proposed as Gay-related immune deficiency (GRID). Only a few months later, on 3 January 1983, Luc Montagnier's team[8] obtained a lymph node biopsy from a patient (BRU) with a lymphadenopathy in the neck where they found the first traces of a reverse transcriptase (RT) in the supernatant of the lymphocyte culture, indicating the presence of a retrovirus. The only retroviruses then known to date were the human T cell leukemia viruses, HTLV-I and HTLV-II, identified by Robert Gallo's group, but the RT couldn't be recognized by antibodies against HTLV. In the fall of 1983 and the winter of 1984 it was noted the favored tropism of this virus for $CD4^+$ T lymphocytes. In the spring of 1984, Gallo published more convincing evidence[20] that HIV (known at the time as HTLV-III) causes AIDS. The cloning and sequencing of the HIV genome came in 1985 with identification of new open reading frames specific for lentiviruses [21]. This was followed by identification of the HIV large surface glycoprotein [22] and of T cell CD4 as the receptor for HIV. In 1986, the virus was named Human Immunodeficiency Virus type 1 (HIV-1) and a second type called HIV-2 was isolated from West African patients [23]. A year after, in 1987, the first anti-HIV drug, azidothymidine (AZT), which blocks HIV RT activity, was introduced. With the arrival of the HIV protease inhibitors and triple drug therapy in 1995, many patients are alive today who would otherwise have died [24].

Although vast improvements have been achieved with regards to AIDS treatments Luc Montagnier was right when he wrote [24]:

"… we must not be complacent the task ahead is immense. We still do not understand the origin of the AIDS epidemic; the slow destruction of the immune system by factors in addition to HIV infection of $CD4^+$ T cells; the importance of cofactors in AIDS progression and virus transmission; and the nature of the HIV reservoir that resists triple drug therapy. The next wave of advances in the fight against this worldwide scourge will require the contribution and energy of us all."

As a consequence of the Ebola virus pandemic in 2014-2016 the WHO gathered an expert panel to identify emerging infectious diseases that may eventually progress towards a public health emergency and specify actions and measurements to be taken in order to prevent and curate illnesses associated to those diseases at the right time. They created a list of priority diseases regarding their epidemic potential in an effort to recognize country-specific difficulties in their hypothetical responses to an outbreak, promote research and development, and facilitate diagnostic, preventive, and therapeutic products in cases of public health emergencies[25,26]. Among the nine diseases included in the list there were human coronaviruses and a disease called Disease X whose description appropriately fits the COVID-19 outbreak[27].

The timelines of COVID-19 and AIDS outbreaks were however very different, with the COVID-19 pandemic developing much faster than the AIDS one. This was probably due to the human-to-human transmission of HIV being much slower, as contact with infected

bodily fluids is needed in order to spread the virus, while COVID-19, being originated by an airborne coronavirus infection, can spread more easily and faster from one individual to another. Public health organisms and the scientific community have also played an important role on the early evolution of the two pandemics. On 2020 all the efforts applied on testing, contact tracing, and strict health measures over the overall worldwide population paid off by largely managing to contain COVID-19. In addition, the scientific advances allowed for a rapid identification of the causal agent of COVID-19 and its subsequent study. Thanks to that, only several weeks elapsed since the detection of several patients presenting pneumonia of unknown etiology in December 2019 in Wuhan, a city with a population of 11 million inhabitants, to the isolation of the virus causing the infection on January 7th 2020. The virus was then called 2019-nCoV (2019-Novel CoronaVirus) and only three days after its isolation the sequencing of its genome was achieved and released. Almost two weeks later, on January 23rd Wuhan City was locked down. And on January 30th the World Health Organization (WHO), after not having declared the novel coronavirus a Public Health Emergency the week before, reconvened and declared the coronavirus outbreak a Public Health Emergency of International Concern[17,18]. By the end of the year and the beginning of 2021 several vaccine candidates were approved and by the end of 2021 the majority of the population in developed countries had been vaccinated and boosted with a 65% of the world's population being at least partially vaccinated with one dose and 59% with at least two doses as of 30th April 2022. However, it must be noted that there exists a huge inequality between the doses receipt by people living in developed countries, who have been vaccinated with up to three doses, and people living in developing countries, who have in many cases received none, especially in Africa where less than 5% of the population have received at least one dose[28].

1.3 The HIV pandemic and the SARS-CoV-2 pandemic: getting out of the frying pan into the fire.

HIV-1 arose through several independent zoonotic transmissions of simian immunodeficiency viruses during the past century[29]. AIDS pandemic has claimed more than 37 million lives over the past four decades[30]. Untreated HIV ultimately causes progressive $CD4^+$ T cell loss and a broad range of immunological abnormalities, leading to an increased risk of infectious and oncological complications[31]. The state of severe immune deficiency is defined as an HIV infection with either a $CD4^+$ T cell count <200 cells/μL or the occurrence of a specific AIDS-defining illness[31,32].

Although major global focus is currently set on the dramatic Coronavirus Disease 2019 (COVID-19) pandemic, we cannot forget that the HIV/AIDS pandemic is still very active and continues to be one of the world's largest pandemics to date with more than 40 million people currently living with HIV still representing a worldwide health issue[32] (Figure 1.3.1). What is more striking is the upsurge in HIV infections over different populations around the world, such as the outbreak in China's students, where the number of newly diagnosed college students has seen an annual growth rate ranging from 30 to 50% over the past several years[33]. All this together with the fact that HIV newly infects 1.8 million people each year, makes the development of an HIV vaccine a global health priority[34].

But almost 40 years after the discovery of HIV as the causative agent of AIDS we still do not have a licensed vaccine. Progress has been hindered by the extensive genetic variability of HIV and our limited understanding of immune responses required to protect against HIV acquisition[35].

This emphasizes the importance of continuing the development of new therapeutics to treat the infection, despite the fact that existing therapies such as modern Highly Active Antiretroviral Therapy (HAART) have helped to reduce the number of deaths. The absence of an effective vaccine combined with the growing emergence of multi-resistant HIV variants to several of these drugs urges for the development of some new anti-HIV compounds directed against the different stages of the virus life cycle[31,36,37]. Targeting the entry of HIV into the target cell constitutes an important inhibitory strategy because the infection could be prevented even before it happened, as once the virus has entered the cell and infected it, it can evade more easily the host's immune system integrating its genome with that of the cell and even creating reservoirs in certain cells where it remains latent and unsusceptible to drug treatments[32,38].

According to the World Health Organization the ongoing COVID-19 pandemic continues to devastate global health and economy[39,40]. To date, SARS-CoV-2 has infected over 485 million people and caused more than 6,130,000 deaths[28]. These digits may seem overwhelming if we compare them with those of HIV's pandemic nowadays but they are closer to the 1.8 million deaths per year reached at the peak of the HIV pandemic around 2004[41]. The data becomes more disturbing if we analyze the number of deaths among the number of total infected people, this gives us the death rate of each viral infection. For SARS-CoV-2 it approaches 1% while for HIV it exceeds 47%. Moreover, recent studies have claimed that about 1% of the infected people not being treated die because of SARS-CoV-2 infection, whereas almost the totality (>95%) of people infected with HIV who are not under treatment eventually die[42]. There also exist differences in the geographical distribution of the deaths caused by the two viruses (Figure 1.3.1). While HIV is almost geographically restricted to undeveloped countries, especially Africa, SARS-CoV-2 causes deaths across the world affecting to a high degree developed countries. This may be related with the differences in transmission rate between both viruses and the ability to reduce new infections following preventing measures in developed countries.

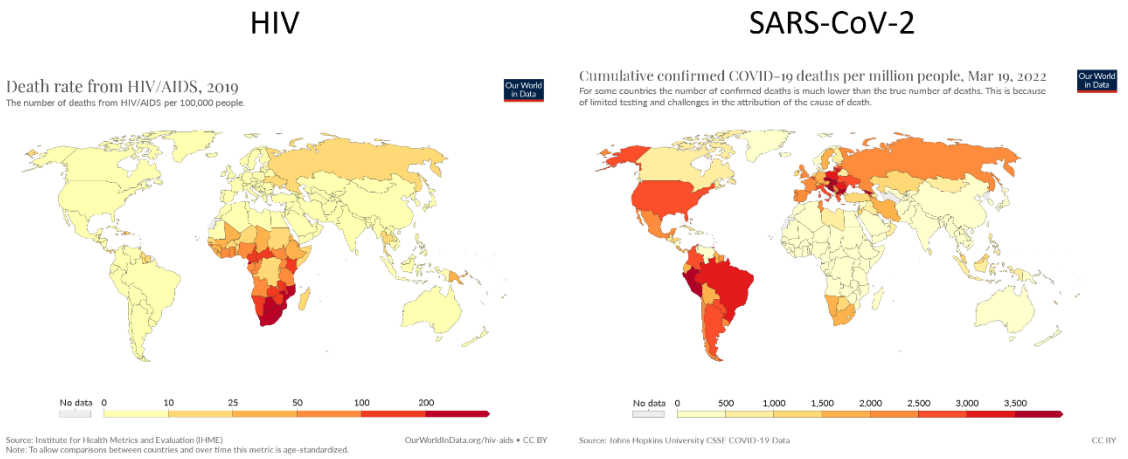


Figure 1.3.1: Comparison between deaths normalized per 100.000 and 1 M inhabitants for HIV and SARS-CoV-2 respectively. Left panel, data of HIV published by Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2021. Institute for Health Metrics and Evaluation. Link: http://ghdx.healthdata.org/gbd-results-tool. Right panel, data of SARS-CoV-2 published by COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. Link: https://github.com/CSSEGISandData/COVID-19. Data extracted using[43].

1.4 The viruses: Assemblies and Structures.

HIV is a Lentivirus belonging to the Retroviridae family and the Orthoretrovirinae subfamily. It is an enveloped virus, which means that it has a double lipid membrane derived from the host cell that houses the viral content. Viral particles or virions exist as roughly spherical nanoparticles about ~120 nm (100-150 nm) in diameter[44,45]. Its genome consists of two identical single-stranded positive-sense RNA molecules that are enclosed within the core of the virus particle, also known as the virion. The double-stranded HIV DNA integrated into the human genome is called proviral DNA and it's generated by the reverse transcription of the viral RNA genome into DNA, degradation of the RNA followed by an integration step[46].

HIV's genome features two LTR (Long Terminal Repeat) sequences at both ends (Figure 1.4.1). The LTRs participate in the integration of the provirus into the host genome. Once the provirus has been integrated, the 5'-LTR acts as the promoter for the entire retroviral genome, while the 3'-LTR acts as a source of RNA polyadenylation and, encodes the accessory protein, Nef[47]. In the 5'-3' reading frame direction the *gag* gene is the first being transcribed encoding the outer core membrane protein (MA, Figure 1.4.1), the capsid protein (CA) and the nucleocapsid protein (NC). In a different reading frame, the *pol* frame, encodes three different enzymes, namely the protease (PR), the reverse-transcriptase (RT) and the integrase (IN). Adjacent to the *pol* gene, the *env* reading frame encodes gp160 polyprotein that will later be split into the two envelope glycoproteins

gp120 (surface protein, SU) and gp41 (transmembrane protein, TM). Besides enzymes and structural proteins, the HIV genome encodes different regulatory proteins such as Tat (transactivator of transcription protein), Rev (regulator of virion protein) which are required for the initiation of HIV replication; and other proteins involved in replication, virus budding and pathogenesis: Nef (negative regulatory factor), Vif (viral infectivity factor), Vpr (viral protein r) and Vpu (viral protein u)[38,48,49]. These proteins and their three-dimensional structure are depicted in Figure 1.4.2. HIV-2 encodes Vpx (virus protein x) instead of Vpu, which is partially responsible for the reduced pathogenicity of HIV-2[50].

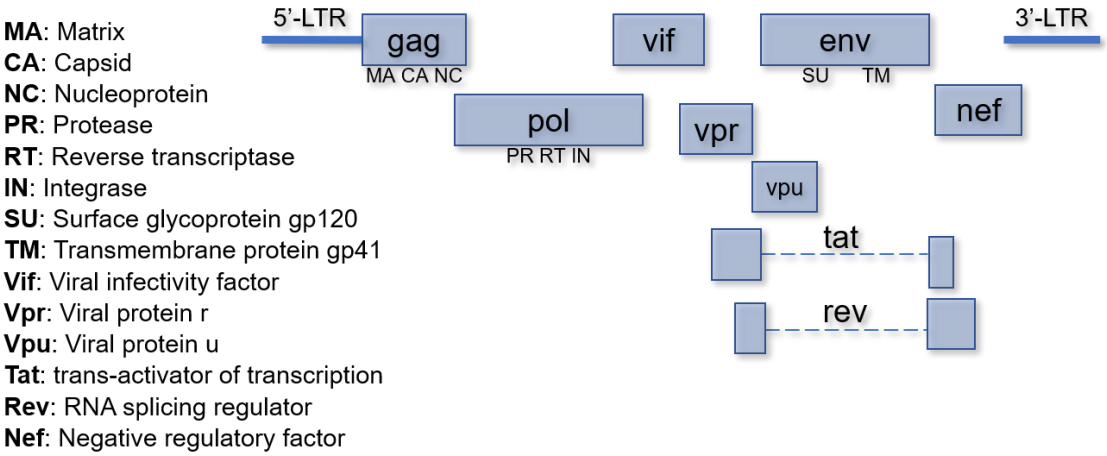


Figure 1.4.1: Schematic representation of the HIV-1 genome. Modified from [51].

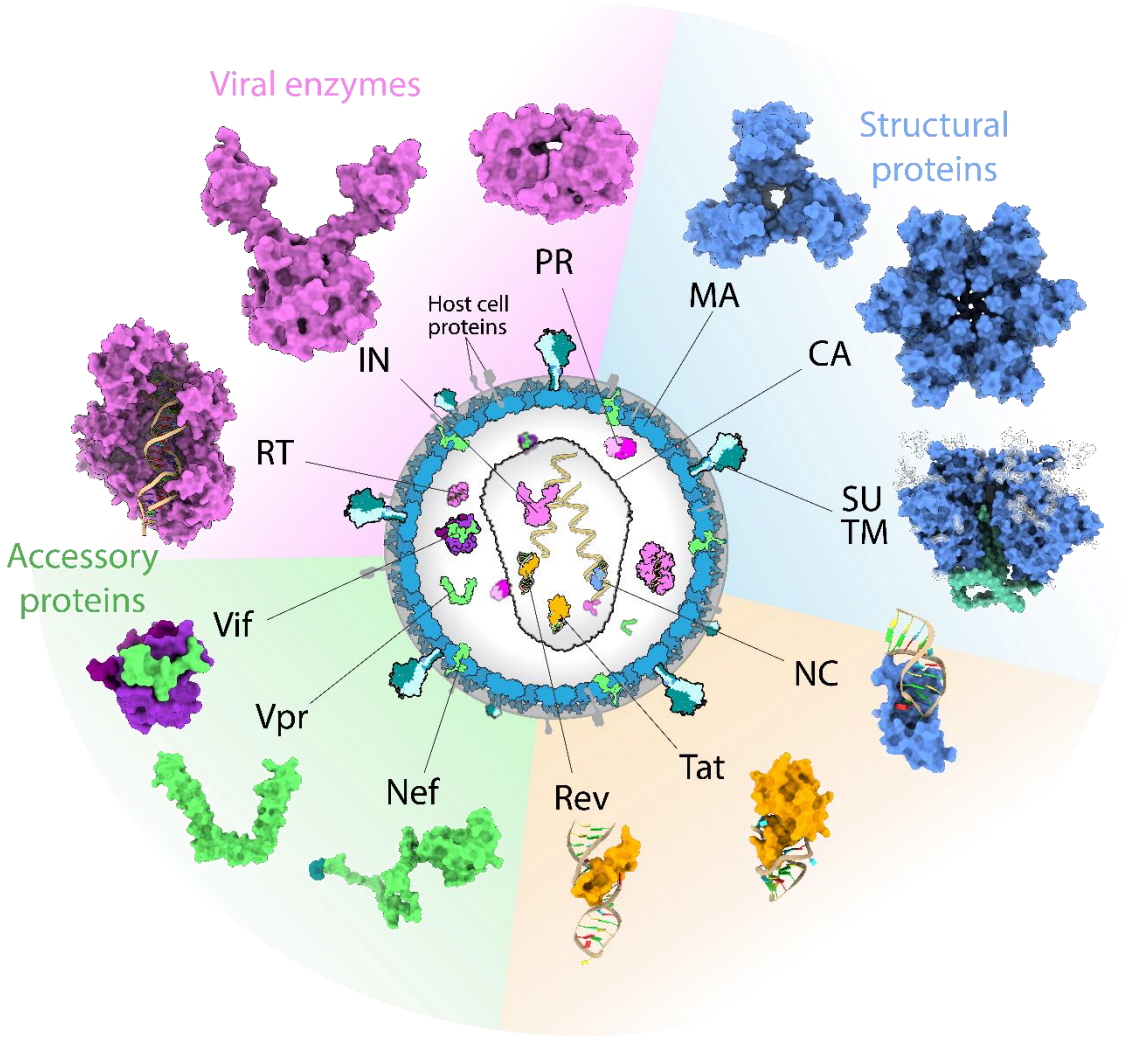
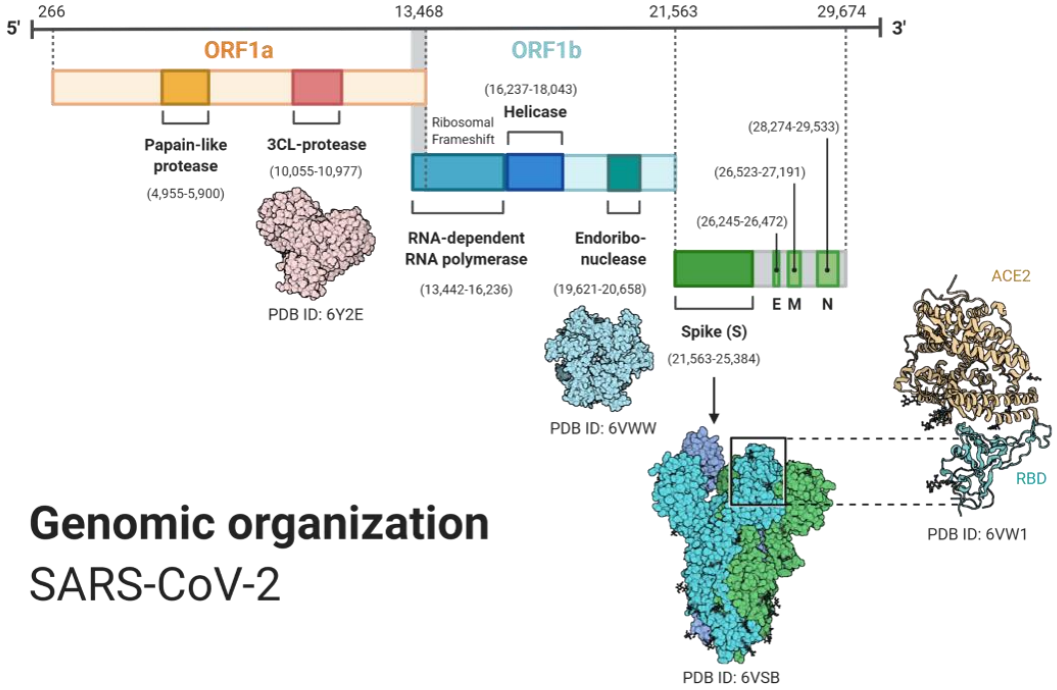


Figure 1.4.2: Diagram of the HIV virion and its constituent proteins showing their distribution in the fully mature virion. 15 proteins compose the HIV virion together with two ssRNA (+) strands and several proteins from the last infected host cell. The proteins whose strucutre is known are represented in the illustration: MA, matrix protein (PDB ID: 1HIW); CA, capsid protein (PDB ID: 3H47); SU and TM, envelope surface and transmembrane proteins (PDB ID: 4NCO); NC, nucleocapsid protein (PDB ID: 1ALT); Tat, trans-activator of transcription (PDB IDs: 1BIV and 1JFW); Rev, regulator of virion (PDB ID: 1ETF); Nef, negative regulatory factor (PDB ID: 1AVV and 1QA5); Vpr, viral protein r (PDB ID: 1ESX); Vif, viral infectivity factor (PDB ID: 3DCG); RT, reverse transcriptase (PDB ID: 1HYS); IN, integrase (PDB ID: 1EX4); PR, protease (PDB ID: 1HPV). The figure was inspired from "The Structural Biology of HIV" Poster illustration by David S. Goodsell, The Scripps Research Institute: shorturl.at/ikwyB and the "Diagram of the HIV virion" by Thomas Splettstoesser (www.scistyle.com).

Coronaviruses exhibit the largest genomes among all known RNA viruses. While HIV genome consists approximately of 9.2-9.6 kb[49], SARS-CoV-2 genome size varies from 29.8 kb to 29.9 kb[52]. SARS-CoV-2, as other related *betacoronaviruses* such as SARS-CoV and MERS-CoV, presents a positive-stranded RNA genome. The 5'-end of the RNA genome encodes two polypeptides that are then proteolytically processed to produce 16 non-structural proteins. The 3'-end encodes the structural proteins S (spike glycoprotein,

Figure 1.4.3), N (nucleocapsid protein), M (membrane protein) and E (Envelope protein)[53,54].



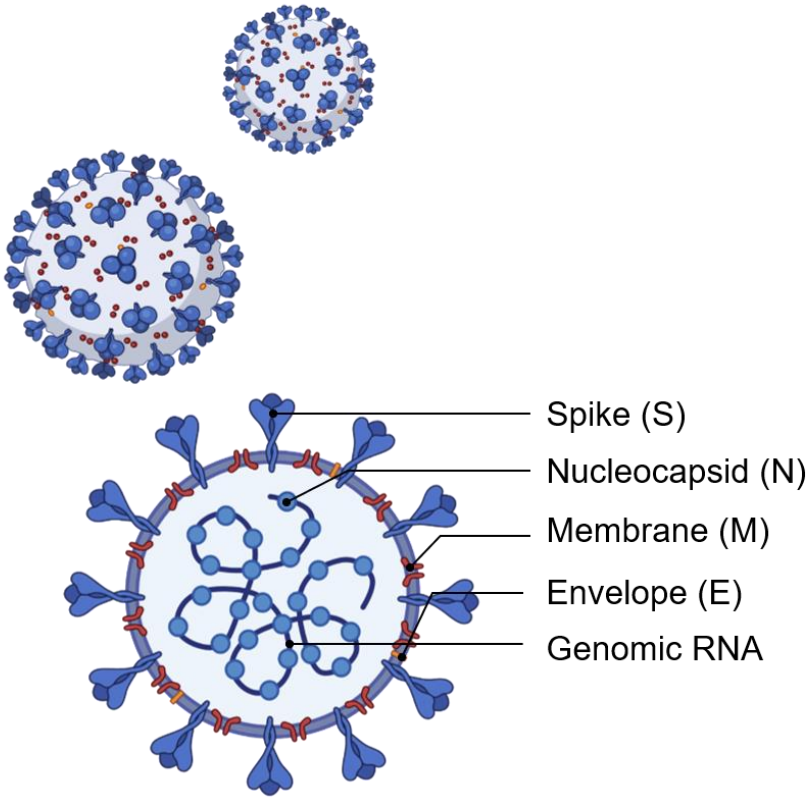


Figure 1.4.3: Genomic organization and protein distribution in SARS-CoV-2. SARS-CoV-2 genome (top panel) presents two genes: ORF1a (yellow), ORF1b (blue). They encode 16 non-structural proteins and several structural proteins (bottom panel, each protein is shown in context of the virion particle). The figure was created with BioRender.

1.5 The life cycle of HIV and SARS-CoV-2

HIV uses the cellular machinery in order to replicate itself. The replication process starts after the virus productively engages the target cell's surface receptors and ends when nascent particles mature into infectious virions (for an overview of the life cycle see Figure 1.5.1). Throughout this entire cycle, HIV exploits a myriad of cellular factors to replicate whereas, at the same time, the host restriction factors fight, many times unsuccessfully, to try to suppress this process[55,56].

The HIV surface envelope spikes are composed of gp120 and gp41 proteins and are responsible of cellular receptor recognition, binding and fusion. HIV-1 primarily infects $CD4^+$ T cells and macrophages. Upon receptor and co-receptor binding the spikes initiate a cascade of conformational changes that culminates in fusion between the viral and host cell membranes and release of the viral core into the cytoplasm causing infection of the human target cell[31].

Following the attachment and fusion steps and once the viral content has entered into the target cell, the cone-shaped capsid starts to partially dissolve releasing its genomic RNA together with other enzymes such as RT and IN. This allows the RNA to be reverse-transcribed via the RT. The RT doesn't have any proofreading activity, thus, in average one incorrect nucleotide per transcription round is incorporated into the proviral DNA leading to a broad spectrum of various quasispecies that develop in a patient in the course of infection[49]. The resultant linear viral DNA molecule, in the form of a pre-integration complex (PIC) associated with NC and IN among other proteins, is then integrated into a host cell chromosome after entering the cell nucleus aided by the host chromatin-binding protein lens epithelium-derived growth factor (LEDGF), thus forming the integrated provirus[57].

The creation of the provirus is a turning point in the life cycle of the retrovirus establishing the transition between the early and late phase of the replication. The transcription step is initiated by Tat and viral mRNAs are produced and exported from the nucleus[57].

In the cytoplasm, the *gag* gene is then translated into a precursor polypeptide protein Gag, which is further processed to generate the MA, CA and NC proteins[58]. HIV-1 virion assembly occurs at the plasma membrane. The Gag polyprotein is in charge of mediating all the essential events in virion assembly[59].

The last step in the HIV life cycle is the maturation of viral particles to infectious virions by PR via proteolysis of the precursor polypeptides Gag and Gag-Pol in order to obtain the final structural proteins (MA, CA and NC) and enzymes (PR, RT and IN) [57,60]. Several steps along the pathway of HIV-1 assembly and particle release from cells have been targeted for antiviral drug development[61].

The life cycle of SARS-CoV-2 includes the following major events: viral entry, replication and transcription, assembly and release. Of note is the lack of viral genome integration into the host's genome[62] as opposed to this relevant step present in HIV's life cycle. Beyond that, SARS-CoV-2 life cycle (Figure 1.5.2) presents many similarities with that of HIV. At the very first steps the S protein on the outer surface of the virion is

responsible for membrane fusion[63]. Once the viral content is inside the cell genomic RNA is released in the cytoplasm where it is immediately translated using the cellular ribosomes into two large open reading frames: ORF1a and ORF1b. The two resulting polyproteins are proteolytically processed by M^{pro} (main protease) and papain-like protease (PL^{pro}) into 16 non-structural proteins (NSPs). The NSPs arrange into the viral replication and transcription complex (RTC, not shown in Figure 1.5.2 for the sake of clarity). This is followed by translation of structural proteins, which then translocate into endoplasmic reticulum (ER) membranes where they will transit through the ER-to-Golgi intermediate compartment (ERGIC). The newly produced viral genomic RNA will then interact with N-protein forming a complex. Meanwhile the S, E and M proteins are inserted into the viral envelope at Golgi Apparatus (GA) intermediate compartments. Finally, the mature viral particles are released from the infected cell via exocytosis to initiate a subsequent round of infection[64–66].

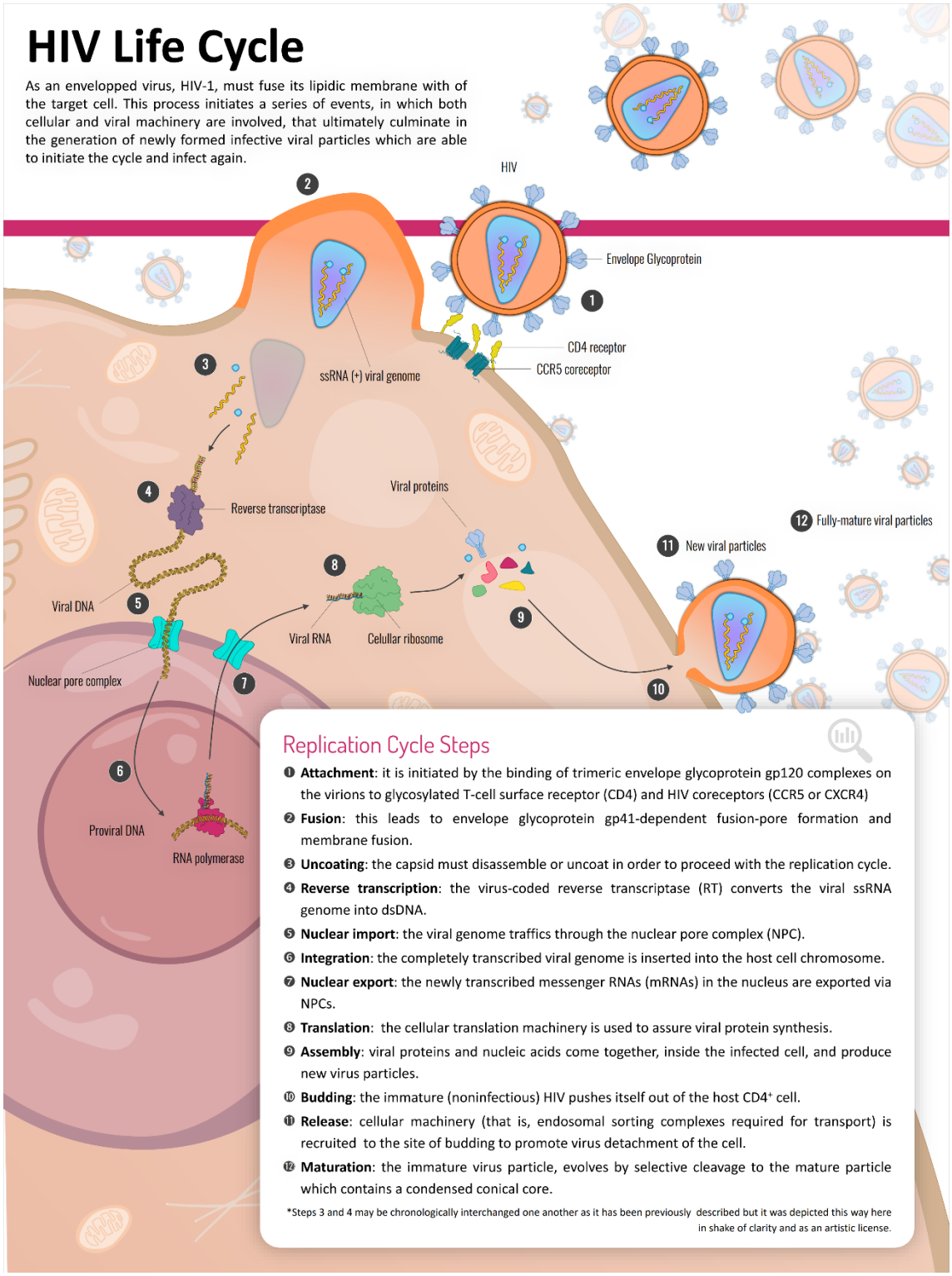


Figure 1.5.1: HIV life cycle. Details are provided in the text and in the figure legend for rapid contextualization. Description and conceptualization are based on several studies detailing the different steps comprised in HIV's life cycle, see text for details.

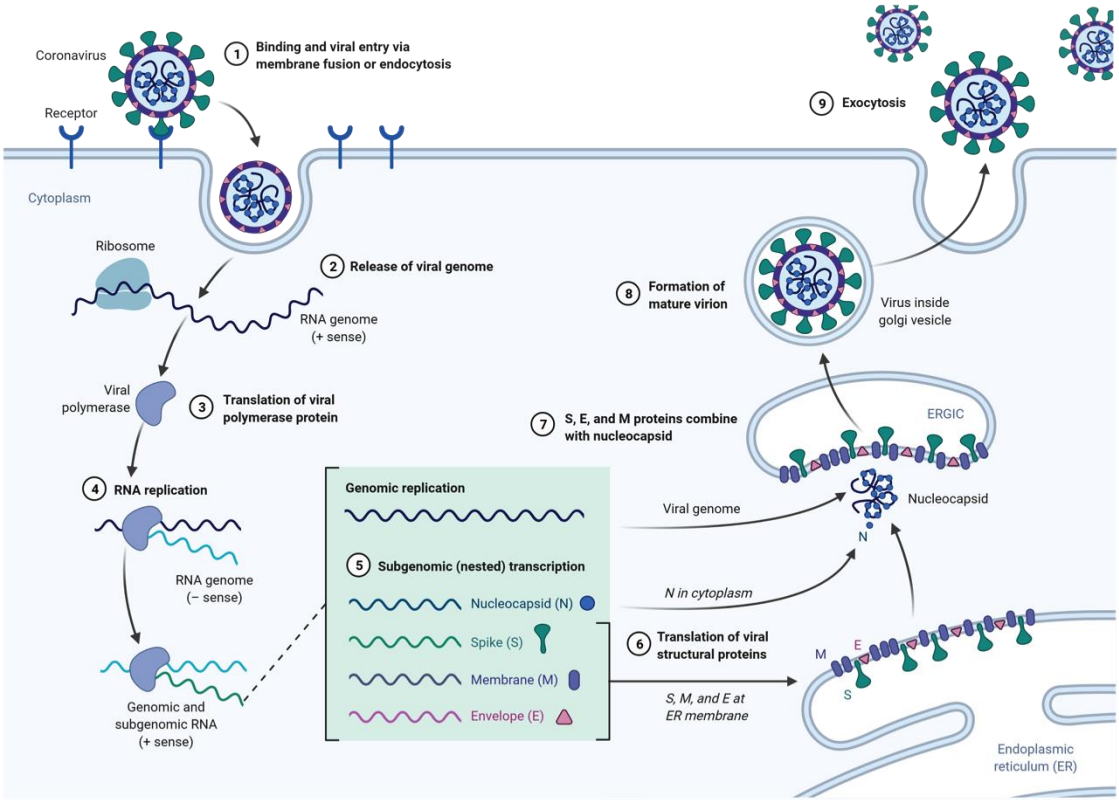


Figure 1.5.2: SARS-CoV-2 life cycle. It consists of several steps numbered in the Figure in order of occurrence, see text for details. The Figure was created with BioRender.

1.5.1 General treatments against both viruses.

Since the approval of the first HIV inhibitor, AZT, in 1987 numerous therapeutics have been developed targeting several important steps in the virus life cycle (Table 1). Those include nucleoside (NRTIs), nucleotide (NtRTIs) and non-nucleoside analogues (NNRTIs), reverse transcriptase inhibitors combined with protease inhibitors[67] (PI or PRI) and/or additionally a fusion or integrase inhibitor (INI)[68]. Moreover, new drugs are currently being developed[69,70]. The initial therapy against HIV infection should start with a highly effective and at the same time safe and well-tolerated combination of two NRTIs with one NNRTI. The combination of several mechanistically different inhibitors should delay the development of resistant HIV, although this is not always achieved; moreover, a percentage of the newly infected patients is infected by drug-resistant HIV variants[49]. Therefore, a genotypic resistance test should be done prior to treatment in order to avoid a reduced effectiveness of the chosen therapy[71].

Table 1: Overview of the drugs available for HIV therapy. Data obtained from [49].

Nucleoside/nucleotide reverse transcriptase inhibitors (NRTI)	Non-nucleoside reverse transcriptase inhibitors (NNRTI)	Protease inhibitors (PRI or PI)	Integrase inhibitors (INI)	Entry inhibitors
Azidothymidine – Zidovudine (1987)	Nevirapine (1996)	Saquinavir (1995)	Raltegravir (2007)	Enfuvirtide (T20) (2003)
Didanosine (1991)	Efavirenz (1998)	Indinavir (1996)	Elvitegravir (2013)	Maraviroc (2007)
Zalcitabine (1992)	Delavirdine (2007)	Ritonavir (1996)	Dolutegravir (2013)	
Stavudine (1994)	Etravirine (2008)	Nelfinavir (1997)		
Lamivudine (1995)	Rilpivirine (2011)	Lopinavir + Ritonavir (2000)		
Abacavir (1998)		Atazanavir (2003)		
Tenofovir (2001)		Fosamprenavir (2003)		
Emtricitabine (2003)		Tripanavir (2005)		
		Darunavir (2006)		

Year of approval in parentheses.

To date, there has been a lack of effective therapies to treat COVID-19. Due to the rampant and continuous spread of COVID-19, it is still relevant to progress in the identification and characterization of new antivirals[72]. Several pharmaceutical companies have attempted to develop therapeutic drugs against different steps in SARS-CoV-2 life cycle, including Merck's Lagevrio© which uses molnupiravir, a polymerase inhibitor, as active compound; and Pfizer's Paxlovid© which is intended to be a treatment composed of nirmatrelvir and ritonavir, a protease inhibitor and a protector of nirmatrelvir which prolongs its half-life in the organism[73]. Unfortunately, to date, there is only one treatment approved by the FDA (Food and Drug Administration) against COVID-19[74]: remdesivir, an adenosine analog that interferes with the RNA polymerase inhibiting viral replication by prematurely terminating RNA transcription[75]. The lack of effective therapies to treat COVID-19 together with the wide and continuous spread of the disease makes urgent the discover and characterization of new drugs that target important steps in SARS-CoV-2 life cycle[72].

On the other hand, three FDA-approved vaccines currently exist: Pfizer-BioNTech[76], Moderna[77] and Janssen[78] COVID-19 vaccines[79] and several others approved in non-USA countries. All them were approved under Emergency Use Authorization protocols to facilitate the availability of vaccines during public health emergencies[80]. Soon after the discovery of HIV the goal was also to develop an effective vaccine to protect the population against it, however, 40 years later there is still no approved vaccine. HIV-1's genetic variability and its ability to evade the host's immune response may be the main factors behind the difficulties of developing an effective HIV vaccine[35,81].

1.6 The Fusion Machinery of SARS-CoV-2 and HIV.

As enveloped viruses, both SARS-CoV-2 and HIV need to fuse its lipidic membrane with that of the target human cell in order to enter it and cause infection. The entry process can be further divided into three stages: attachment, priming and fusion, as I will be describing in this section.

In HIV, the envelope protein complex, Env, mediates attachment of the virus to the cell surface. Each virion has eight to ten Env spikes on its envelope[82] and each one consists of three copies of gp41 and gp120, whose structures are important to understand the fusion process. The Env complex, Figure 1.6.1, is synthesized as a precursor polyprotein, gp160, which is then cleaved in the trans-Golgi network by furin or furin-like proteases.

As a result of the proteolytic cleavage, the surface glycoprotein (SU) gp120 and the transmembrane glycoprotein (TM) gp41 are generated, which are now held together in a non-covalent manner. These two protein subunits are incorporated into the viral particle, or virion, forming a trimer of heterodimers that constitutes the Env complex or the spike on the surface of the virus [82,83,84].

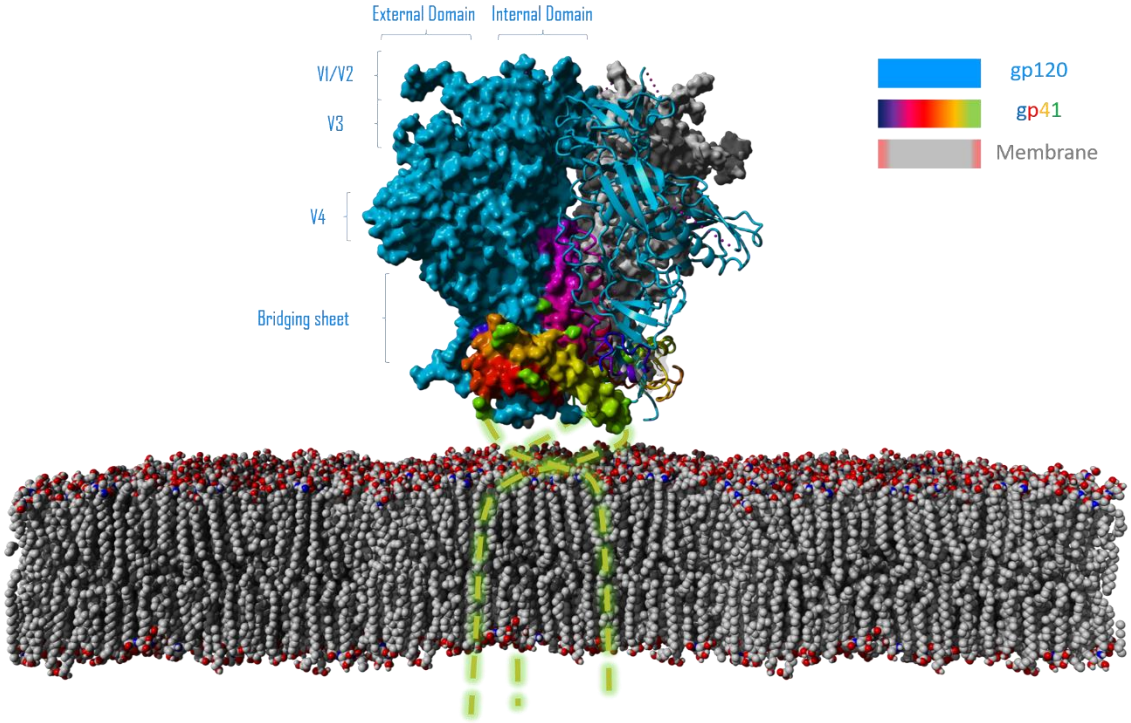


Figure 1.6.1: External part of the Env complex in a pre-fusion state. One protomer is represented with ribbons, another protomer in blue molecular surface and the third one is represented in gray molecular surface. The most important structural elements of gp120 are labelled. The structure of the membrane proximal, transmembrane and cytoplasmic gp41 subdomains is unknown and they are represented schematically as a dashed green line.

The Env complex is the only viral protein structure located on the surface of the HIV virion envelope, which makes it a key target both in the treatment of infection and in the development of a vaccine or in the search for neutralizing antibodies[85]. However, Env presents defense mechanisms such as the dense shield of glycosylated residues that protects it from neutralizing antibodies, which is reflected in the fact that half of the mass of gp120 is composed of carbohydrates, making it one of the most glycosylated proteins ever known[45,83].

Most of the elements exposed on the surface of the Env complex belong to gp120, which is located at the most distal end with respect to the viral membrane and partially occludes the gp41 structure, as it is located around the latter. The trimerization domains of gp120 are located in its internal domain and hold the three gp120 subunits together in the apical region of the complex. But the most important interactions that maintain the native

structure of Env are located at the interface between gp41 and gp120 and are responsible for keeping a gp41 subunit non-covalently bound to another gp120 subunit, being this structural unit called a protomer. Specifically, in this structure each gp41 subunit surrounds the amino and carboxyl termini of a gp120 subunit via four helices that form a collar close to the membrane. This wrapping effect and the collar clasp are depicted in Figure 1.6.2. In addition, hydrophobic interactions between the gp41 subunits hold the three protomers together [85,86].

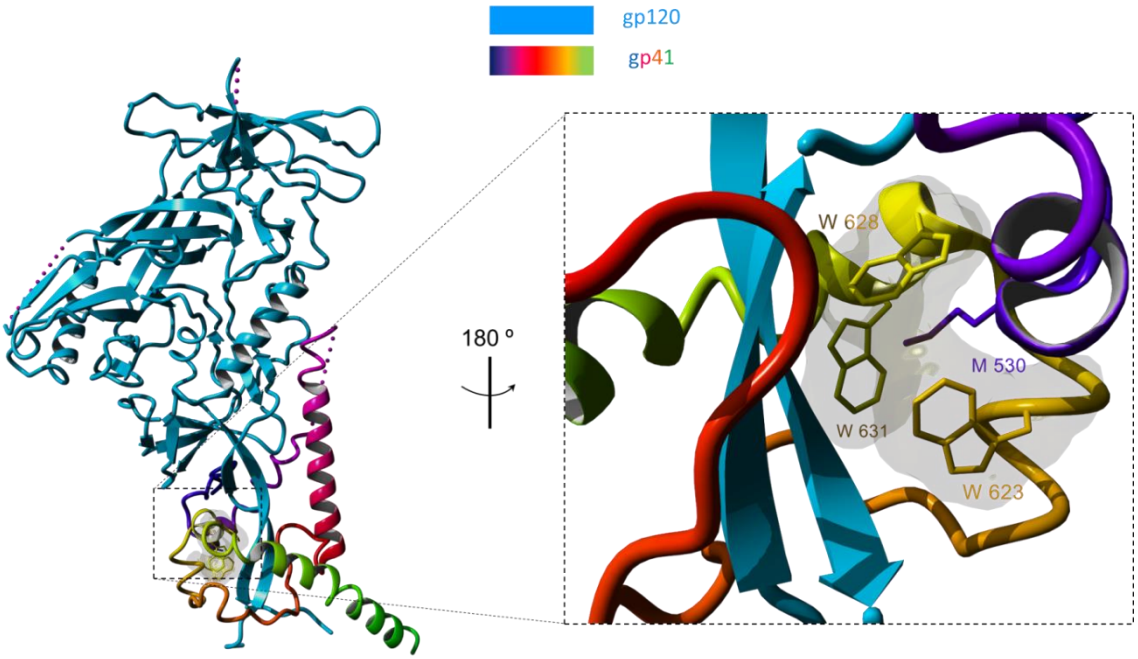


Figure 1.6.2: Three-dimensional arrangement of an Env protomer formed by gp120 and gp41. The non-covalent interaction that holds the heterodimer together is shown in the enlargement.

Gp120 (Figure 1.6.1, blue) mediates the attachment and the binding to the receptor (CD4) and co-receptor (CCR5 or CXCR4) during the virus entry process. It consists of five conserved domains (C1-C5) and five variable domains (V1-V5). The conserved regions make up the core of gp120 and contain the critical domains for receptor binding. On the other hand, the variable domains are close to the surface of gp120 forming exposed loops that play an important role in the binding of Env to the co-receptor[87].

Gp41 (Figure 1.6.3-a) is made up of three domains: the extracellular domain or ectodomain (ED), the transmembrane domain (TMD) and the cytoplasmic domain (CD). The main functions of gp41 are mediated by the ectodomain, which can be, in turn, subdivided into 6 regions: the fusion peptide (FP), followed by the fusion peptide proximal region (FPPR), the N-terminal heptad-repeat (NHR) region, the disulfide bridged loop (DSL) region, the C-terminal heptad-repeat region (CHR) and the membrane-proximal external region (MPER)[37,45,84,88]. Due to the fact that gp41 comes from the proteolytic cleavage of gp160, there are two types of numbering that index the

protein residues differently, in this work I will follow the numbering of gp160 in which the first amino acid of gp41 is designated as Ala512[89].

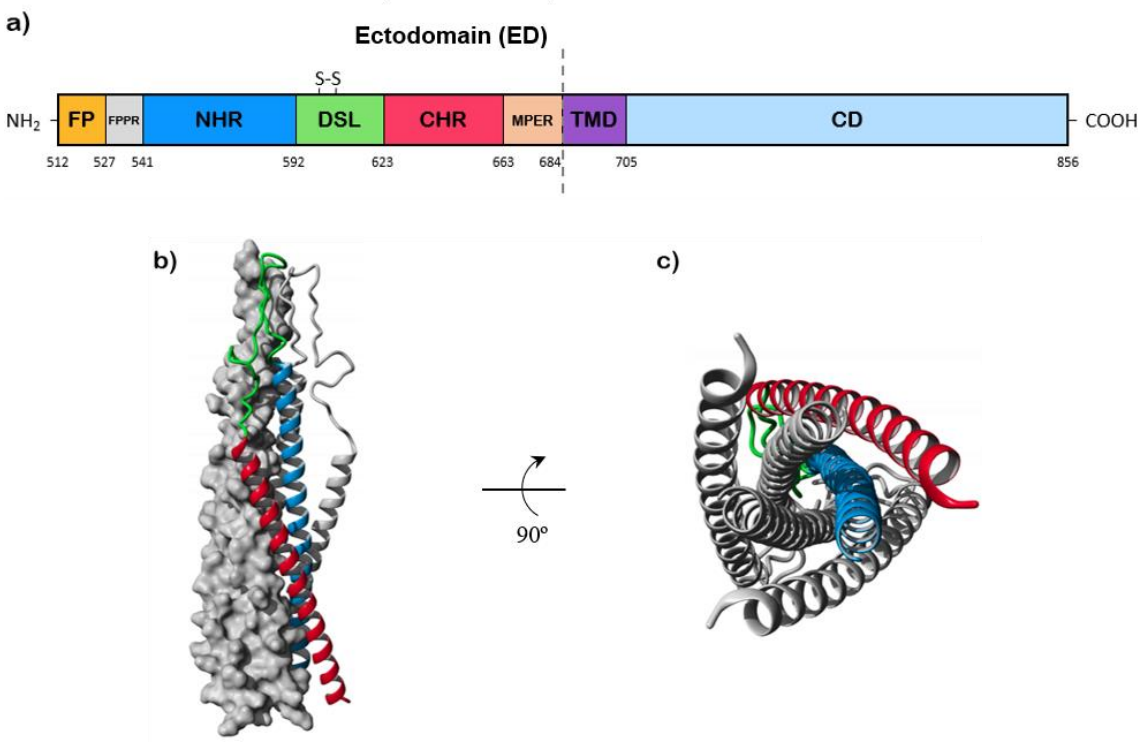


Figure 1.6.3: HIV-1 gp41 structure. a) To-scale representation of gp41 regions and domains within the primary sequence following gp160 numbering. Adapted from[89]. b) Side view of the six-helix bundle of gp41 in its post-fusion structure. One gp41 subunit is shown in molecular surface and colored in gray; a second subunit is also colored in gray but represented in ribbons; and the third subunit is also shown in ribbons and colored according to a). c) Bottom view of the bundle of six gp41 helices in ribbon scheme, colored as in b), which allows the appreciation of the supercoiled helices (PDB ID: 1IF3).

SARS-CoV-2 is closely related to SARS-CoV and shares high sequence homology levels (76% amino acid identity throughout the entire S protein[64]) although they have some differences in their replication cycles and infectious potential. Like the rest of the coronaviruses, they display the distinctive spike-shaped protein structures in their envelope, formed by the so-called S-glycoprotein. The S (Figure 1.5.4) protein is a class-I fusion protein, which is responsible for adhering to the host cell membrane and causing its infection [90–92].

The S protein is one of the largest class-I fusion machines known, with 1273 amino acids. As opposed to HIV's envelope, the native S protein is a homotrimer instead of a trimer of heterodimers because their two subunits, S1 and S2, are linked through the same polypeptide chain until proteolytic priming is accomplished after attachment. The S1 subunit, (Figure 1.6.5) consists of an N-terminal domain (NTD) and a Receptor Binding Domain (RBD) which is responsible for cellular receptor ACE2 binding during the attachment phase. The S2 subunit encompasses a Fusion Peptide (FP), a Fusion Peptide Proximal Region (FPPR), an heptapeptide repeat region 1 (HR1, analogous to NHR in

HIV), a Central Helix (CH), a Connector Domain (CD), an heptapeptide repeat region 2 (HR2, similar to CHR in HIV), a transmembrane domain (TM) and a short cytoplasmic domain (CT)[9394]. Both SARS-CoV-2 and HIV's spikes are highly glycosylated, especially that from SARS-CoV-2 with up to 22 N-glycosilation sites per protomer which act as a glycan shield to evade the host's immune system[95].

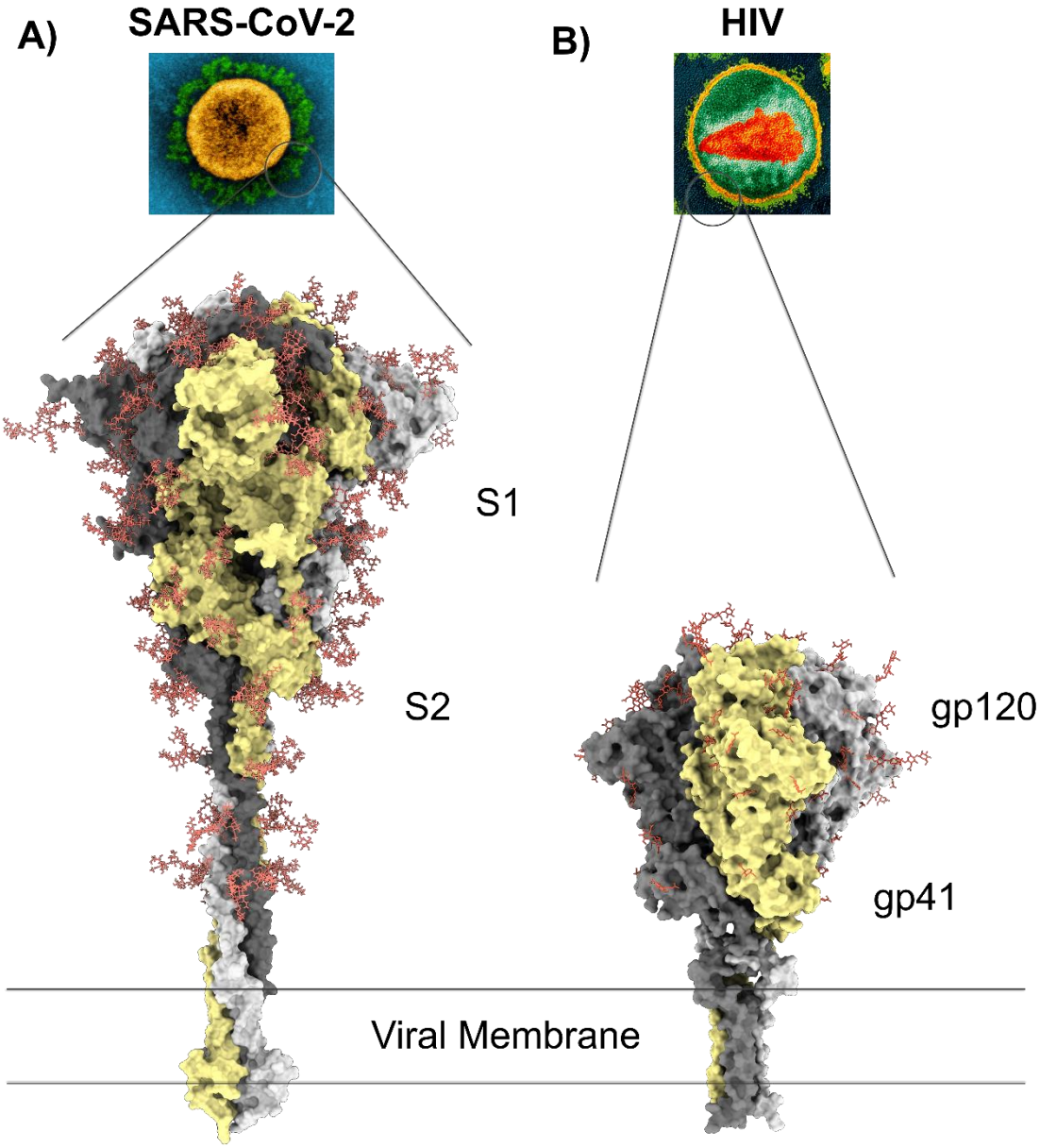


Figure 1.6.4: Structure of the class-I fusion proteins of SARS-CoV-2 and HIV. Both panels show electron tomographs of isolated viral particles with their respective characteristic crown-like thorns on their surface composed of their envelope and spike proteins, respectively. A) SARS-CoV-2 is surrounded by a larger amount of spike proteins whose three-dimensional structure is represented below as a model of the fully-glycosylated full-length S protein complex. The model was taken from [96]. B) HIV particle is enclosed by a fewer amount of envelope proteins whose structure is depicted below the electron micrograph image of the virion. It is an in-house model of the structure of a native Env with the ectodomain fused to the transmembrane and part of the C-terminal

domain made ad-hoc to compare with that from A) and joined as described elsewhere [97] (PDB IDs: 4TVP and 6E8W respectively).

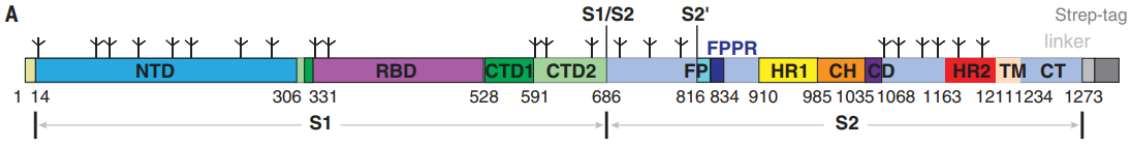


Figure 1.6.5: Diagram representation of the sequence of SARS-CoV-2 spike. Numbers refer to the amino acids that delimitate each functional region. Figure reproduced from [94].

1.7 Entry mechanism and Fusion.

The HIV entry process (Figure 1.7.1) begins with the binding of gp120 to the cellular CD4 receptor, whereby contacts are established between the D1 domain of CD4 and the CD4-binding site of gp120, a highly conserved and carbohydrate-free region. CD4 binding causes large conformational changes in gp120: first, the gp120 bridging sheet region becomes an ordered structure consisting of 4 β-sheets, which alters the position and flexibility of the V1/V2 loops. Second, the V3 loop extends and projects in the opposite direction of the virion towards the cell membrane. Third, the gp120 subunits rotate upon the Env complex so that the orientation of gp120 is altered in such a way that the bridging sheet and V3 loop are positioned toward the cell membrane to favor interaction with the coreceptor. Thus, CD4 binding induces a conformational transition from an unbound state to a CD4-bound state that has the effect of creating and exposing the co-receptor binding site, which is otherwise rarely exposed as a mechanism of immune evasion[84].

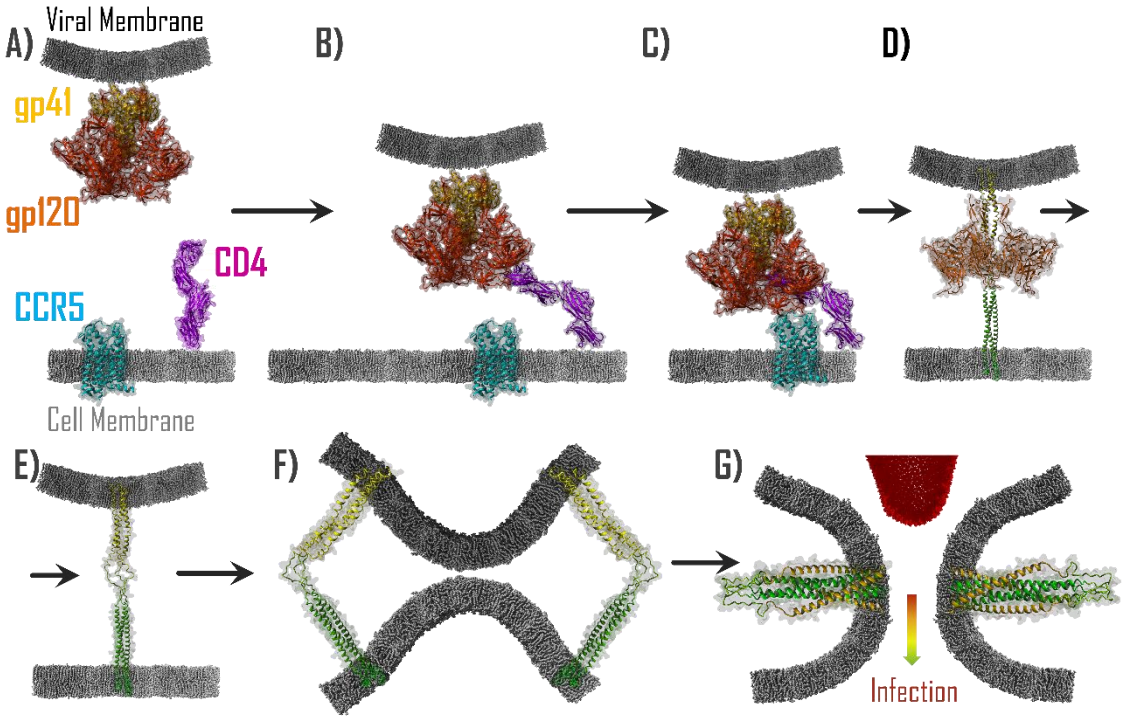


Figure 1.7.1: Fusion mechanism of HIV-1 and the CD4⁺ host cell. A) The Env complex is shown in orange (gp120) and yellow (gp41); CD4 receptor is shown in purple; CCR5 co-receptor is shown in blue. The viral membrane (top) is shown in dark gray, while the cellular membrane (bottom) is shown in light gray. B) Binding to CD4: Binding of gp120 to CD4 brings the viral membrane closer to the cell membrane and causes a conformational change that exposes the co-receptor binding site. C) Binding to the co-receptor. D) The binding of gp120 to the co-receptor promotes a second conformational change that results in the exposure of gp41 FP (in shades of yellow to green from the C- to the N-terminus) and its insertion into the cell membrane. E) Gp41 connects both membranes, this is its most exposed to inhibitors conformation. F) Hemifusion: The folding of the CHR and NHR regions of gp41 brings both membranes closer together. G) The formation of 6-HB (two 6-HB are shown in the image) promotes the fusion of the two membranes and the viral content (represented by the protein capsid in red) is able to penetrate into the cell interior. The cytoplasmic domain of gp41 has been removed in shake of clarity. Figure inspired from [87].

Unlike most viruses with class-I fusion machines, HIV requires binding of gp120 to a second receptor, the co-receptor, which may be either CCR5 or CXCR4. Due to the smaller extracellular projection that the co-receptor exhibits, the CD4 receptor undergoes a subsequent conformational change whereby it folds at its hinge region between the D2 and D3 domains (Figure 1.7.1-C). This approach favors the interaction between gp120 and the co-receptor causing a subsequent conformational change that results in the exposure of the gp41 fusion peptide (FP) and its insertion into the host cell membrane (Figure 1.7.1-D)[84]. At some point after FP projection, gp120 shedding occurs (Figure 1.7.1-E), temporarily leaving gp41 in its most exposed conformation, the pre-hairpin or pre-bundle intermediate, accessible to inhibitor binding, with its C-terminal end inserted into the viral membrane and the N-terminal end into the cell membrane thus establishing a connection bridge between both membranes[98].

This pre-fusion intermediate structure undergoes further conformational changes that culminate in the fusion of both the viral and cellular membranes (Figure 1.7.1-F-G). At this point, the NHR regions of gp41, each exhibiting α-helix secondary structure, are assumed to be organized as a coiled-coil trimer exposing three highly-conserved hydrophobic grooves on the surface of the supercoiled NHR. The CHR regions have a high affinity for these grooves. In this way, the CHR regions rearrange and approach to NHR packing themselves in an antiparallel α-helix on the hydrophobic NHR grooves, so that a supercoiled structure in the form of a six-helix bundle (6-HB) is formed, which constitutes the post-fusion structure of gp41 (Figures 1.5.3-B-C). This folding of the CHR regions on the NHR trimeric nucleus is energetically very favorable, which is why it constitutes the driving force necessary to bring the viral and cellular membranes close together, thus facilitating their fusion[82]. The final membrane fusion steps are facilitated by interaction of MPER and TM with the FPPR and FP[99].

The entry mechanism of SARS-CoV-2 is on the whole quite similar to that of HIV but also to other viruses such as the influenza virus or the Ebola virus, which also use class-I fusion proteins[100,101]. Viral fusion proteins are organized as trimers of heterodimers formed by two subunits associated with each other (Figure 1.6.4). The external subunit (S1 in SARS-CoV-2 analogous to gp120 in HIV) is responsible for adhering to the cell membrane through its binding to a specific receptor (ACE2 in SARS-CoV-2 or CD4 in HIV)[102]. Subsequently, an activation process occurs, which in HIV consists in the binding of gp120 to a co-receptor. While in SARS-CoV-2 the activation or priming consists in a proteolytic cut by the transmembrane serine protease TMPRSS2 in its S protein, which causes a large conformational change in the second transmembrane subunit (gp41 or S2 if we refer to HIV or SARS-CoV-2 respectively)[63] (Figure 1.7.2). S2 protrudes its N-terminal end (the fusion peptide) inserting it into the cell membrane and subsequently folding itself bringing the viral and cell membranes closer to each other and promoting their fusion with the opening of a pore through which the genetic material will be inserted causing the infection[103].

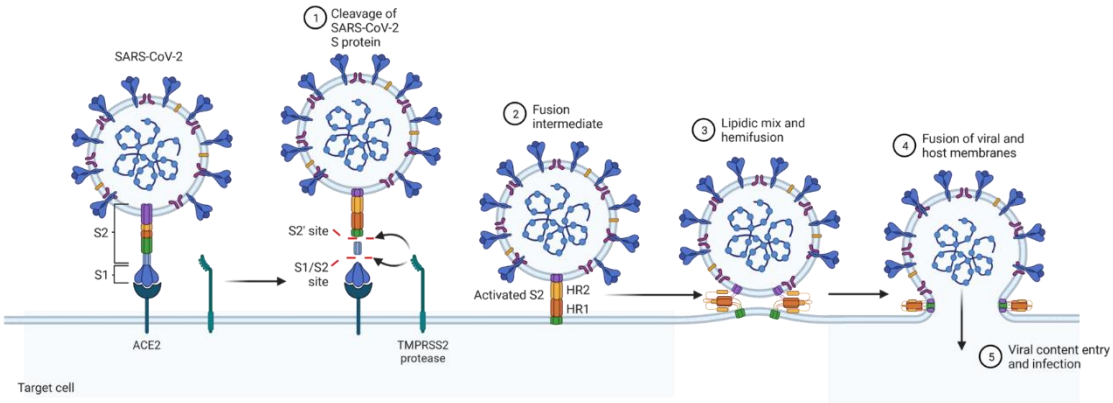


Figure 1.7.2: SARS-coV-2 fusion mechanism. The transition from the prefusion conformation of the Spike is conceptualized as an ensemble of steps through which the post-fusion conformation is accomplished and membrane fusion is induced. See the text for more details. Figure made using BioRender.

In this conformational change, the HR1 and HR2 regions must interact with each other forming a very stable coiled-coiled 6-helix bundle structure (Figure 1.7.3), whose formation is thought to generate the energy needed for membrane fusion[104]. Although both 6-HB structures are very similar, the reduced α-helix content in the HR2 region from SARS-CoV-2 stands out. This does not necessarily make the HR1-HR2 interaction less tight but the absence of a prominent hydrophobic cavity has some implications in the discovery of inhibitors targeting this region as we discuss in following sections.

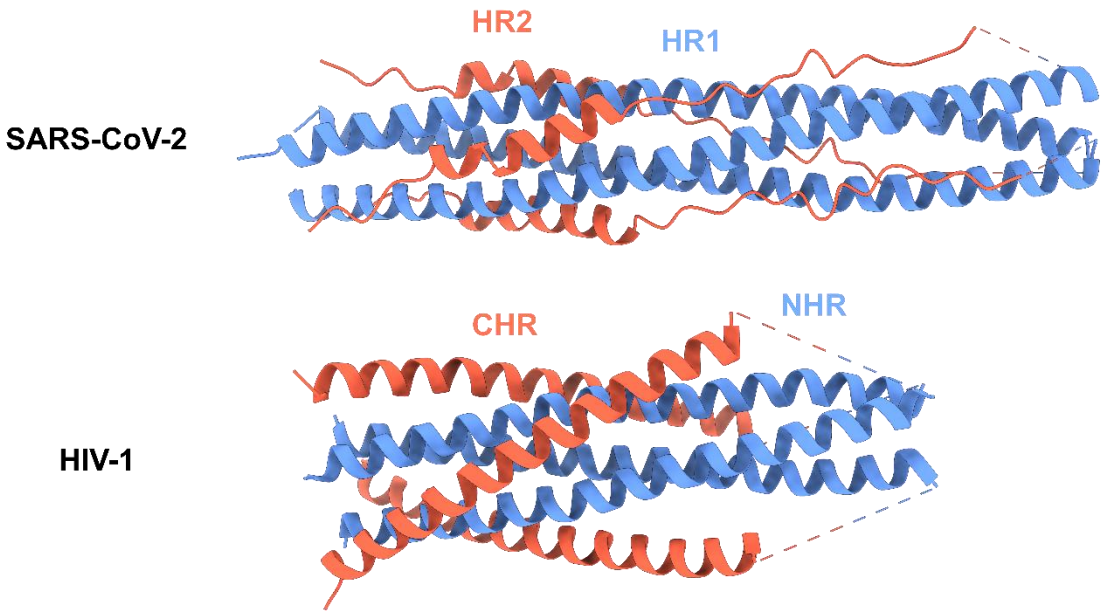


Figure 1.7.3: 6-helix bundle structure of gp41 and S2 proteins (PDB IDs: 1IF3 and 6LXT respectively). HR1 regions are colored in blue and HR2 in red. In both structures the regions connecting HR1/NHR and HR2/CHR (the Central Helix and Connector Domain in SARS-CoV-2 and the Disulfide Loop in HIV) have been removed in shake of clarity and the dashed lines indicated continuity of the chain.

1.7.1 Fusion inhibitors directed against gp41 and S2. State of the Art.

The process of fusing the viral and cellular membranes represents a target for drugs that inhibit both the HIV and SARS-CoV-2 entry process. The main target is the pre-hairpin intermediate structure where both NHR (or HR1) and CHR (or HR2) are exposed and can be accessed by a diversity of molecules, collectively called ‘fusion inhibitors’.

Their mechanism of action lies in their binding to the pre-hairpin intermediate within the viral spike during the fusion process. In this way, they compete with the genuine viral regions, thus, preventing the formation of the actual productive viral 6-HB and precluding membrane fusion and infection[105,106].

Fusion inhibitors may be divided in classes depending on whether they target NHR or CHR regions.

1.7.2 Class-1 fusion inhibitors.

This class of inhibitor acts by binding to the hydrophobic crevice exposed by the trimeric NHR coiled-coil structure. Because auto-association between NHR regions is needed in order to correctly form and expose the hydrophobic crevice, class-1 fusion inhibitors tend to act sometime after the pre-hairpin intermediate conformation is formed and prevent its evolution to the fusion active state[107]. Many approximations have been used to design this kind of inhibitors during the lasts 30 years since the first ones were discovered at the early 1990s[108,109].

CHR-derived peptides are an important type of inhibitors belonging to this class. But it does not restrict to them as many approximations have been developed trying to overcome some of their drawbacks, like starting from a CHR peptide and designing mutations to improve helicity and stability.

To date, T-20 (previously known as DP-178[110]), whose generic and commercial names are Enfuvirtide and Fuzeon, respectively, has been since 2003 the first and only fusion inhibitor approved by the FDA[88,110–112]. T-20 is a peptidic inhibitor whose sequence derives from the CHR and MPER regions of gp41[82]. However, this inhibitor accumulates a series of drawbacks associated with its use, such as:

- As a synthetic peptide, it exhibits a high production cost[88,113].
- It presents low oral bioavailability[104,113].
- It has a short half-life *in vivo*, since it is rapidly degraded by proteolytic enzymes in the blood stream[104].
- Both its short half-life and its lack of oral bioavailability influence its dosage mode: it must be injected subcutaneously twice a day at very high doses (90 mg/day), causing low adherence of patients to treatment [37,104,114].
- Due to the high doses it produces severe local reactions when injected[115].
- The treatment itself generates virus resistance[82].
- Due to all its drawbacks it is only used as salvage therapy for patients with multiresistance to other treatments[37].

Using structure-based design techniques, improved versions of T-20 have been developed, such as sifuvirtide[116], which incorporates the hydrophobic pocket binding domain of gp41. In addition, sifuvirtide harbors regions in its sequence that favor the formation of intra-helical salt bridges, thus improving its stability, half-life and reducing its resistance[112].

Strategies to increase the *in vivo* half-life of peptide inhibitors by synthesizing them using D-amino acids have also been described[117]. Natural peptides are constituted of L-amino acids, thus, artificial D-peptides[118–120] re synthesized from the D-enantiomers of naturally occurring amino acids. Their main advantage resides in that they are highly resistant to degradation, as they cannot fit into the active site of regular proteases because of their

high stereo-specificity. Meanwhile, they adopt the exact same mirror three-dimensional structure of their L-counterparts, thus retaining their activity[121,122].

Other strategies fuse membrane anchoring domains with C-peptides in order to augment their effective concentration closer to their target[123]. Recently, a novel fusion inhibitor called Albuvirtide (ABT), whose structure is derived from the CHR sequence and modified to contain a single 3-maleimimidopropionic acid (MPA) group. This MPA group promotes its irreversible conjugation with serum albumin in vivo allowing for a more extended half-life. ABT exhibited in vitro activity against various HIV-1 strains and certain variants that are resistant to enfuvirtide. The only approved country for ABT commercial use to date, is China, where the access to several other oral medications may be more limited [124–127].

To overcome the difficulties associated with the therapeutic use of CHR-peptides, small-molecule compounds have attracted much attention. Many of those small-molecules were developed trying to replicate a similar mechanism of action to that of CHR-peptides[113]. However, because their target, the hydrophobic groove in NHR, is too large for a small compound, the area of interest was restricted to the deep hydrophobic pocket, a conserved hydrophobic cavity on the surface of the NHR trimer that accommodates the side chains of three conserved hydrophobic residues in the gp41 CHR region (Figure 1.7.1)[128,129].

The most important characteristics of small-molecule fusion inhibitors can be summarized as follows: they should display drug-like properties, such as solubility, permeability, oral bioavailability, lack of toxicity, low molecular weight, etc.[130,131]; they should present both polar and non-polar groups in their structure and at least one must be negatively charged in order to interact with experimentally determined relevant amino acids, e.g., Lys 574 or Arg 579[37,110,132].

Although, small-molecule inhibitors present a relevant potential, none of them has made its way into actual therapeutic drugs for reasons not well understood. Several causes have been proposed, namely, the difficulty for a small-molecule that targets a single binding pocket to compete with a protein-protein interaction with a large interface; the conformational flexibility of the target interface; and the difficulty in faithful representations of the NHR groove with emphasis in the most prominent hot-spot, the hydrophobic pocket[37].

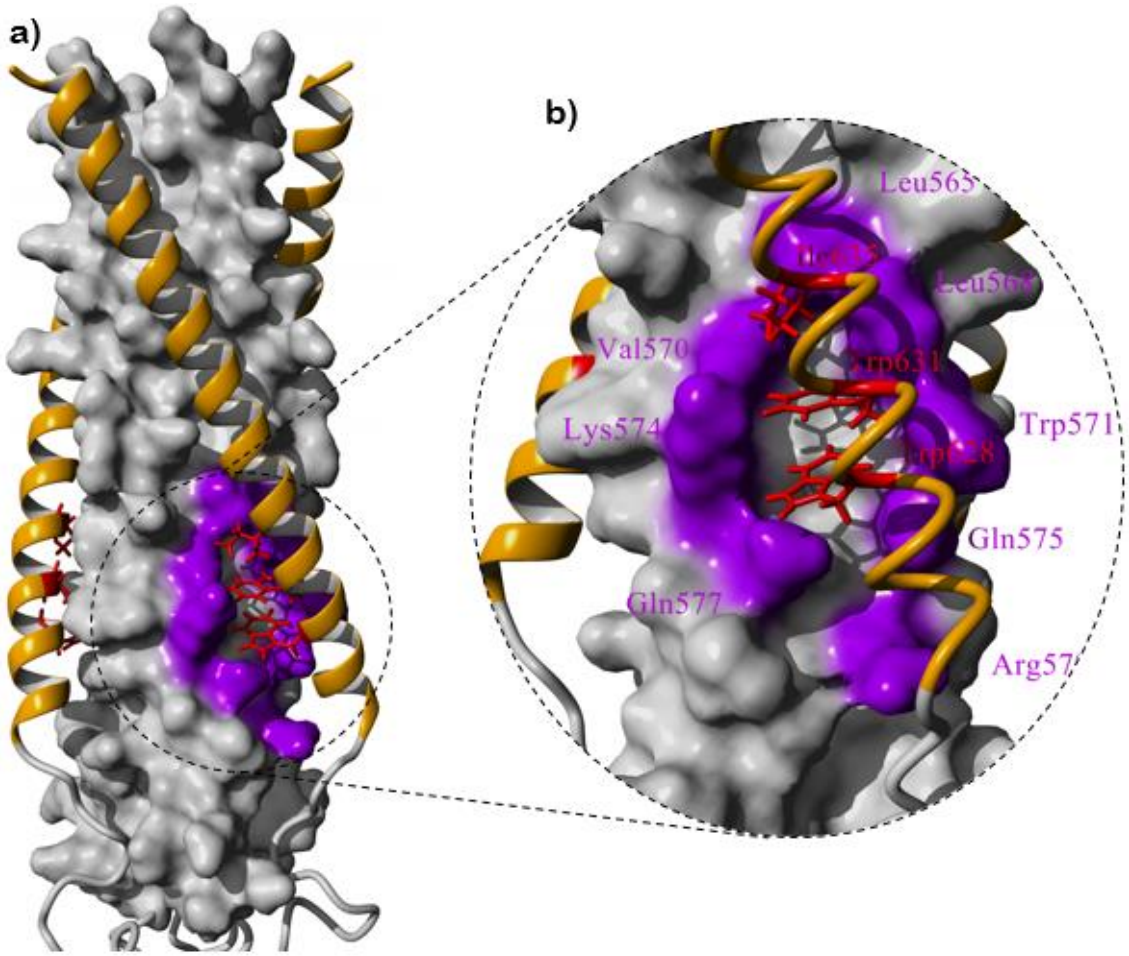


Figure 1.7.1: Hydrophobic pocket (HP) of the NHR region of gp41 in the post-fusion conformation (PDB ID: 1IF3). a) The molecular surface of the trimer of NHR coiled-coil regions is shown in grey. The CHR regions are shown in yellow ribbons. Conserved hydrophobic residues that interact with HP residues in the 6-HB are shown in red and in sticks. Residues delimiting the HP are colored in purple. b) Close-up view of the HP interaction. Same color scheme as in a) but the CHR backbone is represented as smooth tubes in shake of clarity and residues lining the HP are identified with purple labels.

Besides peptidomimetic and small compound therapeutics, antibodies have also been developed in an effort to try to neutralize viral infection by passive immunization. Because the NHR region is not accessible at its native state, NHR-peptides and constructs mimicking a trimeric NHR bundle had to be used in order to select a human antibodies, such as D5[133], whose epitope is the hydrophobic pocket reaffirming its role as a hot spot in viral inhibition[134,135].

1.7.3 Class-2 fusion inhibitors.

Class-2 inhibitors target the CHR region of gp41 and are based on the same principle as those that bind to the NHR region: preventing fusion of the viral and cellular membrane by blocking the formation of the productive 6-HB[98]. They also target the prefusion intermediate but are not only restricted to that particular stage, as they are directed against

the CHR regions, which may constitute a target at some point after priming is achieved and before the formation of the pre-hairpin intermediate[136].

Compared to CHR peptides, NHR peptides and peptidomimetics have a generally lower anti-HIV-1 activity, being able to inhibit the fusion process only in the micromolar range, in contrast to the nanomolar range of activity exhibited by the former. This limited potency may be due to the low solubility and aggregation tendency of NHR peptides in solution. However, they have certain advantages, such as their activity against strains resistant to CHR-inhibitors[117,137]. The first N-peptide synthesized, DP-107 (residues 558–595 from gp41; see Figure 1.6.3 for reference), displayed a highly α-helical secondary structure compatible with the acquisition of coiled-coil structure expected for this amino acid sequence and was able to block virus mediated cell-cell fusion and cell infection by viruses[138]. Soon after, Peter Kim's group identified several other N-peptides: N51 (residues 540–590), N36 (residues 546–581) and N34 (residues 546-579), moderately able to inhibit viral fusion. In order to reduce N-peptides aggregation and enhance their helical structure improving their activity, chimeric peptides were designed by linking those peptides with more stable and autonomous folding units. Examples of this approach are IQN17 and its derivatives IQN23 and IQN36 and IZN17 together with IZN23 and IZN36. They are all chimeric peptides, in which N-peptides of different length were fused with trimerization domains such as GCN4 or IZ_m to stabilize the trimeric structure of the molecule[120,139]. At the same time, Peter S. Kim and M. J. Root's group developed a protein construct named 5-helix, which is in fact an inhibitor that presents both NHR and CHR sequences covalently linked through small loops. Specifically, it presents 3 NHR regions and only 2 CHR regions, an arrangement that facilitates its folding, forming a structure similar to 6-HB but lacking one CHR region, which instead leaves an NHR groove exposed. 5-helix can bind to the CHR region of gp41 and in this way prevent the formation of 6-HB. This stabilization strategy accounts for the increment in inhibitory activity achieved by these type of constructs which is in the low nanomolar range, representing up to 3 orders of magnitude higher than the N-peptides without modification[117]. A similar strategy has also been very recently implemented for the design of SARS-CoV-2 inhibitors by Shibo Jiang's group but with lower inhibitory potency compared to HIV[140].

Marius Clore's group used another strategy to design a trimeric coiled-coil comprising exposed NHR regions by recombinantly fusing them to a small 6-HB structure (composed of N34 and C26 peptides) further stabilized by engineered intersubunit disulfide bonds. This construct, named NCCG-gp41, also exhibits *in vitro* anti-HIV activity in the low nanomolar range[141]. P. Tien's group designed a chimeric linear polypeptide called HR_{121} containing consecutive HR1, HR2 and HR1 regions tethered by flexible amino acid linkers, so that the first HR1 and the HR2 sequences are able to interact and associate exposing an HR1 region. Three HR_{121} constructs can trimerize, thus exposing a trivalent HR1 region. Another chimeric polypeptide was called HR_{212} and it was designed following the same principle but exposing a trivalent HR2 region, and it would thus fall into the class-1 inhibitor's category[142,143].

A different strategy worth noting is inhibiting HIV fusion by precluding productive NHR coiled-coil homotrimer formation. This was firstly described also by M. Clore's group by designing a modified N36 peptide, named $N36^{Mut(e,g)}$ with mutations in the e and g heptad-

repeat position so that it could not interact with C-peptides. It could, nevertheless, auto-associate forming trimers in solution and *in vitro* acts by disrupting the homotrimeric coiled-coil of N-terminal helices in the pre-hairpin intermediate to form non-productive heterotrimers[107].

1.8 CovNHR proteins.

An innovative strategy, developed by our group, takes advantage of the versatile structures that coiled-coil motifs exhibit to design small and stable protein constructs that mimic the surface of NHR in gp41 and therefore fall in the class-2 of inhibitors. Those protein constructs are called covNHR and their rational design stands out for the conversion of the topology from the trimeric parallel arrangement of NHR chains of gp41 into a single polypeptide chain that folds its helices in an antiparallel manner into a helix-loop-helix-loop-helix conformation. See Figure 1.6.1 for a schematic view of the molecule's topology[114].

These protein constructs spontaneously fold as predicted, strongly interact with CHR peptides and exhibit strong antiviral activity (IC_{50} in the low nanomolar range) against a wide variety of HIV-1 strains. Like other class-2 inhibitors described above, the inhibition mechanism of these protein constructs resides in the similarity that they present with the structure of an exposed hydrophobic groove in the fusion intermediate conformation of NHR, in such a way that they expose binding determinants that exhibit great binding affinity for CHR peptides, thus competing with the NHR domain of gp41 and preventing the formation of 6-HB[114,144,145].

The covNHR proteins are easy to produce recombinantly by *E. coli* overexpression with high yields. Because they fold spontaneously in their expected structure without the need of any addition of exogenous foldons or posttranslational modifications, they have a relatively small size and have therefore less steric hindrances. They are also very stable and soluble at physiological conditions and expose a mono-valent NHR groove, which makes easier the analysis of their interactions with CHR.

These characteristics make the covNHR proteins particularly attractive as fusion inhibitors, especially compared to class-I fusion inhibitors. However, some of them may still be improved in terms of conformational stability to achieve even higher inhibitory activities.

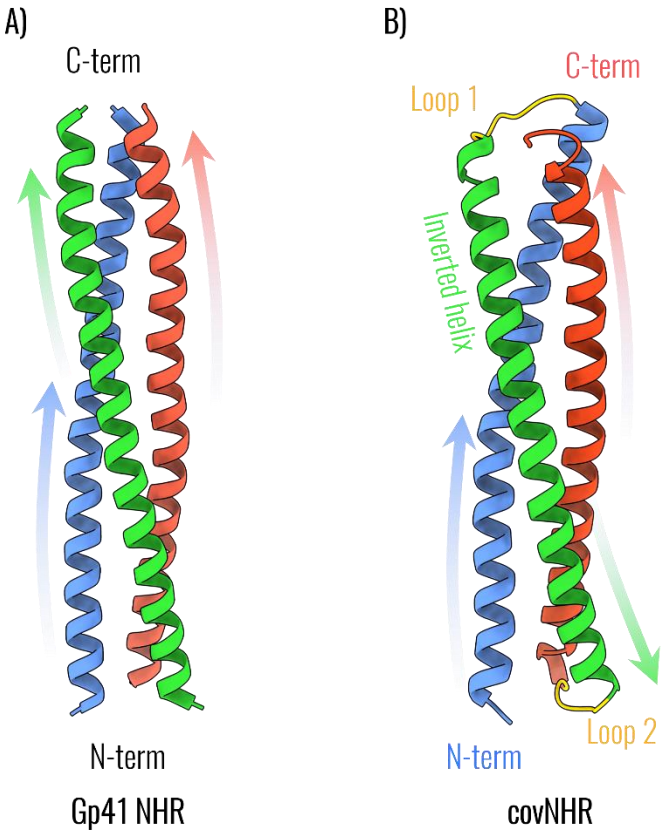


Figure 1.8.1: Comparison of gp41 and covNHR topologies. Both proteins are displayed in ribbons and arrows indicate the sense of the polypeptide chain from the N-terminus to the C-terminus of the respective protein. The inverted helix is colored in green in panel B. Loops are colored in yellow.

We previously determined the crystallographic structure of a covNHR protein in free form (PDB ID: 4R61)[114] and in complex with the C34 peptide (PDB ID: 6R2G, Figure 1.8.2) [144]. Taking advantage of the high degree of structural similarity (with an RMSD <1 Å) between the covNHR structures and gp41 in its post-fusion structure (PDB ID: 1IF3), we have previously used them to characterize thermodynamically the binding interface with CHR peptides which consists mainly of the hydrophobic groove itself and the residues adjacent to it[144,145]. An extensive calorimetric characterization of the binding energetics between covNHR and CHR peptides of different lengths allowed us to define a detailed map of gp41's NHR surface corresponding to different sub-pockets that conform the hydrophobic groove (Figure 1.8.2). The most prominent one is the above describe HP. Another important site of interaction occurs at an N-terminal pocket (NTP), where polar CHR and NHR residues establish a network of polar interactions and hydrogen bonds mediated by two interfacial water molecules[145]. Mutations at this region disrupt this interaction network and are related to the emergence of resistance to CHR peptides. Between these two pockets, shallow middle pocket (MP), on which several aliphatic residues interact. Finally, a smaller C-terminal pocket (CTP), where additional CHR residues upstream the HP-binding motif interact with NHR.

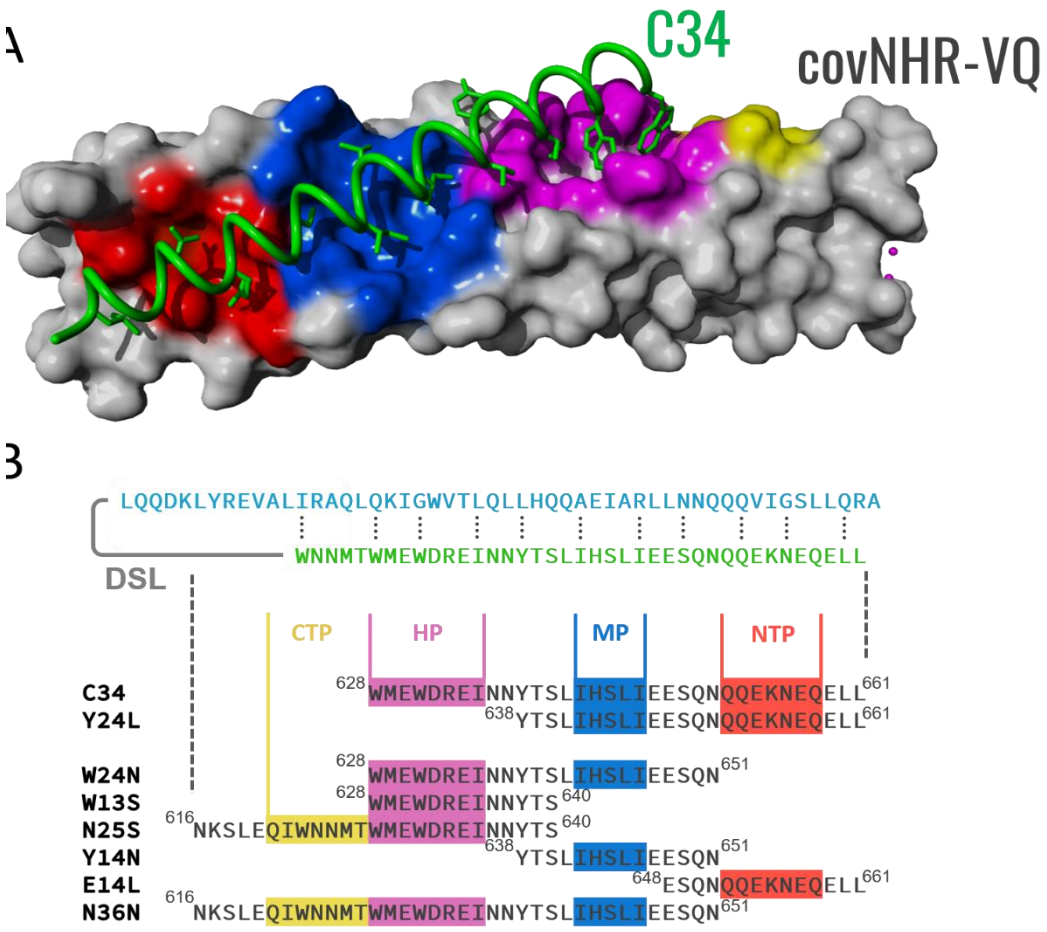


Figure 1.8.2: CovNHR mapping of the different sub-pockets present in its surface. A) Complex between covNHR and C34 peptide. CovNHR is shown in molecular surface and each pocket is colored differently according to B). C34 is shown in tubes and colored in green with the interacting amino acids shown in sticks. B) Sequences of the different peptides used in previous studies and in this thesis. NHR sequence is colored in light blue, followed by a schematic representation of the DSL region and the sequence of the CHR region colored in light green. CHR derived peptide sequences are shown next to each other's name and each sub-pocket is colored and labeled with its name initials, see text for more details. Figure modified from[145].

CHR-peptides targeting only one sub-pocket showed little or no binding interaction. On the other hand, peptides encompassing two or more sub-pockets exhibited submicromolar affinities, evidencing that inter-pocket cooperativity was crucial in high-affinity covNHR-CHR peptide complex formation. Thus, the middle pocket (MP) plays a crucial role to achieve tight binding compared with only targeting the hydrophobic pocket, which turns out to be insufficient by itself to achieve significant affinity[145].

This capability to bind CHR peptides is the base of the high inhibition activity of covNHR proteins, in the low-nM range. However, covNHR and an NTP-abolished mutant (named covNHR-ER which stands for its two mutations V10E and Q123R) differ largely on their affinity towards C34 (up to 5 orders of magnitude, 0.08 pM for covNHR and 15 nM for covNHR-ER) while exhibiting relatively similar IC_{50} values against virus infection in the low-nM range[144]. This points at an issue formerly described by M. J. Root and coworkers, by which the binding affinity of several NHR inhibitors is only correlated to a certain degree with their respective inhibitory capacity, meaning that higher affinity values are only translated into lower IC_{50} values until low-nM K_d values are reached. Further improvements in K_d values below nM do not result in reductions in IC_{50} below the low-nM range. They explained this fact with a model that takes into account that the transient and limited exposure of the target region, such as gp41 CHR region, may kinetically limit its interaction with inhibitors such as 5-helix (or covNHR here). In this model, CHR, which is not fully exposed in native Env, (Figure 1.6.2) should undergo conformational changes that lead it to a transiently more exposed conformation in the pre-hairpin intermediate, to become accessible to NHR-derived inhibitors. Further evolution of gp41 to the post-fusion 6-HB would occlude the exposed CHR, thus kinetically limiting the inhibitory activity. In this binding and inhibition process, the steric hindrances caused by inhibitors' size may also be taken in account, such as the relatively high molecular size of 5-helix[146,147].

To explore the possibility of targeting CHR with smaller, less sterically-hindered protein constructs, together with the opportunity of studying the capacity of isolated pairs of sub-pockets to bind CHR peptides and inhibit HIV in vitro, led us to design and produce two different mini-protein covNHR variants[148]. Each of the mini-proteins contained a pair of pockets, as only one had previously been seen as not capable enough to bind peptide counterparts. Namely, covNHR-N contained the NTP and MP pockets; and covNHR-C contained the HP and CTP pockets. Despite both constructs were able to bind their respective complementary CHR peptides with similar affinities, only covNHR-N exhibited inhibition against HIV-1 infection. However, due to its low intrinsic stability it needed to be stabilized with an intra-chain disulfide bond[148]. This modification translated into a higher inhibition activity although binding affinity towards its CHR counter peptide remained unaltered. These results suggested a relationship between conformational stability and inhibitory potency. Moreover, they suggested a special vulnerability of the flexible C-terminal end of the gp41 CHR region. Nevertheless, these findings encouraged us to further improve these and other constructs to try to achieve even higher inhibition activities.

All this research shows that the covNHR molecules constitute an excellent representation of a functionally exposed NHR groove that may not only serve as a great tool to study gp41's post-fusion structure and the interactions established with complementary peptides but also as suitable molecular templates to use in the search of new small-molecule therapeutics.

2. Hypothesis and Objectives.

The development of effective therapeutics against HIV and SARS-CoV-2 fusion and infection is regarded as one of the main goals for treatment. However, gp41 and S2 proteins constitute elusive subjects of study because they undergo several dramatic conformational changes through the fusion process, transiently exposing different functionally relevant structures. In addition, their intrinsic properties make very challenging the *in vitro* study of their interactions and function. This is predominantly due to the fact that they are aggregation-prone transmembrane proteins with high structural complexity, as they are divided in several sub-domains with different activities establishing a variety of interactions at distinct steps during the fusion process. Therefore, it is of vital importance to have simplified models, such as the covNHR proteins, that imitate functionally relevant proteins' states, in this particular case the exposed NHR region representative of the fusion intermediate state.

On the light of all the background summarized above, this Thesis is based upon the following hypotheses:

Hypothesis 1:

The discovery and development of small-molecule HIV inhibitors directed to the NHR region and particularly to the HP has been especially elusive for reasons still not well understood. A possible reason is the relatively large size of the HP compared to this kind of inhibitors, precluding the achievement of sufficient affinity to compete against CHR peptides for the interaction with the HP. A feasible inhibitor may need additional interactions with the HP and possibly with adjacent pockets to achieve potent activity. Therefore, additional research is needed for a more complete understanding of the conformational and dynamic properties of the HP and its relationships with other gp41 pockets. Since the covNHR proteins mimic accurately the structure of the exposed NHR region of the gp41 fusion intermediate, and particularly a fully exposed HP, they may serve as appropriate target templates for a detailed structural and thermodynamic study of small-molecule HP ligands that may guide the identification and validation new HIV fusion inhibitors.

Traditionally, the search of small-molecule compounds has been restrained to the HP in the context of the post-fusion structure, but the opportunity that the covNHR molecules offer goes even further, as they expose the whole NHR hydrophobic groove. Studying the linkage established between small HP-binding ligands and other pocket's ligands, such as peptides, may shed light into their allosteric communication and how it influences their binding.

8-Anilinonaphtalene-1-sulfonic acid (ANS) is a small-molecule widely known for its use as a fluorescent probe that can bind to solvent-exposed hydrophobic patches in proteins. ANS could be a suitable HP ligand because its drug-like physicochemical properties

exhibited by other gp41 inhibitors described so far, namely, its low molecular weight, an amphiphilic structure and the presence of an anionic group.

Hypothesis 2:

Besides their usefulness as target model templates, the covNHR proteins constitute themselves powerful HIV inhibitors because of their capacity to bind CHR and interfere with fusion. However, to further improve the inhibitory activity of the covNHR molecules it is necessary to deeply understand the structural and thermodynamic determinants of the NHR-CHR interaction. In particular, it is of interest understanding how the internal coiled-coil stability of the covNHR proteins influences their interaction with CHR, as well as the aforementioned inter-pocket allosteric communication. An interesting feature of the NHR coiled-coil is the presence of several non-canonical buried polar residues, which are thought to be responsible for specificity and structural uniqueness at the expense of stability. Moreover, we hypothesized that buried polar residues may also be involved in the mechanisms by which the NHR–CHR interaction is regulated. Accordingly, introducing polar-to-hydrophobic mutations in the covNHR proteins may allow us to study the molecular and energetic role of those polar buried residues in the NHR-CHR interaction, as well as in their resulting inhibitory activity.

Hypothesis 3:

We previously showed that HIV could be inhibited using covNHR miniproteins encompassing only the NTP and MP NHR pockets (called covNHR-N as it mimics the N-terminal subdomain of the NHR region) and that the HP is not strictly required. However, the intrinsic low stability of these miniproteins could be influencing their relatively low inhibitory potency. CovNHR-N was stabilized by engineering an intra-chain disulfide bond showing a moderate improvement in inhibitory potency against different HIV strains. Thus, we hypothesized that a further increase in stability could lead to an improvement of the affinity to its CHR peptide counterparts ultimately resulting in a further improved anti-HIV inhibitory activity. This conformational stabilization may be achieved by designing a second disulfide bond tethering the N-terminal region of the miniprotein.

Hypothesis 4:

As described in the Introduction, SARS-CoV-2 and HIV, being enveloped viruses with class I fusion machines, exhibit a similar fusion mechanism. Gp41 and S2 proteins share a high degree of structural similarity between their post-fusion structures, Thus, we hypothesized that similar single-chain chimeric proteins could be designed to mimic an exposed SARS-CoV-2 HR1 trimer and that these mimetic proteins could also have antiviral activity. Likewise, those constructs may also serve as useful tools to explore the details behind the molecular interactions responsible for the completion of the viral fusion

mechanism, potentially yielding information about the infection biology and biochemistry of SARS-CoV-2 at the molecular level.

This Thesis is presented as a collection of four chapters, each corresponding to a published article (one of them submitted to publication) that covers each of the four objectives as follows:

1. To investigate the structural, thermodynamic and dynamic properties of the interactions established at the hydrophobic pocket by ANS, a suitable fluorescent probe with drug-like properties, and analyze how CHR peptides influence the binding of ANS to the HP to discern if allosteric communication exists between NHR pockets.

2. To study the role of buried polar residues in the interior of the NHR coiled-coil and identify how they affect, not only conformational stability, but also NHR-CHR binding, as well as inhibitory activity.

3. To deepen our understanding of the effect of conformational stability on CHR binding and HIV inhibitory potency of covNHR miniproteins engineered to further improve their stability by intramolecular disulfide bonds.

4. To draw on the experience and previous results acquired with covNHR proteins to design, produce and characterize functional mimetic proteins of the HR1 region in the S2 SARS-CoV-2 protein so that they can bind with effective affinities to the complementary HR2 region and in that way inhibit the fusion process of the virus with the target cell.

3. Results.

The results in this Thesis are presented in this chapter as a compendium of publications.

Publication 1:

Cano-Muñoz, M., Jurado, S., Morel, B., & Conejero-Lara, F. (2021). Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors. International Journal of Biological Macromolecules, 192, 90-99.

Impact Factor (2020): 6.953

Category name / Quartile in Category: Polymer Science / D1

Publication 2:

Cano-Muñoz, M., Cesaro, S., Morel, B., Lucas, J., Moog, C., & Conejero-Lara, F. (2021). Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity. Biomolecules, 11(4), 566.

Impact Factor (2020): 4.879

Category name / Quartile in Category: Biochemistry & Molecular Biology / Q2

Publication 3:

Cano-Muñoz, M., Lucas, J., Lin, L. Y., Cesaro, S., Moog, C., & Conejero-Lara, F. (2022). Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion. International Journal of Molecular Sciences, 23(5), 2794.

Impact Factor (2020): 5.924

Category name / Quartile in Category: Biochemistry & Molecular Biology / Q1

Publication 4:

Cano-Muñoz, M., Polo-Megías, D., Cámara-Artigas, A., Gavira, J. A., López-Rodríguez, M. J., Laumond, G., Schmidt, S., Demiselle, J., Bahram, S., Moog, C. & Conejero-Lara, F. Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity (submitted for publication to P.N.A.S. USA).

Publication I

Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors.

International Journal of Biological Macromolecules 192 (2021) 90–99

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors

Mario Cano-Muñoz [*], Samuel Jurado [1], Bertrand Morel [2], Francisco Conejero-Lara [*]

Departamento de Química Física, Instituto de Biotecnología y Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain

ARTICLE INFO

Keywords:
Binding cooperativity
Coiled-coil
Allosterism
Calorimetry
Antiviral therapy

ABSTRACT

During HIV-1 infection, the envelope glycoprotein subunit gp41 folds into a six-helix bundle structure (6HB) formed by the interaction between its N-terminal (NHR) and C-terminal (CHR) heptad-repeats, promoting viral and cell membranes fusion. A highly preserved, hydrophobic pocket (HP) on the NHR surface is crucial in 6HB formation and, therefore, HP-binding compounds constitute promising therapeutics against HIV-1. Here, we investigated the conformational and dynamic properties of the HP using a rationally designed single-chain protein (named covNHR) that mimics the gp41 NHR structure. We found that the fluorescent dye 8-anilino-naphtalene-1-sulfonic acid (ANS) binds specifically to the HP, suggesting that ANS derivatives may constitute lead compounds to inhibit 6HB formation. ANS shows different binding modes to the HP, depending on the occupancy of other NHR pockets. Moreover, in presence of a CHR peptide bound to the N-terminal pockets in gp41, two ANS molecules can occupy the HP showing cooperative behavior. This binding mode was assessed using molecular docking and molecular dynamics simulations. The results show that the HP is conformationally flexible and connected allosterically to other NHR regions, which strongly influence the binding of potential ligands. These findings could guide the development of small-molecule HIV-1 inhibitors targeting the HP.

1. Introduction

Enveloped viruses with class-I fusion type machines, such as the Human Immunodeficiency Virus (HIV), Ebola Virus (EboV), Human Respiratory Syncytial Virus (HRSV), Influenza virus, and coronaviruses like Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) or the recently emerged SARS-CoV-2, currently constitute an undoubtedly major threat to humanity (public health and economic issue) [1–3]. Although the major focus is currently set on the dramatic Coronavirus Disease 2019 (COVID-19) pandemic, we cannot forget that the HIV pandemic is still very active and continues to be one of the world's largest pandemics with near than 1 million deaths annually [4].

Regarding HIV, the causative agent of Acquired Immune Deficiency Syndrome (AIDS), modern Highly Active Antiretroviral Therapies (HAART) have been developed and have helped reducing the mortality and morbidity associated with the infection. Nevertheless, the emergence of multi-resistant HIV variants to the current treatment and the appearance of side effects and drug-drug interactions promote the necessity of continuing to discover and produce new inhibitors directed towards different stages in the virus' life cycle [5,6].

The entry of virion particles into the host cell by fusing both membranes is a key step in the infection of HIV, thus, it has long been considered as an attractive target for the development of new anti-HIV drugs [1]. Conceivably, inhibition of the fusion process prevents the infection by an active viral particle and thus, investigation in this respect may yield compounds that can potentially serve as prophylactics and antivirals.

The fusion process is mediated by the envelope glycoprotein (Env), a non-covalently associated trimer of heterodimers composed of two glycoprotein subunits, gp120 and gp41 [7]. The binding of HIV-1 gp120 to CD4 receptor and a co-receptor, CCR5 or CXCR4, on surface of the target cell, initiates the fusion process. This event in gp41 triggers a

Abbreviations: ANS, 8-anilino-naphtalene-1-sulfonic acid; DSC, differential scanning calorimetry; ITC, isothermal titration calorimetry; CD, circular dichroism; DLS, dynamic light scattering; SLS, static light scattering; NHR, N-terminal heptad repeat; CHR, C-terminal heptad repeat; HP, gp41 hydrophobic pocket.

* Corresponding authors.
E-mail addresses: mariocano@ugr.es (M. Cano-Muñoz), conejero@ugr.es (F. Conejero-Lara).
[1] Present address: Diater Laboratorio farmacéutico, Av. Peces Barba n° 2, Parque tecnológico Leganés, 28918, Spain.
[2] Present address: Angany Innovation, 1 voie de l'innovation, Pharmaparc II, 27100 Val de Reuil, France.

https://doi.org/10.1016/j.ijbiomac.2021.09.198
Received 3 August 2021; Received in revised form 24 September 2021; Accepted 29 September 2021
Available online 5 October 2021

series of large conformational changes, in which three N-terminal heptad repeat (NHR) regions form an inner helical coiled-coil core trimer and three C-terminal heptad repeat (CHR) regions associate externally with the NHRs in an antiparallel fashion, ultimately adopting a six-helix bundle conformation (6HB) [8,9]. This energetically favorable interaction between NHR and CHR brings viral and host-cell membranes into close proximity promoting fusion and eventually causing infection. Consequently, compounds that interact with either CHR or NHR may interfere with this key process and thereby constitute HIV fusion inhibitors [10].

Traditionally, peptides with sequence derived from those regions have been developed as promising inhibitors. However, the CHR-derived peptide T20 is still the only FDA approved fusion inhibitor [11]. This is in part due to the drawbacks associated with the use of peptide-based compounds, such as its lack of oral bioavailability and short half-life in vivo due to rapid proteolytic cleavage and renal filtration, which forces it to be administered by subcutaneous injection twice daily in very high dosage. These problems, together with its high cost of production, make T20 only used as salvage therapy [12]. Therefore, the identification of new small-molecule compounds targeting gp41 with high oral bioavailability and lower production costs is urgently needed.

On this basis, small-molecule compounds still constitute promising therapeutics agents. This search for highly active compounds could be made either by optimizing existent inhibitors or by finding new pharmacophore structures not evaluated before, or even following both strategies, i.e., finding new structures to further optimize them.

In the gp41 6HB structure, a deep hydrophobic pocket (HP) on the NHR surface has proven to be of high importance in the NHR-CHR interaction [9] and its occupancy interferes with 6HB formation and thereby with membrane fusion [13]. Although small-molecule based strategies targeting gp41 have been widely explored in the past and a large number of small-molecule ligands have been described [10,14,15], such as NB-2 and NB-64 including its derivatives [13] or indole-based compounds [16], the compounds identified to date exhibit limited in vitro potency and their affinity towards their target is usually moderate. Many of these strategies have made use of stabilized chimeric variants of the NHR region of gp41, such as IZN17 [17] and 5-helix [18], as templates exposing the HP in order to sieve compound libraries (following both virtual and experimental approaches) [19]. Nevertheless, they have not achieved the expected success for reasons that remain not well understood. A major challenge is to find a model that can truly mimic the dynamic conformation of the HP in the prefusion intermediate of gp41. It has also been pointed out that the HP may be too big or too deep for small-molecule inhibitors to achieve a sufficiently high affinity that can compete with the internal CHR interaction, and that a potent inhibitor may need additional interactions outside the HP to acquire potent activity [15]. For these reasons, a more complete understanding of the conformational and dynamic properties of the HP and its relationships with other gp41 pockets would shed light on these issues and guide the discovery of new and more active molecules.

Recently, our group has designed, produced and characterized single-chain chimeric protein constructs (named covNHR, Fig. 1) that accurately mimic the structure of NHR in gp41 [20,21]. In contrast to gp41, which is a transmembrane protein, highly insoluble in aqueous solvents at physiological pH [22], covNHR is a very stable, highly soluble and easy-to-produce recombinantly in *E. coli* strains. The solved crystallographic structure of covNHR in complex with the CHR-derived peptide C34 (PDB ID: 6R2G) shows a virtually identical binding interface to that present in other gp41 constructs in post-fusion conformation [21]. Thus, it constitutes an interesting target to be employed as a template for the search of small-molecule compounds directed against gp41.

8-Anilinonaphtalene-1-sulfonic acid (ANS) is a well-known fluorescent probe that binds to solvent-exposed hydrophobic patches in proteins and is widely used to characterize exposed hydrophobicity and

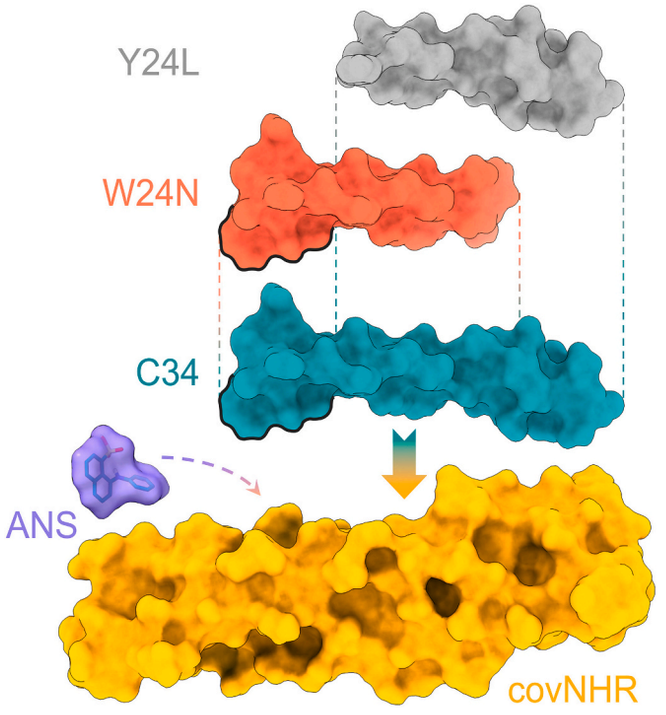


Fig. 1. Overall layout of the molecules involved in this study. The structure of the protein and peptides involved is taken from the crystal structure of the complex between covNHR and C34 peptide (PDB ID: 6R2G), depicted in yellow and cyan surface respectively. The structure of W24N and Y24L peptides, depicted in red and gray surfaces, is derived from that of C34 in the complex structure. The pocket-binding domain is highlighted in black lines for C34 and W24N peptides. ANS molecule is represented in sticks and in purple semi-transparent surface. CovNHR is a highly accurate mimic of the trimeric NHR coiled-coil, with the advantage of exposing only one single interface for CHR binding, see details in the text.

partial unfolding or misfolding in proteins [23]. ANS exhibits drug-like physicochemical properties [24] present in many other small-molecule gp41 inhibitors inspected to date, such as a low molecular weight (<500 Da), an amphiphilic structure and the presence of a negative charge at biological pH [12,16]. Due to these special properties, ANS could constitute an excellent probe to investigate the conformational properties and interactions at the HP.

In this study, we investigated the binding of ANS with the covNHR protein in the presence and absence of interactions with CHR-derived peptides (Fig. 1). We found that ANS can bind specifically to the HP but its binding mode is strongly modulated allosterically by other distant NHR-CHR interactions, which are relevant in the gp41-mediated fusion process. Our results confirm a previously hypothesized allosteric communication between different NHR pockets and may help in guiding the identification of new small-molecule inhibitors targeting the gp41 NHR region.

2. Materials and methods

2.1. Protein and ligands samples

The DNA encoding the covNHR protein sequences was synthesized by Thermofisher Scientific (Waltham, USA). In order to facilitate purification by Ni-Sepharose affinity chromatography, the protein sequences were histidine tagged at the C-terminus. CovNHR proteins were produced by overexpression in *E. coli*, as described previously [20]. Synthetic CHR peptides, both *N*-acetylated and C-amidated, were acquired from Genecust (Luxembourg), with a purity >95%. Protein and peptide concentrations were determined by UV absorption measurements at 280 nm with extinction coefficients calculated theoretically according to

their amino acid sequences with the ExPasy ProtParam server (https://web.expasy.org/protparam/) [25]. 8-Anilino-naphtalene-1-sulfonic acid (ANS) was purchased from Honeywell Fluka (Sigma-Aldrich, St. Louis, Missouri, United States). The concentration of ANS was quantified by UV–visible spectrophotometry using a molar extinction coefficient of 7800 $M^{-1}{\cdot}cm^{-1}$ at 370 nm [26].

2.2. Fluorescence measurements

Fluorescence spectra were measured on a Varian Cary Eclipse spectrofluorimeter (Mulgrave, Victoria, Australia) equipped with a four-cell block temperature controller using a 3 mm path length quartz cuvette. ANS binding to the protein constructs was monitored by measuring the fluorescence emission spectra between 400 and 600 nm following excitation at 370 nm at 25 °C. Slit widths were set to 5 nm for both the excitation and emission monochromators. The resulting spectra were the average of 5 individual scans.

All measurements were performed in 50 mM sodium phosphate buffer at pH 7.4 at 25 °C with a protein concentration of 10 μM and samples were equilibrated for 5 min prior to measurement. No alteration in pH of the protein or complex solutions was observed in the presence of up to 800 μM ANS. For the titration experiments with ANS the total protein and peptide (if present) concentrations were kept constant whereas the ANS concentration was varied between 0 and 250 μM. ANS titration experiments in presence of CHR-derived peptides were carried out at a fixed 1:2 molar ratio between the protein and the peptide; Peptide titrations in presence of ANS were performed at a 1:3 molar ratio (protein:ANS), unless stated otherwise. Due to the relatively high absorbance of ANS reached in the titrations, the fluorescence emission was corrected from the inner filter effect as described elsewhere [27].

2.3. Circular dichroism

Circular dichroism (CD) experiments were conducted on a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) equipped with Peltier-thermostatic cell holder using a bandwidth of 1 nm, a scan rate of 100 nm/min and a response time of 1 s. Near-UV CD spectra (450–250 nm) were measured at a protein concentration of ~40 μM using a 5 mm quartz cuvette and are the result of the average of 20 individual scans. The interaction experiments were carried at a 1:2 molar ratio between the protein and each CHR-derived peptide and at a 1:3 molar ratio between the protein and ANS ligand. Each spectrum was corrected by baseline subtraction using the blank spectrum obtained with the buffer and finally the CD signal was normalized to molar ellipticity ([θ], in $deg{\cdot}dmol^{-1}{\cdot}cm^{2}$).

2.4. Isothermal titration calorimetry

ITC experiments were performed in a Microcal VP-ITC calorimeter (Malvern Instruments, Worcestershire, UK). The protein solutions were typically titrated with 25–30 injections of 5 μL or 10 μL ANS solution using an interval of 480 s, the stirring rate was 300 rpm and the reference power for the baseline was set to 10 μcal/s. Concentrations of the protein used for the titrations were 10 μM to 40 μM, while ANS in the syringe was between 450 μM and 1000 μM. When the CHR-derived peptide was present it was added at a ratio of 1:2 between the protein and the peptide to reach saturation. All experiments were carried out in 50 mM phosphate buffer, pH 7.4 at 25 °C.

The experimental thermograms were baseline corrected and the peaks were integrated to determine the heats produced by each ligand injection. As a blank, an independent experiment with only buffer in the calorimeter's cell was performed with the same ANS solution to determine the corresponding heats of dilution. Finally, each heat was normalized per mole of added ligand. The resulting binding isotherms were fitted using different binding models, as described in the Results. Additional details of the binding models are provided in the Supplementary Text S1.

2.5. Dynamic and static light scattering

The apparent hydrodynamic radii of the protein and protein-ligand complexes were measured using dynamic light scattering (DLS) in a DynaPro MS-X DLS instrument (Wyatt, Santa Barbara, CA, USA). Dynamics v6 software (Wyatt Technology Corporation, Santa Barbara, CA, USA) was used in data collection and processing. Sets of DLS data were measured at 25 °C with an average number of 50 acquisitions and an acquisition time of 10 s. Measurements were carried out in 50 mM sodium phosphate buffer pH 7.4. Static scattering intensities were also measured at different concentrations of protein in a range of 0.2 to 3.5 mg mL^{-1}. The intensities were analyzed using the Debye plot as represented by Eq. (1),

$$Kc/R_{90} = 1/M_w + 2A_2c \quad (1)$$

valid for particles significantly smaller than the wavelength of the incident radiation, where K is an optical constant of the instrument, c is the particle mass concentration, R_{90} is the Rayleigh ratio of scattered to incident light intensity, M_w is the weight-averaged molar mass, and A_2 is the 2nd virial coefficient that is representative of inter-particle interaction strength. M_w can be determined from the intercept of the plot.

2.6. Molecular modeling and docking studies

The high-resolution crystal structure of covNHR in complex with the CHR-derived peptide C34, previously determined by our group, was retrieved from the Protein Data Bank (PDB ID: 6R2G) [21]. To model the structure of the free covNHR protein, we removed the C34 peptide, the water molecules and the phosphate ion present in the structure, and added missing hydrogens and heavy atoms in the coordinate file. After that, we modeled the first loop, which was undefined in the crystal structure, using the Build Loop tool of YASARA Structure (v17.12.24) [28]. The model was then subjected to energy minimization to remove bumps and correct the covalent geometry; the structure was energy-minimized with the NOVA force field [29], using an 8 Å distance cut-off and the Particle Mesh Ewald algorithm [30] to treat long-range electrostatic interactions. After removal of conformational stress by a short steepest descent minimization, the procedure continued by simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e. the energy improved by less than 0.05 kJ mol^{-1} per atom during 200 steps.

AutoDock 4.2 [31] was used to examine how ANS interacts with covNHR in both the free and peptide-bound forms. Docking was performed using the default docking parameters and point charges initially assigned according to the AMBER03 force field [32] and then the partial charges of atoms were calculated using the less polar Gasteiger-Marsili method to optimize the AutoDock scoring function. The 2D structure of the ANS ligand was drawn using the ChemDraw software, adding hydrogen atoms. The ANS structure was energy minimized by PM3 semiempirical techniques and the molecule was saved as a PDB format file prior to docking as described elsewhere [33]. A box 5 Å larger than the protein receptor in every dimension was generated with covNHR as the centroid. The Lamarckian genetic algorithm (LGA) [34] was used to search 100 independent ligand conformations and during the docking experiment the ligand was set as "flexible" meaning that it could be oriented within the grid box to generate different binding conformations while the protein was kept as "rigid", i.e. in a fixed conformation.

At the end of the docking experiments, the structures were ranked by energy and ligand-protein interactions were all visualized on the basis of docking results. The best docking poses were selected based on the estimated binding energy of the lowest energy conformation of the highest populated cluster and the best interactions between ligand and protein. The setup, modeling, visualization and analysis were performed

with the YASARA molecular modeling program [28].

2.7. Molecular dynamics simulations

All-atom molecular dynamics simulations were performed using YASARA Structure (v.17.12.24) [28] under the same parameters with explicit solvent (TIP3P water, the solvent density was equilibrated to a final value of 0.997 g/mL) in a periodic box with size 10 Å larger than the protein in every dimension. In order to describe long-range electrostatics, Particle Mesh Ewald (PME) [30] method was used with a cutoff distance of 8 Å at physiological conditions (0.9% NaCl, pH 7.4), constant temperature (298 K) using a weakly-coupled Berendsen thermostat and constant pressure (1 bar). Ewald summation was used to assign amino acids charge according to their predicted side chain pK_a and was neutralized by adding counterions (NaCl) [35]. The AMBER14 [36] force field was used together with multiple time step integration where intra-molecular forces were calculated every 2 fs and inter-molecular forces every 2.5 fs. The structures were initially energy-minimized using first steepest descent without electrostatics to remove steric clashes and conformational stress, and subsequently relaxed by steepest descent minimization and simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e. the energy improved by less than 0.05 kJ mol^{-1} per atom during 200 steps. The minimized system was slowly heated up during an equilibration phase until the target temperature and density were reached. Every system was simulated for a minimum of 100 ns and coordinates were saved every 10 ps, yielding 10,000 time points for each trajectory.

3. Results and discussion

3.1. ANS binds specifically to the gp41 hydrophobic pocket

ANS is an amphiphilic fluorescent dye that is commonly used to assess the presence of solvent-exposed hydrophobic patches in proteins [23]. To determine whether ANS can bind to covNHR we carried out ANS fluorescence emission experiments upon excitation at 370 nm. We found out that ANS binding to covNHR produced a fluorescence emission band with a maximum intensity at ~468 nm (Fig. 2A), characteristic of ANS binding to hydrophobic surfaces in proteins. Due to the reported ability of ANS to bind nonspecifically to a variety of surfaces in proteins we wanted to verify whether this fluorescence signal accounted for specific or nonspecific binding of ANS to covNHR.

We previously demonstrated that covNHR binds with very high affinity to several CHR derived peptides (such as the well-known C34 peptide) [21,37]. C34 (gp41 residues 117–150) interacts with three NHR sub-pockets, a N-terminal polar pocket (NTP), a shallow middle pocket (MP), as well as the HP. When added to the covNHR-ANS mixture, C34 decreased strongly the ANS fluorescence emission to a residual value (Fig. 2A), implying that this HP binding peptide can displace ANS from its binding location. A shorter peptide, W24N (residues 117–140), which also binds to the HP and the MP but does not cover the NTP, produced an identical fluorescence emission decrease.

In contrast, the Y24L peptide, which also binds to covNHR with high affinity (~90 nM) covering the NTP and the MP but lacking the HP binding motif, produced a lower decrease in ANS fluorescence emission, although the maximum ANS fluorescence intensity was red-shifted by ~9–10 nm. This bathochromic shift indicates a change to a more polar environment of bound ANS in presence of this peptide, suggesting a change in the binding mode of ANS to the HP.

To delimitate more precisely the location of the ANS binding site, we tested ANS binding against two covNHR miniproteins, each mimicking the N-terminal and C-terminal halves of the NHR region in gp41 (covNHR-N and covNHR-C respectively) [38]. CovNHR-N displays in its surface the NTP and MP sub-pockets, while covNHR-C exposes the HP and a minor C-terminal pocket (CTP). As expected, ANS fluorescence

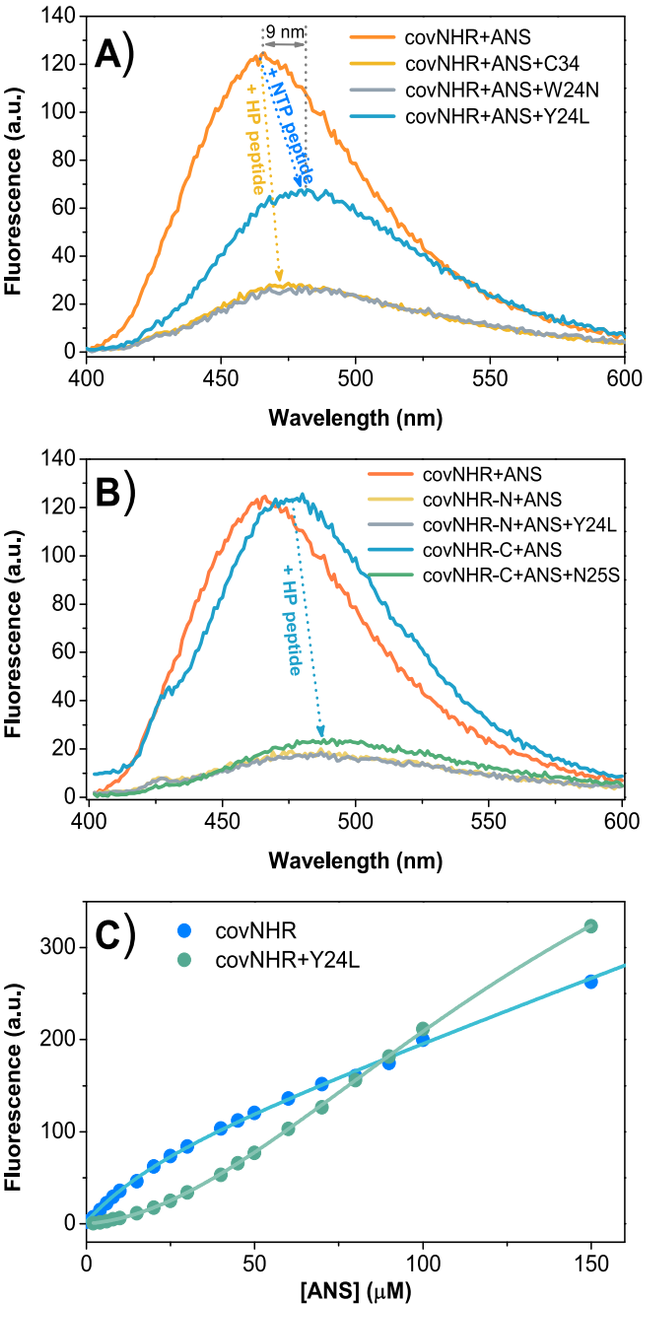


Fig. 2. Binding of ANS monitored by fluorescence. A) ANS fluorescence spectra in its 3:1 mixture with 10 μM covNHR in its free form and in presence of different CHR peptides (2:1 peptide to protein ratio). W24N and C34 peptides bind to the HP, whereas Y24L does not. B) Fluorescence spectra of ANS mixed at 3:1 ratio with shortened versions of covNHR [38], displaying different pockets of the NHR crevice. CovNHR-N displays the NTP and MP, whereas covNHR-C displays the HP and CTP. Addition of the N25S peptide abolishes the interaction of ANS with the HP pocket of covNHR-C. C) Fluorescence titration experiments of covNHR and ANS in presence (green) and absence (blue) of Y24L peptide showing the effect that this peptide causes in the mode of binding of ANS.

was only marginal in its mixture with covNHR-N and did not change in presence of bound Y24L peptide (Fig. 2B). However, ANS bound to covNHR-C induced a similar fluorescence emission band to that found upon ANS binding to covNHR. Interestingly, in presence of the N25S peptide, complementary to covNHR-C, the ANS fluorescence emission decreased to the residual intensity. This confirms that only peptides harboring the HP binding motif are able to efficiently reduce ANS fluorescence emission and evidences that the HP is the actual site for specific binding of ANS.

It is interesting to see how the maximum in the emission spectra of ANS in complex with covNHR-C is also red-shifted (~9 nm from 468 to 477 nm) compared to ANS bound to covNHR. This red-shift effect, frequently associated to a change in polarity of the environment of the fluorophore, also emerges when ANS binds to the covNHR-Y24L complex, suggesting a similar change in the ANS bound conformation.

To further investigate the binding of ANS to the HP, we carried out a direct titration of covNHR with ANS measuring fluorescence emission using a protein concentration of 10 μM (Fig. 2C). The titration shows an initial parabolic dependence of fluorescence intensity with ANS concentration, corresponding to a singular binding event, followed by a linear dependence that could be attributed to unspecific binding. Intensity does not reach a plateau within the concentration range explored. Preliminary fitting of the titration curve using a simple model of 1:1 binding followed by a linear fluorescence increase complies with relatively low apparent binding affinity (K_d ~13 μM).

Next, we studied the titration of 10 μM covNHR with ANS in presence of 2:1 molar ratio of the Y24L peptide, sufficient to saturate the protein with the peptide. The fluorescence titration curve (Fig. 2C green) does not comply with a simple binding event and there seems to be a sort of cooperative behavior in ANS binding at low concentrations giving rise to a sigmoidal shape. Saturation of the fluorescence intensity at high ANS concentration is not reached in this experiment.

3.2. Dual binding mode of ANS

Circular dichroism (CD) in the near-ultraviolet (near-UV) wavelength region can provide information on the local asymmetric environment of aromatic groups in proteins due to interactions established within the tertiary structure and is sensitive to the relative orientation and proximity of other aromatic groups [39], such as those present in ligands like ANS. Previously, we have shown that the free covNHR protein presents a weak negative band at ~280 nm, corresponding to the Trp143 sidechain (Trp60 in gp41 sequence) present in the HP, but in complex with CHR peptides stacking of two CHR tryptophan side chains (gp41 residues Trp117 and Trp120) into the HP causes the appearance of a characteristic negative ellipticity band centered at 293 nm [20,21,37].

Here, we have found that binding of ANS to covNHR also produces a specific sharp negative ellipticity band at 281 nm (Fig. 3), comparable in intensity to that caused by the interaction of CHR peptides with the HP. This band may have contributions from the aniline moiety of ANS and the Trp sidechain at the HP. In addition, a weaker and broader positive band at ~340 nm is also developed, attributable to immobilization of the naphthalene moiety [40]. The CD bands, therefore, seem to be induced by the ANS aromatic groups interacting with Trp60, consistently with the idea of ANS specific binding to the HP. These specific CD bands produced upon ANS binding to covNHR are completely abolished by the addition of C34 or W24N in 1:2 molar ratio, which instead elicit the appearance of the characteristic band at 293 nm, meaning that the pocket binding motif of the CHR peptides fully displaces ANS from the HP.

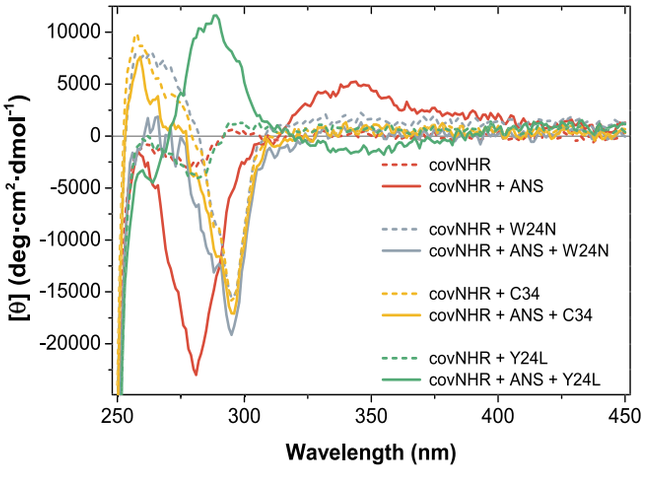


Fig. 3. Effect of ANS binding on circular dichroism spectra. Near-UV CD spectra of ANS mixtures (3:1 molar ratio) with free covNHR protein and in presence of CHR peptides (2:1 molar ratio). The protein concentration was 40 μM. Spectra in absence of ANS (light red, light yellow and light gray) were taken from reference [37].

Remarkably, when ANS binds to covNHR in presence of Y24L peptide, devoid of the HP binding motif, a positive CD band appears centered at around ~285 nm and a negative band at ~343 nm, approximately mirroring the CD spectrum of ANS bound to free covNHR and suggesting a change in the spatial orientation of bound ANS. This finding is in agreement with the hypothesis that the presence of Y24L bound to covNHR alters ANS mode of binding to the HP. This implies an allosteric communication between the different binding pockets along the NHR crevice, as previously proposed [37].

3.3. Thermodynamics and stoichiometry of ANS binding

We used ITC to investigate the interaction between ANS and covNHR, both in its free form and in presence of Y24L peptide. First, we carried out a direct ANS titration of free covNHR at 10 μM and 36 μM (Fig. 4A). The negative peaks obtained in the ITC thermogram indicate an exothermic interaction. In addition, at low protein concentration the normalized heats were very small but increase strongly at high protein concentration. The shape of the binding isotherms is not sigmoidal but hyperbolic, and the heats do not tend to zero at high ANS to protein ratios.

Preliminary analysis of the ITC isotherm at high concentration using a binding model of n-independent and identical sites does not allow a good fitting. Fixing the binding stoichiometry to values lower than 1:1 produces a reasonable description of the ITC isotherm (Fig. S1), suggesting that ANS may be binding to more than one covNHR molecule or even favoring its self-association. This would explain the observed concentration dependence of the ITC isotherms.

We recently reported the characterization of several covNHR

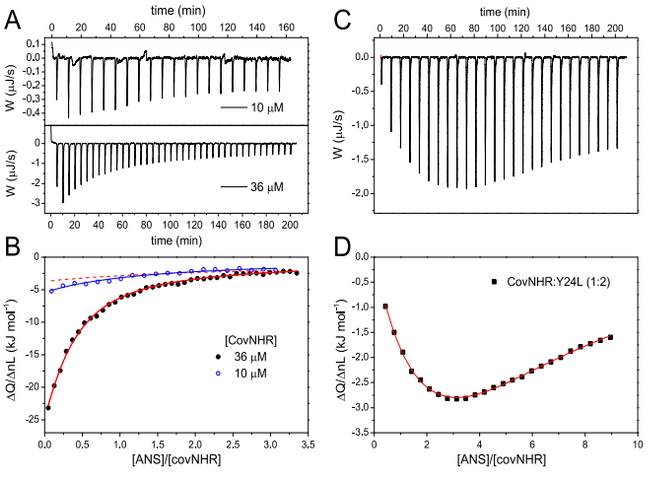


Fig. 4. ITC analysis of the interaction between ANS and covNHR in presence and absence of bound Y24L peptide. A) ITC thermograms for the ANS titration of covNHR at 10 μM and 36 μM; B) binding isotherms analyzed with a model of binding coupled to dimer formation (2 M + A ↔ M_2A). Dashed red line represents the simulated isotherm for the titration at 10 μM using the parameters of the titration at 36 μM. C) ITC thermograms for the ANS titration of covNHR at 35 μM in presence of 70 μM Y24L peptide; D) binding isotherm analyzed with a two-site sequential binding model. Symbols represent the experimental heats per mole of added ANS. The solid line corresponds to the best fits with the model.

mutants, in which buried polar residues at core positions in the coiled-coil where replaced by isoleucine [41]. These mutants show a significantly higher propensity to self-associate as dimers compared to the reference covNHR protein, which is mainly monomeric at low concentrations although it also maintains a residual propensity to dimerize at high concentrations (above ~75 μM). To test the possibility that ANS binds preferentially to covNHR dimers, we fitted the ITC isotherm using a model of ligand binding coupled to dimer formation ($2M + A \leftrightarrow M_2A$; see Supplementary Text S1 for details of this analysis). The model fits satisfactorily the ITC data supporting a binding stoichiometry of two covNHR proteins to one ANS molecule (Fig. 4B). To explain the protein concentration dependence of the binding isotherm, we simulated the binding curve of the ITC titration at 10 μM using the fitting parameters obtained with the titration at 36 μM. Although the correspondence is not perfect, the curve is very close and can be perfectly adjusted with quite similar parameters.

To support this binding mode, we carried out a similar ANS titration of the mutant protein covNHR-Q23I, which contains three core Q-I mutations that promote a higher dimerization propensity [41]. The ITC data could also be very well fitted with the model of binding to the dimer (Fig. S2). The thermodynamic parameters obtained from these analyses (Table 1) indicate a strongly exothermic binding for both protein variants, which suggests a considerable stabilization of the proteins induced by the coupled processes of ANS binding and dimerization.

To further confirm the anomalous binding stoichiometry of ANS, we carried out dynamic and static scattering measurements to elucidate the influence of ANS upon the molecular size of the protein. DLS measurements (Table S1) show an increment in the apparent hydrodynamic radii ($R_h = 3.1$ nm) of the covNHR protein in presence of ANS in 2:1 molar ratio, compared to that of the free protein (2.7 nm), which exhibits a value expected for a monomer, indicating an event of self-association favored by ANS. This effect is reverted in the complex covNHR:Y24L (2.8 nm) and in the complex covNHR:Y24L:ANS (2.7 nm) indicating that peptide binding shifts the equilibrium towards the monomeric state even in presence of ANS.

Static light scattering measurements confirmed that ANS binding favors the dimerization of covNHR, compared to the free protein (Fig. 5), since the Debye plot intercepts provide absolute values of the molecular weights. Only at very low concentrations, the binding of ANS to the monomeric protein could perhaps be relevant but the low sensitivity of the ITC data at these low protein concentrations do not allow an accurate determination of the binding parameters to the monomer. The binding mode of ANS to the covNHR dimer may imply the mutual insertion of the Trp143 side chains of each covNHR protomer into the reciprocal HP and aromatic stacking with the ANS groups. We observed a similar interaction in the crystal structure of the free covNHR molecule (PDB code: 4R61) [20].

In sharp contrast, the covNHR:Y24L complex showed a monomeric state in all the concentration range studied, as demonstrated in the Debye plot of Fig. 5. Moreover, ANS binding to this complex does not appear to induce any dimerization. These results further support a mechanism of ANS-mediated dimerization of the free protein as proposed above. In addition, occupancy of the NHR crevice by the Y24L peptide is incompatible with formation of ANS-mediated dimers.

Table 1
Thermodynamic parameters of binding of ANS to covNHR protein in free form and bound to the Y24L peptide. Uncertainties in the parameters correspond to standard errors of the fits.

Binding model: $2M + A \leftrightarrow M_2A$		
Protein	K_2 (M^{-2})	ΔH_2 (kJ mol^{-1})
covNHR	$(5.7 \pm 0.3) \times 10^8$	-56.4 ± 1.8
covNHR-Q23I	$(8.4 \pm 0.7) \times 10^8$	-43.7 ± 2.2

Binding model: $M + A \leftrightarrow MA$; $MA + A \leftrightarrow MA_2$				
Protein	K_1 (M^{-1})	ΔH_2 (kJ mol^{-1})	K_2 (M^{-1})	ΔH_2 (kJ mol^{-1})
covNHR+Y24L	$(5.6 \pm 1.1) \times 10^3$	$+0.3 \pm 0.2$	$(8.0 \pm 1.2) \times 10^3$	-36.5 ± 1.5

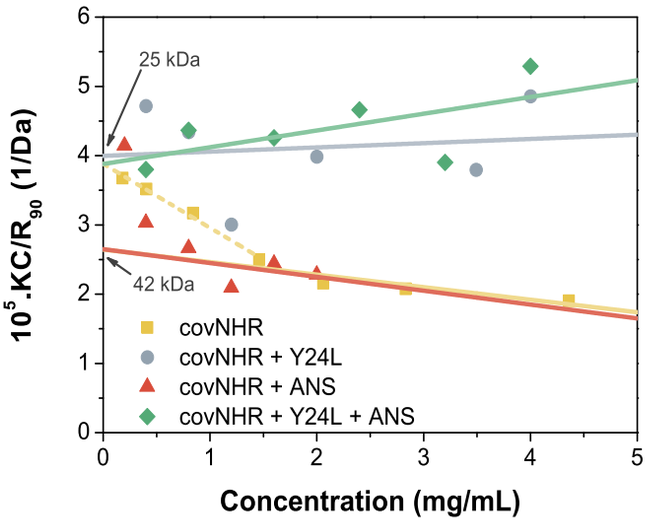


Fig. 5. Debye plots for the static light scattering (SLS) intensity of covNHR with Y24L peptide and ANS in different mixture solutions measured at different protein concentrations. Symbols indicate the experimental results and lines are drawn only for visual purposes.

The ITC titration of covNHR bound to Y24L shows a dramatic change in behavior (Fig. 4C–D). The data comply well with cooperative binding, as suggested by the aforementioned fluorescence titrations. We analyzed the ITC isotherm using a sequential binding model, in which two ANS molecules bind to the HP ($M + A \leftrightarrow MA$; $MA + A \leftrightarrow MA_2$; see Supplementary material). The model reproduces very well the binding isotherms and indicates that two ANS molecules can occupy the HP with positive binding cooperativity.

This dramatic change in the binding mode of ANS to the HP is induced by occupancy of the distant NTP and MP by the Y24L peptide. The thermodynamic parameters of binding derived from this analysis (Table 1) indicate a low affinity binding of a first ANS molecule with slightly positive enthalpy and a second binding event with significantly higher negative enthalpy. The first ANS binding is therefore entropically driven and complies with a mainly hydrophobic interaction of ANS with the HP. In contrast, the second binding event is driven by enthalpy, suggesting a considerable contribution of polar/electrostatic interactions. The cooperativity ratio [42], $\rho = 4 \cdot K_2/K_1 = 5.7 \pm 2.0$, indicates a moderately positive cooperativity. These results clearly show that the HP can hold two ANS molecules establishing energetically favorable communication.

The global enthalpy of ANS binding is much smaller in magnitude than that of binding to the free protein, likely because the bound CHR peptide has already stabilized the protein, removing part of the structural tightening contribution.

3.4. Computational modeling and molecular docking

To analyze the ANS binding mode in more detail, we performed a molecular docking simulation study using as template the crystallographic structure of covNHR in complex with C34 (Fig. 6A) (PDB ID: 6R2G) [21]. First, we carried out a blind-docking simulation on the entire protein surface to explore the binding mode of ANS. The top ranked poses in each cluster, regarded as those having the highest binding energy, correspond to binding poses of ANS in the HP. The top ranking conformation for the covNHR-ANS complex (Fig. 6D) evidences

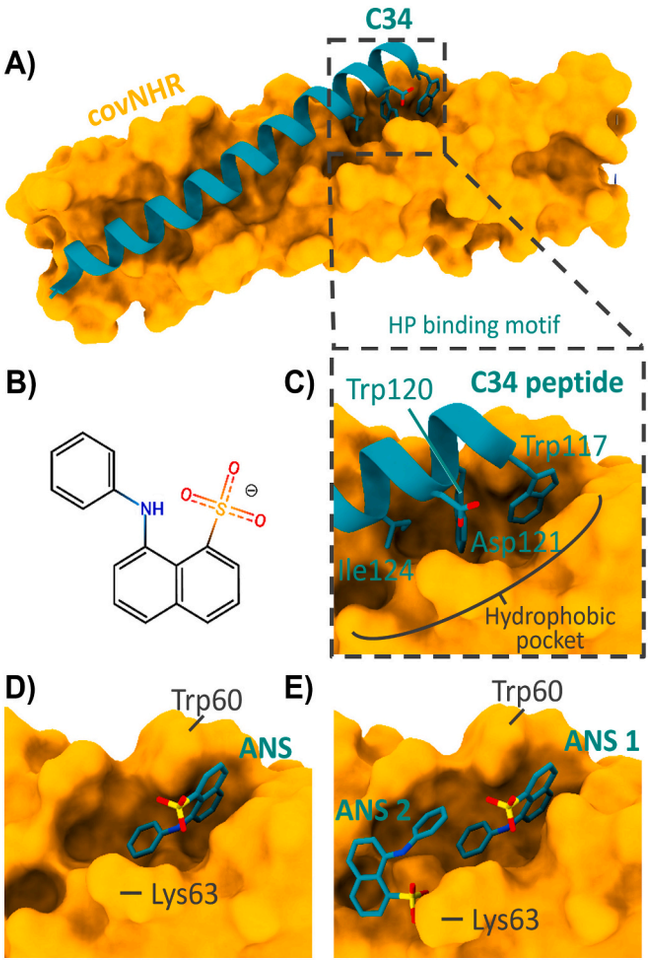


Fig. 6. Structural insight into the ANS binding to the HP pocket of covNHR. A) covNHR in complex with C34 peptide (PDB ID: 6R2G). B) 2D structural representation of ANS molecule colored by atom type. C) Close up view of the HP pocket and the HP binding motif. D) Top scored ANS molecule docked into covNHR showing that it binds to the HP. E) Top scored cluster of ANS molecules sequentially docked into the HP of covNHR in presence of bound Y24L peptide.

that the ANS aromatic rings adopt a similar orientation to that of Trp sidechains of the HP binding motif of C34 (Fig. 6C). This binding pose of ANS within the HP helps to rationalize the results obtained in the previous fluorescence and CD experiments, as peptides owning the HP binding motif displace ANS due to the fact that they share an overlapping binding region (Figs. 1 and 2A and 3).

To inspect the binding mode of ANS within covNHR in complex with Y24L, we conducted a similar docking procedure but this time with the structure of Y24L present in the protein receptor. Because the results obtained in the previous sections reveal that not only one but two molecules of ANS bind cooperatively to covNHR in the presence of Y24L, we carried out a subsequent sequential docking round in order to assess this dual binding mode.

The structure of the complex between covNHR and Y24L was generated as described above but, instead of eliminating the whole C34 peptide, only 10 N-terminal amino acids were removed so that the sequence of Y24L was maintained. After that, as the commercially acquired peptides are *N*-acetylated and C-amidated those residues where added to the respective N- and C-terminal ends of Y24L and the complex structure was energy minimized as described earlier.

We performed a first docking simulation of ANS with the covNHR: Y24L complex structure with the grid box centered on the HP and a box size of 7x7x7 $Å^3$, ensuring a sufficient exploration of the HP to generate favorable interactions. The resulting docking structure was energy minimized as described above. This structure was then used to assess the 2:1 cooperative binding between ANS and covNHR when Y24L peptide is present, by performing a sequential docking experiment with the same parameters as described before. The results of the docking simulations in the presence of Y24L are shown in Fig. 6E. As expected, the first HP-bound ANS molecule (ANS 1) in presence of Y24L adopts a similar conformation to that found in the absence of the peptide. It is interesting to find that two molecules of ANS can be accommodated within the hydrophobic pocket when Y24L peptide is present, in good agreement with the results of the ITC experiments discussed above. Moreover, the spatial orientation of the second bound ANS (ANS 2) is completely upturned relative to ANS 1 and its aromatic groups are less deeply inserted within the HP. These differences are fully consistent with the spectral changes in the near-UV CD bands and with the red-shift in the fluorescence of bound ANS observed in presence of Y24L.

The interactions established between ANS and the residues at the hydrophobic pocket are summarized in Table S2. It is interesting to note that ANS 1 exhibits a higher number of hydrophobic contacts with the surrounding amino acids that constitute the HP. A total of 8 hydrophobic interactions were listed. On the other hand, ANS 2 shows only 3 hydrophobic contacts with the residues at the HP, but it establishes a hydrogen bond with its sulfate group and the amino group of Lys35 (structurally equivalent to Lys63 in gp41 sequence; Table S2). Moreover, both ANS molecules mutually establish a π-π stacking interaction between their aniline moieties. This interaction justifies the observed positive cooperativity of binding between the two ANS molecules.

3.5. Molecular dynamics simulations

To further validate and refine the docking results and assess the stability and dynamic behavior of the covNHR-ANS and covNHR-Y24L-ANS complexes, we carried out 100 ns all-atom explicit-solvent MD simulations using the top scored docking pose of each complex as starting point. Additionally, the structures of free covNHR and the covNHR-Y24L complex were also evaluated with the same MD protocol. The ANS binding poses obtained by docking were stable throughout the entire simulation time without significant differences between the initial and the final ligand poses (Fig. 7).

Fig. 7A shows the time evolution of backbone root mean square deviations (RMSD) of the covNHR protein from each initial structure. All the systems reached equilibrium within the first 20 ns of MD simulation and remain stable throughout the whole simulation time reaching a similar plateau. When comparing the probability distribution function of each system's RMSD (Fig. 7B), small overall reductions in RMSD values can be observed when ANS molecules occupy the HP, suggesting a higher stability of the ANS-bound structures.

Larger differences corroborating these findings can be observed by comparing the evolution of the radius of gyration (Rg), which is a good estimate of the structure compactness (Fig. 7C). Rg values of the covNHR complexes with the different ligands are consistently smaller than the values observed for the free protein during the simulation time. Those differences become more evident when the probability distribution functions of the Rg are compared (Fig. 7D). Binding of Y24L and/or ANS increased progressively the compactness of the system.

Moreover, the root mean square fluctuation (RMSF) values, which are a measurement of the average atomic mobility of the protein in the MD simulations, show an increase in the stability of the protein when it is in presence of ANS and/or Y24L peptide (Fig. 7E). The RMSF mean values decrease with the binding of Y24L peptide or ANS compared to the free protein, and become further decreased for the complex with Y24L and two ANS molecules. This reinforces the idea that the protein complexes are less dynamic than the protein in its free form. The residues at the binding interfaces with the ligands show little fluctuations during the MD simulation time indicating that they form stable interactions with the protein. Interestingly, the main fluctuations correspond to the first and second loop of the protein (together with the N- and C-terminal ends of the protein), which are far from the ligand

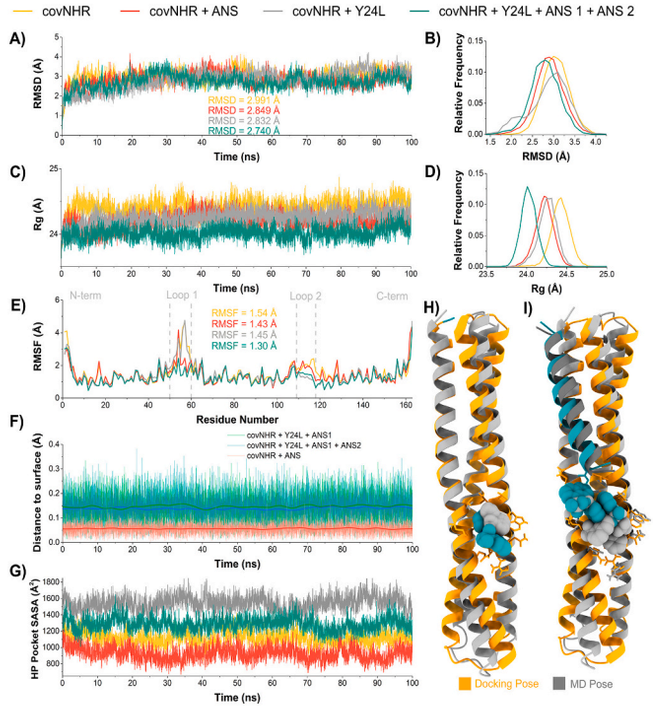


Fig. 7. Molecular dynamics simulations analyses of the four different systems: yellow, free covNHR; red, covNHR in complex with one ANS molecule; gray, covNHR in complex with Y24L; green, covNHR in complex with Y24L and two ANS molecules. A) Evolution of mean backbone root mean square deviation (RMSD) for each simulated system. B) Probability distribution function of RMSD values shown in A). C) Evolution of the radius of gyration (Rg) for each system. D) Probability distribution functions of Rg values shown in C). E) Root mean square fluctuation (RMSF) per residue for each system. F) Evolution of the distance of ANS molecules to the closest point on the surface of covNHR, green and blue lines indicate the distance of ANS 1 and ANS 2 respectively in the covNHR:Y24L; red line indicates the distance of the only ANS molecule present in its complex with covNHR. G) Hydrophobic pocket solvent accessible surface area (SASA) throughout the MD trajectories. The HP area in covNHR was calculated as the sum of SASA of residues: Gln24, Leu27, Gln28, Val31, Ser32, Ile34, Lys35, Gln38, Leu137, Leu140, Thr141, Trp143, Gly144, Ile145, Gln147, Leu148 and Arg151. H, I) Superposed structures between the docking poses and the average MD poses throughout the entire simulation time for H) covNHR and ANS and I) covNHR in complex with both Y24L and two ANS molecules.

interfaces. In addition, the different complexes show reduced fluctuations in these regions, in agreement with an increased stability and a long-range connection between different regions produced by ligand binding, and also related to the observed decrease in Rg values.

The ability of ANS to establish stable interactions is further corroborated by the close contact maintained by the ANS molecules with the protein surface during the whole simulation time (Fig. 7F). ANS molecules bind to the HP and remain in it with little fluctuations during the simulations not dissociating and maintaining a short distance between each ANS molecule and the protein as can be seen in Fig. 7H and I (for the complex covNHR-ANS and covNHR-Y24L-ANS1-ANS2 respectively).

Fig. 7G shows evolution of the HP size throughout the MD trajectories according to the solvent accessible surface area (SASA) of the HP residues in each system calculated removing the ligands. The HP is enlarged in presence of bound Y24L compared to free covNHR, thus facilitating the binding of a second molecule of ANS. Binding of ANS to the HP in either the free protein (red vs yellow in Fig. 7G) or the covNHR-Y24L complex (green vs gray) produces a reduction in SASA of the HP, indicating an induced fit binding of ANS molecules to the receptor regardless of the presence of Y24L peptide. These findings highlight a considerable conformational flexibility of the HP that is modulated by the occupancy of the rest of the hydrophobic groove.

4. Discussion

Finding small-molecule inhibitors targeting the gp41 HP has been a challenging task despite extensive efforts both from scientific community and pharmaceutical companies. One of the reasons is the difficulty to develop a suitable molecular model of gp41 NHR for combined computational and experimental research, as well as for high-throughput screening. We have previously shown that our covNHR proteins are very useful models displaying a functionally relevant HP, as demonstrated by their strong capacity to compete with internal gp41 interactions and to broadly inhibit HIV-1 [20,21,37,38,41]. The results presented here show how this mimetic protein allows a combined experimental-computational study to characterize structurally and energetically two different binding modes of ANS to the HP.

While ANS has been extensively used as a probe for surface exposed hydrophobicity in proteins, it shares several characteristics with other small molecule inhibitors, including aromatic groups and the presence of a negatively charged group. This has been considered a key feature in other inhibitors to interact with the preserved Lys63 side chain at the HP [43]. Also, the docking poses found here imitate the hydrophobic interactions established by the Trp-Trp-Ile pocket-binding motif of CHR. However, we demonstrate that binding of a single ANS molecule to the HP shows an irrelevant affinity because the HP is too large for such a small compound to establish sufficient interactions. This might be applicable to similarly sized compounds that have been tested in the past as potential fusion inhibitors targeting the HP. It has been observed that peptide-based drugs are most effective in inhibiting protein-protein interactions that bury more than 2000 Å^2 of accessible surface area, whereas small-molecule compounds usually target smaller interfaces [44]. Therefore, blocking the gp41 6HB formation with small-molecule inhibitors is a great challenge because the NHR-CHR interaction extends along a quite large interface burying more than 3000 Å^2 [21] and the binding energy is largely distributed between different pockets on the NHR crevice [37,45]. Nevertheless, the HP continues to be a hotspot in this interaction because its contribution to the overall NHR-CHR binding energy is predominant [37]. Therefore, larger compounds may still constitute good HP binders interfering with 6HB formation provided they fill more completely the HP cavity and establish sufficiently favorable interactions.

We demonstrate here that the HP is conformationally flexible and dynamic and shows allosteric communication with the rest of the NHR is in agreement with our previous work [37]. Therefore, binding at other places of the NHR crevice alters the conformation and interactions at the HP opening it and allowing it to hold two ANS molecules.

The conformational flexibility of the HP reported here might have a functional relevance for the fusion mechanism. Recently, it has been reported that the Trp-rich region of the MPER can also interact with the HP [46]. This binding promiscuity of the HP implies a significant conformational plasticity. However, these MPER-HP interactions should only be transient, as gp41 must progress to its post-fusion conformation for a productive pore formation. The establishment of additional interactions with the CHR along the NHR crevice, possibly involving the MP and NTP pockets, may be key to modulate allosterically the HP conformation, regulate transient interactions, and finally to lock the HP in its high-affinity post-fusion conformation. A more profound understanding of this allosterically regulated process would help to better understand the gp41-driven fusion mechanism and guide the search of more potent inhibitors.

Our results show that the HP can hold up to two ANS molecules in its conformation stabilized by CHR-peptides. The thermodynamic signature of the two sequential binding events together with the docking and MD analysis also helps to understand the mode of interaction. Binding of the first ANS molecule is entropically driven, indicating a mainly hydrophobic interaction. This is consistent with the geometric orientation of the first bound ANS molecule found in our docking study, in which the two aromatic rings insert profoundly into two cavities of the pocket. In

contrast, the binding of the second molecule is enthalpically driven, consistently with a more important contribution of polar interactions. In fact, in the docking pose of the second ANS molecule, its spatial orientation is reversed, so that the sulfate group can interact electrostatically with the side chain of Lys63 (Lys35 in covNHR sequence). This interaction replaces a highly conserved salt bridge between Asp121 and Lys63 in the post-fusion gp41 structure, which plays a critical role in 6HB formation [43]. Moreover, there is a clear stacking between the benzyl rings of the two ANS molecules, which explains the observed binding cooperativity. These results suggest that compounds mimicking this particular molecular geometry and interactions may constitute good leads for new inhibitors targeting the HP. A considerable number of small compounds targeting the HP have been identified or designed previously (recently reviewed in [47]). Many of them were identified by computational screening involving docking analysis or were submitted later to docking protocols to identify the chief interactions. The main features of these compounds with high affinity for the HP have been summarized in several characteristics [48]; namely, an extended hydrophobic scaffold that fills completely the HP [49]; a negatively charged group that can interact electrostatically with Lys63 or Arg68 [50] and/or establish H-bonds with Gln56 or Gln64 [51]; aromatic-ring stacking interactions involving ligand groups and Trp60 in gp41. All these features are perfectly met in the ANS-HP interactions observed in our docking analysis. In fact, the residues lining the HP and the type of interactions involved in ANS binding (Table S2) are largely coincident with those reported in the literature for the most relevant compounds.

The results of our docking and MD studies facilitate potential strategies to identify or design new ANS derivatives with improved binding affinity for the HP. A possible strategy is to reduce the rotational-translational entropy of binding by tethering with an appropriate linker two ANS fragments in a similar orientation to that found in our docking pose. Another possibility is to perform structure similarity virtual screening of compound libraries to identify drug-like compounds with analogous structure to that of the ANS dimer found in our docking poses. Moreover, the discovery of new HP targeting inhibitors may also be facilitated using our covNHR protein by competitive fluorescence-based screening assays, based on the results presented here, to detect HP binding compounds. This potential screening methodology stands out by the ease of using a recombinant protein, which is purified easily and with high yield at a laboratory scale and a widespread fluorescent probe such as ANS.

5. Conclusions

In this study we show that ANS is able to bind specifically to the HP of the gp41 NHR region, a promising target for fusion inhibitors. An experimental structural and thermodynamic characterization combined with computational methods allowed us to discern between two different binding modes of ANS to the HP. The mode of ANS binding is modified allosterically by the occupancy of other NHR pockets by a CHR peptide, demonstrating that the HP shows a marked conformational flexibility. In its stabilized form, the HP can hold two ANS molecules that bind cooperatively, mimicking functionally relevant interactions at the HP.

ANS-based derivative molecules may thus constitute potential drug candidates, providing a valuable class of new lead structures for the design of potent HIV fusion inhibitors. These new structures may be generated upon similarity search within compound libraries or by rational design with the data gathered here. The rational design approach is facilitated by the fact that in this study we exposed a possible way of interaction between this small-molecule and its protein target, which may guide the designing step towards more potent inhibitors. Discovery of the novel scaffolds and binding mechanism suggests avenues for extending the interaction surface and improving the potency of HP-targeting inhibitors.

CRediT authorship contribution statement

Mario Cano-Muñoz: Conceptualization, Investigation, Methodology, Software, Data curation; Formal analysis; Writing - original draft. **Samuel Jurado**: Investigation. **Bertrand Morel**: Methodology, Supervision. **Francisco Conejero-Lara**: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Writing - review & editing.

Declaration of competing interest

Authors declare there is no conflict of interest.

Acknowledgements

We thank Dr. Angel Pey for his fruitful comments and discussion. This research was funded by the Spanish State Research Agency, SRA/10.13039/501100011033 (grants BIO2016-76640-R and PID2019.107515RB.C21), co-funded by ERDF/ESF, "A way to make Europe"/"Investing in your future". The results shown are included as part of MCM doctoral thesis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2021.09.198.

References

[1] D.C. Chan, P.S. Kim, HIV entry and its inhibition, Cell 93 (1998) 681–684, https://doi.org/10.1016/s0092-8674(00)81430-0.
[2] J.M. White, S.E. Delos, M. Brecher, K. Schornberg, Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme, Crit. Rev. Biochem. Mol. 43 (2008) 189–219, https://doi.org/10.1080/10409230802058320.
[3] A.C. Walls, Y.J. Park, M.A. Tortorici, A. Wall, A.T. McGuire, D. Veesler, Structure, Function, and antigenicity of the SARS-CoV-2 spike glycoprotein, Cell 181 (2020) 281–292 e286, https://doi.org/10.1016/j.cell.2020.02.058.
[4] UNAIDS, UNAIDS Data 2020, 2020. http://www.unaids.org/sites/default/files/media_asset/2020_aids-data-book_en.pdf. (Accessed 18 April 2021).
[5] P. Iyidogan, K.S. Anderson, Current perspectives on HIV-1 antiretroviral drug resistance, VirusBasel 6 (2014) 4095–4139, https://doi.org/10.3390/v6104095.
[6] B.F. Silva, G.M.L. Peixoto, S.R. da Luz, S.M.F. de Moraes, S.B. Peres, Adverse effects of chronic treatment with the Main subclasses of highly active antiretroviral therapy: a systematic review, HIV Med. 20 (2019) 429–438, https://doi.org/10.1111/hiv.12733.
[7] A.B. Ward, I.A. Wilson, The HIV-1 envelope glycoprotein structure: nailing down a moving target, Immunol. Rev. 275 (2017) 21–32, https://doi.org/10.1111/imr.12507.
[8] W. Weissenhorn, A. Dessen, S.C. Harrison, J.J. Skehel, D.C. Wiley, Atomic structure of the ectodomain from HIV-1 gp41, Nature 387 (1997) 426–430, https://doi.org/10.1038/387426a0.
[9] D.C. Chan, D. Fass, J.M. Berger, P.S. Kim, Core structure of gp41 from the HIV envelope glycoprotein, Cell 89 (1997) 263–273, https://doi.org/10.1016/s0092-8674(00)80205-6.
[10] H.A. Yi, B.C. Fochtman, R.C. Rizzo, A. Jacobs, Inhibition of HIV entry by targeting the envelope transmembrane subunit gp41, Curr. HIV Res. 14 (2016) 283–294, https://doi.org/10.2174/1570162x14999160224103908.
[11] J.M. Kilby, S. Hopkins, T.M. Venetta, B. DiMassimo, G.A. Cloud, J.Y. Lee, L. Alldredge, E. Hunter, D. Lambert, D. Bolognesi, T. Mathews, M.R. Johnson, M. A. Nowak, G.M. Shaw, M.S. Saag, Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry, Nat. Med. 4 (1998) 1302–1307, https://doi.org/10.1038/3293.
[12] L. Cai, S. Jiang, Development of peptide and small-molecule HIV-1 fusion inhibitors that target gp41, ChemMedChem 5 (2010) 1813–1824, https://doi.org/10.1002/cmdc.201000289.
[13] S.B. Jiang, H. Lu, S.W. Liu, Q. Zhao, Y.X. He, A.K. Debnath, N-substituted pyrrole derivatives as novel human immunodeficiency virus type 1 entry inhibitors that interfere with the gp41 six-helix bundle formation and block virus fusion, Antimicrob. Agents Chemother. 48 (2004) 4349–4359, https://doi.org/10.1128/aac.48.11.4349-4359.2004.
[14] K. Lu, M.R. Asyifah, F. Shao, D. Zhang, Development of HIV-1 fusion inhibitors targeting gp41, Curr. Med. Chem. 21 (2014) 1976–1996, https://doi.org/10.2174/0929867321666131218094559.
[15] L. Lu, F. Yu, L. Cai, A.K. Debnath, S. Jiang, Development of small-molecule HIV entry inhibitors specifically targeting gp120 or gp41, Curr. Top. Med. Chem. 16 (2016) 1074–1090.

[16] G.Y. Zhou, D. Wu, B. Snyder, R.G. Ptak, H. Kaur, M. Gochin, Development of indole compounds as small molecule fusion inhibitors targeting HIV-1 Glycoprotein-41, J. Med. Chem. 54 (2011) 7220–7231, https://doi.org/10.1021/jm200791z.
[17] D.M. Eckert, P.S. Kim, Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 11187–11192, https://doi.org/10.1073/pnas.201392898.
[18] M.J. Root, M.S. Kay, P.S. Kim, Protein design of an HIV-1 entry inhibitor, Science 291 (2001) 884–888, https://doi.org/10.1126/science.1057453.
[19] F. Yu, L. Lu, L.Y. Du, X.J. Zhu, A.K. Debnath, S.B. Jiang, Approaches for identification of HIV-1 entry inhibitors targeting gp41 pocket, VirusBasel 5 (2013) 127–149, https://doi.org/10.3390/v5010127.
[20] S. Crespillo, A. Camara-Artigas, S. Casares, B. Morel, E.S. Cobos, P.L. Mateo, N. Mouz, C.E. Martin, M.G. Roger, R. El Habib, B. Su, C. Moog, F. Conejero-Lara, Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti-HIV-1 drugs, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 18207–18212, https://doi.org/10.1073/pnas.1413592112.
[21] S. Jurado, M. Cano-Munoz, B. Morel, S. Standoli, E. Santarossa, C. Moog, S. Schmidt, G. Laumond, A. Camara-Artigas, F. Conejero-Lara, Structural and thermodynamic analysis of HIV-1 fusion inhibition using small gp41 mimetic proteins, J. Mol. Biol. 431 (2019) 3091–3106, https://doi.org/10.1016/j.jmb.2019.06.022.
[22] F. Manssour-Triedo, S. Crespillo, B. Morel, S. Casares, P.L. Mateo, F. Notka, M. G. Roger, N. Mouz, R. El-Habib, F. Conejero-Lara, Molecular and physicochemical factors governing solubility of the HIV gp41 ectodomain, Biophys. J. 111 (2016) 700–709, https://doi.org/10.1016/j.bpj.2016.07.022.
[23] A. Hawe, M. Sutter, W. Jiskoot, Extrinsic fluorescent dyes as tools for protein characterization, Pharm. Res. 25 (2008) 1487–1499, https://doi.org/10.1007/s11095-007-9516-9.
[24] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 23 (1997) 3–25, https://doi.org/10.1016/s0169-409x(96)00423-1.
[25] M.R. Wilkins, E. Gasteiger, A. Bairoch, J.-C. Sanchez, K.L. Williams, R.D. Appel, D. F. Hochstrasser, Protein identification and analysis tools in the ExPASy server, in: Methods in Molecular Biology; 2-D proteome analysis protocols, 1999, pp. 531–552.
[26] R.F. Latypov, D. Liu, K. Gunasekaran, T.S. Harvey, V.I. Razinkov, A.A. Raibekas, Structural and thermodynamic effects of ANS binding to human interleukin-1 receptor antagonist, Protein Sci. 17 (2008) 652–663, https://doi.org/10.1110/ps.073332408.
[27] A.V. Fonin, A.I. Sulatskaya, I.M. Kuznetsova, K.K. Turoverov, Fluorescence of dyes in solutions with high absorbanceInner Filter Effect Correction, PLoS One 9 (2014), e103878, https://doi.org/10.1371/journal.pone.0103878.
[28] E. Krieger, G. Vriend, YASARA view-molecular graphics for all devices-from smartphones to workstations, Bioinformatics 30 (2014) 2981–2982, https://doi.org/10.1093/bioinformatics/btu426.
[29] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA - a self-parameterizing force field, Proteins 47 (2002) 393–402, https://doi.org/10.1002/prot.10104.
[30] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh ewald method, J. Chem. Phys. 103 (1995) 8577–8593, https://doi.org/10.1063/1.470117.
[31] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A. J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, J. Comput. Chem. 30 (2009) 2785–2791, https://doi.org/10.1002/jcc.21256.
[32] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G.M. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J.M. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012, https://doi.org/10.1002/jcc.10349.
[33] M. Medina-O'Donnell, F. Rivas, F.J. Reyes-Zurita, M. Cano-Munoz, A. Martinez, J. A. Lupianez, A. Parra, Oleanolic acid derivatives as potential inhibitors of HIV-1 protease, J. Nat. Prod. 82 (2019) 2886–2896, https://doi.org/10.1021/acs.jnatprod.9b00649.
[34] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A. J. Olson, Automated docking using a lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662, https://doi.org/10.1002/(sici)1096-987x(19981115)19:14<1639::aid-jcc10>3.0.co;2-b.
[35] E. Krieger, J.E. Nielsen, C.A.E.M. Spronk, G. Vriend, Fast empirical pK(a) prediction by ewald summation, J. Mol. Graph. 25 (2006) 481–486, https://doi.org/10.1016/j.jmgm.2006.02.009.
[36] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple amber force fields and development of improved protein backbone parameters, Proteins 65 (2006) 712–725, https://doi.org/10.1002/prot.21123.
[37] S. Jurado, M. Cano-Munoz, D. Polo-Megias, F. Conejero-Lara, B. Morel, Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects, Arch. Biochem. Biophys. 688 (2020), 108401, https://doi.org/10.1016/j.abb.2020.108401.
[38] S. Jurado, C. Moog, M. Cano-Munoz, S. Schmidt, G. Laumond, V. Ruocco, S. Standoli, D. Polo-Megias, F. Conejero-Lara, B. Morel, Probing vulnerability of the gp41 C-terminal heptad repeat as target for miniprotein HIV inhibitors, J. Mol. Biol. 432 (2020) 5577–5592, https://doi.org/10.1016/j.jmb.2020.08.010.
[39] S.B. Jasim, Z. Li, E.E. Guest, J.D. Hirst, DichroCalc: improvements in computing protein circular dichroism spectroscopy in the near-ultraviolet, J. Mol. Biol. 430 (2018) 2196–2202, https://doi.org/10.1016/j.jmb.2017.12.009.
[40] E. Klimtchuk, S. Venyaminov, E. Kurian, W. Wessels, W. Kirk, F.G. Prendergast, Photophysics of ANS. I., Protein-ANS complexes: intestinal fatty acid binding protein and single-trp mutants, Biophys. Chem. 125 (2007) 1–12, https://doi.org/10.1016/j.bpc.2006.07.016.
[41] M. Cano-Munoz, S. Cesaro, B. Morel, J. Lucas, C. Moog, F. Conejero-Lara, Extremely thermostabilizing Core mutations in coiled-coil mimetic proteins of HIV-1 gp41 produce diverse effects on target binding but do not affect their inhibitory activity, Biomolecules 11 (2021) 566, https://doi.org/10.3390/biom11040566.
[42] E. Freire, A. Schön, A. Velazquez-Campoy, Isothermal titration calorimetry: general formalism using binding polynomials, Methods Enzymol. 455 (2009) 127–155, https://doi.org/10.1016/s0076-6879(08)04205-5.
[43] Y. He, S. Liu, W. Jing, H. Lu, D. Cai, D.J. Chin, A.K. Debnath, F. Kirchhoff, S. Jiang, Conserved residue Lys574 in the cavity of HIV-1 Gp41 coiled-coil domain is critical for six-helix bundle stability and virus Entry*, J. Biol. Chem. 282 (2007) 25631–25639, https://doi.org/10.1074/jbc.M703781200.
[44] X. Ran, J.E. Gestwicki, Inhibitors of protein-protein interactions (PPIs): an analysis of scaffold choices and buried surface area, Curr. Opin. Chem. Biol. 44 (2018) 75–86, https://doi.org/10.1016/j.cbpa.2018.06.004.
[45] L.M. Johnson, W.S. Horne, S.H. Gellman, Broad distribution of energetically important contacts across an extended protein interface, J. Am. Chem. Soc. 133 (2011) 10038–10041, https://doi.org/10.1021/ja203358t.
[46] Y. Zhu, X. Ding, D. Yu, H. Chong, Y. He, The tryptophan-rich motif of HIV-1 gp41 can interact with the N-terminal deep pocket site: new insights into the structure and function of gp41 and its inhibitors, J. Virol. 94 (2019), https://doi.org/10.1128/jvi.01358-19 e01358-01319.
[47] T.M. Rad, L. Saghaie, A. Fassihi, HIV-1 entry inhibitors: a review of experimental and computational studies, Chem. Biodivers. 15 (2018), e1800159, https://doi.org/10.1002/cbdv.201800159.
[48] S. Sepehri, L. Saghaie, A. Fassihi, Anti-HIV-1 activity prediction of novel Gp41 inhibitors using structure-based virtual screening and molecular dynamics simulation, Mol. Inf. 36 (2017) 1600060, https://doi.org/10.1002/minf.201600060.
[49] R. Munnaluri, S.K. Sivan, V. Manga, Molecular docking and MM/GBSA integrated protocol for designing small molecule inhibitors against HIV-1 gp41, Med. Chem. Res. 24 (2015) 829–841, https://doi.org/10.1007/s00044-014-1185-8.
[50] S.B. Jiang, A.K. Debnath, A salt bridge between an N-terminal coiled coil of gp41 and an antiviral agent targeted to the gp41 cove is important for anti-HIV-1 activity, Biochem. Biophys. Res. Commun. 270 (2000) 153–157, https://doi.org/10.1006/bbrc.2000.2411.
[51] S. Sepehri, S. Soleymani, R. Zabihollahi, M.R. Aghasadeghi, M. Sadat, L. Saghaie, A. Fassihi, Synthesis, biological evaluation, and molecular docking studies of novel 4–4-arylpyridin-1(4H)-yl benzoic acid derivatives as anti-HIV-1 agents, Chem. Biodivers. 14 (2017), e1700295, https://doi.org/10.1002/cbdv.201700295.

Supplementary Material to Publication I

Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors.

Supplementary Material to:

Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors.

Mario Cano-Muñoz[*], Samuel Jurado[†], Bertrand Morel[‡] and Francisco Conejero-Lara[*]

Departamento de Química Física, Instituto de Biotecnología y Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain.

*__Corresponding authors__:
M. Cano-Muñoz; e-mail: mariocano@ugr.es.
F. Conejero-Lara; tel. +34 958 242371; e-mail: conejero@ugr.es.

[†] Present address: Diater Laboratorio farmacéutico, Av. Peces Barba nº 2, Parque tecnológico Leganés, 28918, Spain; samueljuur@gmail.com
‡ Present address: Angany Innovation, 1 voie de l'innovation, Pharmaparc II, 27100 Val de Reuil, France; bmorel76@gmail.com

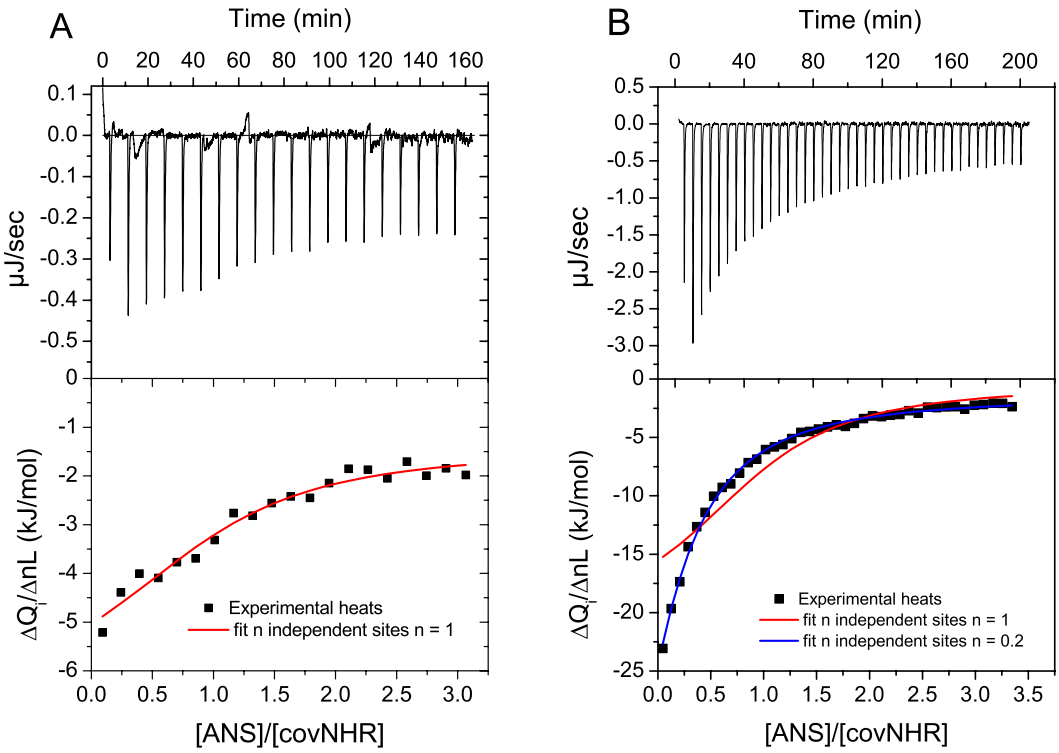


Figure S1. ITC titrations of free covNHR with ANS at 25°C. A) CovNHR at 10 µM was titrated with 420 µM ANS; B) CovNHR at 36 µM was titrated with 800 µM ANS. The upper panels show the baseline-corrected experimental thermograms. The bottom panels represent the corrected binding isotherms showing the normalized heats per mole of injected ANS as a function of the ANS-to-protein molar ratio. Symbols are the experimental values and the red solid curves correspond to the fits using a binding model of n independent and identical sites.

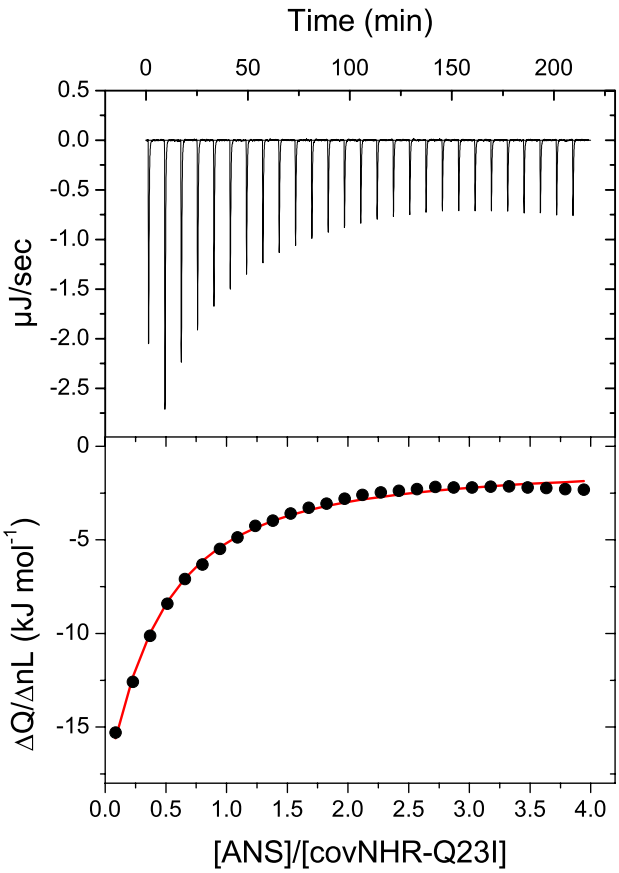


Figure S2. ITC titration of covNHR-Q23I with ANS at 25°C. The protein at 28 µM was titrated with 1.1 µM ANS. The upper panel shows the baseline-corrected experimental thermogram. The bottom panel represents binding isotherm fitted to a model of ligand binding to a protein dimer ($2M + L \leftrightarrow M_2L$).

Table S1: Apparent hydrodynamic radii (Rh) of covNHR protein and its complexes with Y24L peptide and ANS measured by dynamic light scattering at a protein concentration of ~20 µM.

Protein/complexes	Rh (nm)
covNHR	2.7
covNHR:ANS	3.1
covNHR:Y24L	2.8
covNHR:Y24L:ANS	2.7

Table S2. Interactions established with the residues at the hydrophobic pocket by both ANS molecules. Calculated with the *ListInt* command in YASARA software with a cutoff distance of 5 Å.

Hydrophobic Interactions				
First Partner	Second Partner	Residue Number in covNHR	Equivalent Residue Number in gp41	Residue Number in full envelope
ANS 1	Val	31	59	570
	Ile	34	62	573
	Lys	35	63	574
	Gln	38	66	577
	Trp	143	60	571
	Gly	144	61	572
	Gln	147	64	575
	Leu	148	65	576
ANS 2	Gln	28	56	567
	Val	31	59	570
	Leu	137	54	565
Hydrogen Bonds				
First Partner	Second Partner	Residue Number in covNHR	Residue Number in full envelope	Residue Number in full envelope
ANS 1	-	-	-	-
ANS 2	Lys	35	63	574

Supplementary Text S1: Binding models for analysis of ITC thermograms

Here we describe the mathematical development of the binding models used to analyze the ITC isotherms corresponding to the binding of ANS to covNHR proteins.

A) Binding coupled to protein dimerization.

In this model, we consider that the ligand A can bind to a protein monomer M and to a dimer M_2. The binding constants are defined by:

$$K_1 = \frac{[MA]}{[M][A]} \qquad K_2 = \frac{[M_2A]}{[M]^2[A]} \qquad (1)$$

Total concentrations of macromolecule and ligand:

$$[A]_T = [A] + [MA] + [M_2A] = [A] + K_1[A][M] + K_2[M]^2[A] \qquad (2)$$

$$[M]_T = [M] + [MA] + 2[M_2A] = [M] + K_1[A][M] + 2K_2[M]^2[A] \qquad (3)$$

This set of non-linear equations in [M] and [A] can be solved by Newton-Raphson method, given the total concentrations and the equilibrium constants. Naming [A] and [M] as x and z:

$$F_1(x,z) = x + K_1xz + K_2xz^2 - [A]_T = 0 \qquad (4)$$

$$F_2(x,z) = z + K_1xz + 2K_2xz^2 - [M]_T = 0 \qquad (5)$$

Starting from initial estimates of the variables x and z, the N-R method finds iteratively the increments in the variables dx and dz that approximate these functions to zero.
In the neighborhood of the optimal values:

$$F_i(x,z) + \frac{\partial F_i}{\partial x}\cdot dx + \frac{\partial F_i}{\partial z}\cdot dz + O(dx^2) + O(dz^2) + \ldots \qquad (i = 1, 2) \qquad (6)$$

$$\approx F_i + F'_{ix}\cdot dx + F'_{iz}\cdot dz = 0$$

Which leads to the following set of linear equations:

$$F'_{1x}\cdot dx + F'_{1z}\cdot dz = -F_1$$

$$F'_{2x}\cdot dx + F'_{2z}\cdot dz = -F_2 \qquad (7)$$

where the derivatives are calculated from equations (4) and (5):

$$F'_{1x} = 1 + K_1z + K_2z^2$$

$$F'_{1z} = K_1x + 2K_2xz$$

$$F'_{2x} = K_1z + 2K_2z^2$$

$$F'_{2z} = 1 + K_1x + 4K_2xz \qquad (8)$$

The system of equations (7) is solved for each initial estimate of x and z, to obtain dx and dz, use them to increment x and z, and repeat the process iteratively until convergence. Once converged, the free concentrations, [A] = x and [M] = z, allow us to calculate the binding parameter for this system:

$$\bar{n}_A = \frac{[A]_b}{[M]_T} = \frac{[MA] + [M_2A]}{[M] + [MA] + 2[M_2A]} = \frac{K_1[A] + K_2[M][A]}{1 + K_1[A] + 2K_2[M][A]} \qquad (9)$$

To analyze an ITC binding isotherm we need to express the average enthalpy of the macromolecule system in presence of certain ligand concentration relative to the free macromolecule:

$$\Delta H = \Delta H_1[MA] + \Delta H_2[M_2A] = [M]_T\left(\Delta H_1 F_{MA} + \frac{1}{2}\Delta H_2 F_{M2A}\right) \qquad (10)$$

Where ΔH_1 and ΔH_2 are the enthalpies of each binding process and F_{MA} and F_{M2A} are the mole fractions of M in each bound state. These fractions are given by:

$$F_{MA} = \frac{K_1[A]}{Z} \qquad F_{M2A} = \frac{2K_2[M][A]}{Z} \qquad (11)$$

and Z is the system partition function:

$$Z = 1 + K_1[A] + 2K_2[M][A] \qquad (12)$$

In each ligand injection, considered sufficiently small compared to the total cell volume, the heat involved in the binding is given by:

$$dq = V_c[M]_T\left(\Delta H_1 dF_{MA} + \frac{1}{2}\Delta H_2 dF_{M2A}\right) \qquad (13)$$

where V_c is the cell volume, and dF_{MA} and dF_{M2A} are the changes in the fraction of complexes produced by ligand binding. If we divide by the number of moles of injected ligand:

$$\frac{dq}{dnA} = \frac{1}{V_c}\frac{dq}{d[A]_T} = [M]_T\left(\Delta H_1 \frac{dF_{MA}}{d[A]_T} + \frac{1}{2}\Delta H_2 \frac{dF_{M2A}}{d[A]_T}\right) \qquad (14)$$

The derivatives in equation (14) can be calculated as:

$$\frac{dF_{MA}}{d[A]_T} = \frac{K_1}{Z}\frac{d[A]}{d[A]_T} - \frac{K_1[A]}{Z^2}\frac{dZ}{d[A]_T} \qquad (15)$$

$$\frac{dF_{M2A}}{d[A]_T} = \frac{2K_2[M]}{Z}\frac{d[A]}{d[A]_T} + \frac{2K_2[A]}{Z}\frac{d[M]}{d[A]_T} - \frac{2K_2[M][A]}{Z^2}\frac{dZ}{d[A]_T} \qquad (16)$$

$$\frac{dZ}{d[A]_T} = K_1 \frac{d[A]}{d[A]_T} + 2K_2[M]\frac{d[A]}{d[A]_T} + 2K_2[A]\frac{d[M]}{d[A]_T} \qquad (17)$$

The derivatives of [M] and [A] can be obtained by taking the derivative from the total concentrations, equations (2) and (3), as follows:

$$1 = \frac{d[M]}{d[A]_T}\left(K_1[A] + 2K_2[A][M]\right) + \frac{d[A]}{d[A]_T}\left(1 + K_1[M] + K_2[M]^2\right) \qquad (18)$$

$$0 = \frac{d[M]}{d[A]_T}\left(1 + K_1[A] + 4K_2[A][M]\right) + \frac{d[A]}{d[A]_T}\left(K_1[M] + 2K_2[M]^2\right) \qquad (19)$$

Using the values of [A] and [M] obtained above, this system of equations can be solved to obtain the derivatives and from them the normalized heats of the ITC isotherm, equation (14).

These equations were implemented in the non-linear least squares fitting tool of Origin v8.5 (Originlab, Northampton, MA) using Origin C language.

Fitting attempts were tried using the full model, with K_1, K_2, ΔH_1 and ΔH_2 as fitting parameters. However, the parameter dependency was too high and the four parameters could not be determined independently.

The curves could be fitted satisfactorily setting $\Delta H_1 = 0$ and K_1 fixed to a very low value (10^{-9}) (see main manuscript). This effectively converts the model to a single binding event with dimer formation ($2M + A \leftrightarrow M_2A$).

B) Sequential binding of two ligand molecules.

In this model two ligand molecules can bind sequentially to the same macromolecule. The binding constants for each binding step are defined by:

$$K_1 = \frac{[MA]}{[M][A]} \qquad K_2 = \frac{[MA_2]}{[MA][A]} \qquad (20)$$

And the global binding constants:

$$\beta_1 = \frac{[MA]}{[M][A]} \qquad \beta_2 = \frac{[MA_2]}{[M][A]^2} \qquad (21)$$

The total M and A concentrations:

$$[A]_T = [A] + [MA] + 2[MA_2] = [A] + \beta_1[A][M] + 2\beta_2[M][A]^2 \qquad (22)$$

$$[M]_T = [M] + [MA] + [MA_2] = [M] + \beta_1[A][M] + \beta_2[M][A]^2 \qquad (23)$$

Solving for [M] and substituting into $[A]_T$:

$$[M] = \frac{[M]_T}{1 + b_1[A] + b_2[A]^2} \qquad (24)$$

$$b_2[A]^3 + \left(b_1 + 2b_2[M]_T - b_2[A]_T\right)[A]^2 + \left(b_1[M]_T - b_1[A]_T + 1\right)[A] - [A]_T = 0 \qquad (25)$$

This cubic equation in [A] can be solved numerically to obtain [M] and [A]. The fraction of each protein state in complex, MA and MA_2 is given by:

$$F_{MA} = \frac{b_1[A]}{Z} \qquad F_{M2A} = \frac{b_2[A]^2}{Z} \qquad (26)$$

Where the partition function is:

$$Z = 1 + b_1[A] + b_2[A]^2 \qquad (27)$$

In this case the binding parameter is:

$$\bar{n}_A = \frac{[A]_b}{[M]_T} = \frac{[MA] + 2[MA_2]}{[M] + [MA] + [MA_2]} = \frac{b_1[A] + 2b_2[A]^2}{1 + b_1[A] + b_2[A]^2} \qquad (28)$$

As in model (A), we need to formulate the average enthalpy of the system relative to the free macromolecule at certain total concentrations of ligand and macromolecule:

$$\Delta H = \Delta H_1[MA] + \left(\Delta H_1 + \Delta H_2\right)[MA_2] = [M]_T\left(\Delta H_1 F_{MA} + \left(\Delta H_1 + \Delta H_2\right)F_{MA2}\right) \qquad (29)$$

Where ΔH_1 and ΔH_2 are the enthalpies of each binding step.

In each ligand injection, considered sufficiently small compared to the total cell volume, the heat involved in the binding of dn mol of ligand A to form MA and MA_2 complexes is given by:

$$dq = V_c[M]_T\left[\Delta H_1 dF_{MA} + \left(\Delta H_1 + \Delta H_2\right)dF_{MA2}\right] \qquad (30)$$

And dividing by the number of moles of injected ligand, the normalized heat evolved after each injection:

$$\frac{dq}{dnA} = \frac{1}{V_c}\frac{dq}{d[A]_T} = [M]_T\left(\Delta H_1\frac{dF_{MA}}{d[A]_T} + \left(\Delta H_1 + \Delta H_2\right)\frac{dF_{MA2}}{d[A]_T}\right) \qquad (31)$$

The derivatives of the molar fractions can be calculated as:

$$\frac{dF_{MA}}{d[A]_T} = \frac{b_1}{Z}\frac{d[A]}{d[A]_T} - \frac{b_1[A]}{Z^2}\frac{dZ}{d[A]_T} \qquad (32)$$

$$\frac{dF_{MA2}}{d[A]_T} = \frac{2b_2[A]}{Z}\frac{d[A]}{d[A]_T} - \frac{b_2[A]^2}{Z^2}\frac{dZ}{d[A]_T} \qquad (33)$$

$$\frac{dZ}{d[A]_T} = \left(b_1 + 2b_2[A]\right)\frac{d[A]}{d[A]_T} \qquad (34)$$

And taking the derivative with $[A]_T$ of the cubic equation (25) and solving for the derivative:

$$\frac{d[A]}{d[A]_T} = \frac{Z}{3b_2[A]^2 + \left(2b_1 + 2b_2[M]_T - b_2[A]_T\right)[A] + \left(b_1[M]_T - b_1[A]_T + 1\right)} \qquad (35)$$

These equations were implemented in the non-linear least squares fitting tool of Origin v8.5 (Originlab, Northampton, MA) using Origin C language.

Publication II

Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity.

biomolecules

MDPI

Article

Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity

Mario Cano-Muñoz [1,*], Samuele Cesaro [1,†], Bertrand Morel [1,‡], Julie Lucas [2], Christiane Moog [2] and Francisco Conejero-Lara [1,*]

[1] Departamento de Química Física, Instituto de Biotecnología, Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain; samuele.cesaro@univr.it (S.C.); bmorel@ugr.es (B.M.)

[2] INSERM U1109, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, 67084 Strasbourg, France; julie.lucas@etu.unistra.fr (J.L.); c.moog@unistra.fr (C.M.)

* Correspondence: mariocano@ugr.es (M.C.-M.); conejero@ugr.es (F.C.-L.); Tel.: +34-958242371 (F.C.-L.)

† Present address: Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Strada Le Grazie 8, 37134 Verona, Italy.

‡ Present address: Angany Innovation, 1 voie de l'innovation, Pharmaparc II, 27100 Val de Reuil, France.

Abstract: A promising strategy to neutralize HIV-1 is to target the gp41 spike subunit to block membrane fusion with the cell. We previously designed a series of single-chain proteins (named covNHR) that mimic the trimeric coiled-coil structure of the gp41 N-terminal heptad repeat (NHR) region and potently inhibit HIV-1 cell infection by avidly binding the complementary C-terminal heptad repeat (CHR) region. These proteins constitute excellent tools to understand the structural and thermodynamic features of this therapeutically important interaction. Gp41, as with many coiled-coil proteins, contains in core positions of the NHR trimer several highly conserved, buried polar residues, the role of which in gp41 structure and function is unclear. Here we produced three covNHR mutants by substituting each triad of polar residues for the canonical isoleucine. The mutants preserve their helical structure and show an extremely increased thermal stability. However, increased hydrophobicity enhances their self-association. Calorimetric analyses show a marked influence of mutations on the binding thermodynamics of CHR-derived peptides. The mutations do not affect however the in vitro HIV-1 inhibitory activity of the proteins. The results support a role of buried core polar residues in maintaining structural uniqueness and promoting an energetic coupling between conformational stability and NHR–CHR binding.

Keywords: HIV; AIDS; envelope glycoprotein; stability; fusion inhibitors; calorimetry; peptides; binding

Citation: Cano-Muñoz, M.; Cesaro, S.; Morel, B.; Lucas, J.; Moog, C.; Conejero-Lara, F. Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity. *Biomolecules* **2021**, *11*, 566. https://doi.org/10.3390/biom11040566

Academic Editor: Mireille Dumoulin

Received: 22 March 2021
Accepted: 9 April 2021
Published: 12 April 2021

1. Introduction

HIV-1 infection remains a global pandemic with more than 38 million of people living with the virus and near 1.7 million new infections in 2019 [1]. Despite current efforts to develop active and passive immunization strategies [2], no effective vaccines have been approved. Highly active antiretroviral therapies (HAART) have improved considerably the life expectancy of infected patients, but the emergence of multidrug-resistant viral strains [3] and the appearance of adverse effects and drug–drug interactions in some patients [4] highlight the need for new improved antivirals and microbicides to combat the infection.

Cell infection by HIV-1 is mediated by its envelope (Env) glycoprotein [5]. Env is a trimer of heterodimers with three external gp120 and three transmembrane gp41 subunits [6]. Viral attachment to cells is initiated by gp120, which binds to the CD4 cell receptor and a co-receptor. This event triggers gp120 shedding from gp41 and a large conformational change in the latter that inserts its N-terminal fusion peptide in

the cell membrane and forms an extended intermediate that bridges the cell and viral membranes [7]. Subsequently, gp41 folds onto itself forming a highly stable six-helix-bundle (6HB), composed of a central parallel trimeric coiled coil formed by the N-terminal heptad repeat region (NHR) surrounded by three antiparallel C-terminal heptad repeat (CHR) helices [8,9]. This energetically favorable folding process brings the viral and the cell membranes into close proximity facilitating their fusion and thereby the insertion of the viral material into the cell.

Blocking Env-mediated cell-virus membrane fusion has been considered for more than two decades a promising approach to inhibit HIV infection [5,10,11]. Due to its key role in the fusion process, the gp41 extended intermediate is an attractive target and has become a subject of intense research. In this intermediate, both the NHR and CHR regions of gp41 are exposed and can be accessed by a variety of molecules, including complementary peptides derived from NHR or CHR regions, artificial D-peptides, antibodies, and small-molecule compounds [12–16]. The only HIV-1 inhibitor of this type in clinical use is T20, known by its generic name as enfuvirtide [17], but its use is strongly limited by its low efficacy, short half-life, rapid acquisition of drug resistance, and high cost. Other peptides and small-molecule compounds directed against the coiled-coil structure of NHR have been developed, although very few have reached clinical stage [18].

Compared to CHR peptides, NHR-based peptide mimetics have a generally lower anti-HIV-1 activity due to their low solubility and tendency to aggregate. These problems can be alleviated by engineered protein constructs that imitate exposed NHR trimers. These constructs have shown potent inhibition activities and have certain advantages such as their activity against strains resistant to CHR inhibitors [10,19–21].

Despite all these advances, the molecular and energetic determinants of the NHR–CHR interaction are still not fully understood, making it difficult to improve existing inhibitors and develop new molecules, in particular small molecule drugs, targeting gp41.

We previously designed a series of protein molecules, called covNHR, that structurally mimic the trimer NHR helices of gp41 [22]. These proteins consist of single polypeptide chains that fold in a helix–loop–helix–loop–helix topology as antiparallel coiled-coil trimers. These covNHR proteins are very stable, bind CHR-derived gp41 peptides with high affinity, and inhibit infection through a broad variety of HIV-1 pseudoviruses and primary isolates. The crystallographic high-resolution structure of the complex between a covNHR variant and the CHR peptide C34 [23] shows a remarkable similarity with the CHR–NHR binding interface of the post-fusion structure of gp41 [9], including the presence of buried interfacial water molecules. Outstandingly, the change in the coiled-coil topology from parallel to antiparallel of the NHR trimer of helices is very well tolerated, and there is a remarkable preservation of core-packing inter-helical contacts, despite the different spatial orientation of the Cα-Cβ bond in the side chains of the antiparallel helix relative to the parallel ones.

Given that our covNHR proteins are highly soluble and stable, easy to produce in sufficient quantities, and bind the CHR peptides with 1:1 stoichiometry, they are particularly amenable for thermodynamic studies of the CHR–NHR interaction. We have recently made a thermodynamic dissection of this interaction and found that the binding energy is distributed among several binding pockets and that there is allosteric communication between them [24].

An interesting feature of the coiled coils in many proteins is the presence of non-canonical polar residues in a and d core positions of the heptad repeats, which has been related to specificity and structural uniqueness at the expense of stability [25]. The NHR coiled-coil region of gp41 contains three of such buried polar residues, namely Q41, Q51, and T58 (Q562, Q572, and T569 in full Env sequence numbering). These residues occupy positions a, d, and d in their respective heptad repeats, and their polar side chains mutually satisfy part of their hydrogen bonding potential often involving buried water molecules and backbone carbonyl oxygen atoms. In the case of the covNHR protein, the buried polar residues are Q13, Q99, and Q124 (structurally equivalent to Q41); Q23, Q89, and Q134 (equivalent to Q51); and T30, T82, and T141 (equivalent to T58). In our published structure

(Protein Data Bank (PDB) id. 6R2G), a water molecule occupies a small cavity between the side chains of the Q23, Q89, and Q134 triad. Although these hydrogen bonds are expected to be stronger in the apolar environment of the coiled-coil core [26], their number is smaller than in a fully hydrated environment. Moreover, there is a considerable entropy penalty due to the geometric constraints imposed by the formation of these hydrogen bonds and in the immobilization of the buried water molecules. In contrast, canonical hydrophobic residues in trimeric coiled coils (Ile, Val, Leu) are expected to be energetically favored in core positions. This was demonstrated in a study of an NHR(L6)CHR construct of the simian immunodeficiency virus (SIV) gp41 core [27], in which the polar core residues were individually mutated to isoleucine. The authors found that core polar-to-hydrophobic mutations have strongly stabilizing effects. However, since the constructs used were mimics of the post-fusion 6HB conformation, the energetic consequences of these core mutations on the CHR–NHR interaction cannot be separated from their effects on the central NHR coil stability.

We hypothesized that buried polar residues may be involved in the mechanisms by which the NHR–CHR interaction is regulated. Since the covNHR proteins constitute an excellent model to investigate the molecular and energetic details of this interaction, we produced three triple mutants (Figure 1), in which each triad of polar residues was replaced by isoleucine. We selected isoleucine over leucine because trimeric coiled coils can accommodate beta-branched amino acids at both a and d positions of the heptad repeat due to their acute knobs-into-holes packing, whereas leucine is favored at positions with perpendicular packing present in other helical arrangements [28]. For the sake of simplicity, each mutant was named after the first sequence position mutated. We investigated the stability of the three variants in comparison to the reference protein and their capacity to recognize their CHR target sequence, as well as their inhibitory activity against in vitro HIV-1 infection. The results indicate significant influence of the core mutations upon the CHR–NHR interaction.

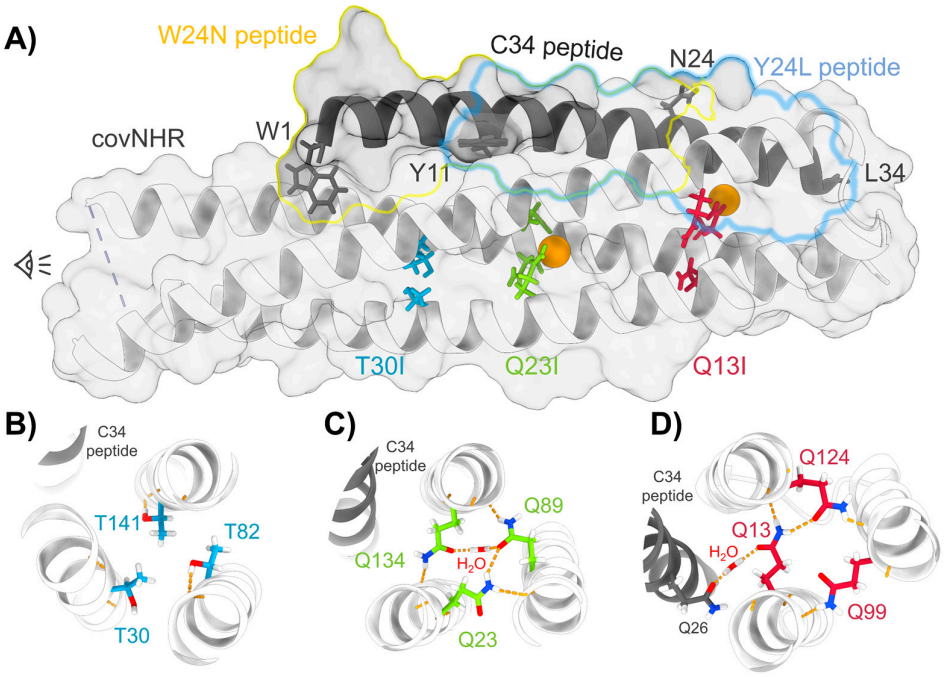


Figure 1. Buried polar amino acids in the coiled-coil structure of covNHR. (**A**) Representation of the X-ray structure of the complex of covNHR with W34L (C34) (PDB id. 6R2G) [23]. The limits of the W24N and Y24L peptides are highlighted on the structure with yellow and blue lines, respectively. The buried polar side chains are represented with sticks and buried water molecules with orange spheres. (**B**–**D**) Cross-sectional representations of each triad of polar side chains showing hydrogen bonds in dashed yellow lines.

2. Materials and Methods

2.1. Proteins and Peptides

The DNA encoding protein sequences were synthesized and cloned into pET303 expression vectors by Thermo Fisher Scientific (Waltham, Massachusetts, MA, USA). To facilitate purification by Ni-Sepharose affinity chromatography, the protein sequences included a C-terminal histidine tag with sequence GGGGSHHHHHH. The sequences are shown in Table S1. Each triple mutant was named as covNHR followed by the identity of first mutation, i.e., covNHR-Q13I, covNHRQ23I, and covNHR-T30I. The proteins were overexpressed in *E. coli* and purified, as described previously [22]. The protein purity was assessed by SDS-PAGE, and the identity of each protein variant was confirmed by mass spectrometry analysis.

Synthetic CHR peptides (Table S1), both N-acetylated and C-amidated, were acquired from Genecust (Mondorf-les-Bains, Luxembourg), with a purity >95%. Protein and peptide concentrations were measured by UV absorption measurements at 280 nm with extinction coefficients calculated according to their respective amino acid sequences with the ExPasy ProtParam server (https://web.expasy.org/protparam/, accessed on 3 March 2020) [29].

2.2. Circular Dichroism

Circular dichroism (CD) measurements were carried out in a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier thermostatic cell holder. Measurements of the far-UV CD spectra (260–200 nm) were made with a 1 mm path-length quartz cuvette at a protein concentration of ~15 μM. Buffer conditions were 50 mM sodium phosphate pH 7.4. Spectra were obtained by averaging 5 scans recorded at a scan rate of 100 nm/min, 1 nm step resolution, 1 s response, and 1 nm bandwidth. Thermal unfolding was monitored by measuring the CD signal at 222 nm as a function of temperature using a scan rate of 1 °C min^{-1}.

2.3. Light Scattering

The apparent hydrodynamic radii of the proteins were measured using dynamic light scattering (DLS) in a DynaPro MS-X DLS instrument (Wyatt, Santa Barbara, CA, USA). Dynamics v6 software (Wyatt Technology Corporation, Santa Barbara, CA, USA) was used in data collection and processing. Sets of DLS data were measured at 25 °C with an average number of 50 acquisitions and an acquisition time of 10 s. Measurements were carried out in 50 mM sodium phosphate buffer pH 7.4 and 50 mM glycine/HCl buffer pH 2.5.

Static scattering intensities were measured in a Malvern μV instrument (Malvern Panalytical, Malvern, UK) at 25 °C, in 50 mM sodium phosphate buffer pH 7.4, at different concentrations of protein in a range of 0.2 to 3.5 mg mL^{-1}. The intensities were analyzed using the Debye plot as represented by Equation (1),

$$K_C/R_{90} = 1/M_W + 2A_2c \quad (1)$$

valid for particles significantly smaller than the wavelength of the incident radiation, where the K is an optical constant of the instrument, c is the particle mass concentration, R_{90} is the Rayleigh ratio of scattered to incident light intensity, M_w is the weight-averaged molar mass, and A_2 is the 2nd virial coefficient that is representative of inter-particle interaction strength. M_w can be determined from the intercept of the plot.

2.4. Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) experiments were carried out in a MicroCal PEAQ-DSC microcalorimeter equipped with an autosampler (Malvern Panalytical, Malvern, UK). Scans were run from 5 to 130 °C at a scan rate of 120 °C·h^{-1}. The experiments were carried out in 50 mM sodium phosphate buffer pH 7.4 and 50 mM glycine/HCl buffer pH 2.5. Protein concentration was typically 40 μM. Instrumental baselines were recorded before each experiment with both cells filled with buffer and subtracted from

the experimental thermograms of the protein samples. Consecutive reheating runs were carried out to determine the reversibility of the thermal denaturation. The partial molar heat capacity (C_p) was calculated from the experimental DSC thermograms using Origin software (OriginLab, Northampton, MA, USA). Reversible unfolding thermograms were fitted using two-state N ↔ U or three-state N ↔ I ↔ U unfolding models. In case of irreversible unfolding, we used a Lumry–Eyring denaturation model [30] (N ↔ U → F) to fit the experimental thermograms. The DSC thermograms measured with mixtures of covNHR proteins and CHR peptides were analyzed with a 1:1 binding model (N + L ↔ NL) coupled to the Lumry–Eyring model, as described elsewhere [23].

2.5. Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) measurements were carried out in a Microcal VP-ITC calorimeter (Malvern Panalytical, Malvern, UK). The proteins were titrated with 25 injections of 5 μL peptide solution at 480 s intervals. Protein concentration in the cell was around 10 μM, while the peptide concentration in the syringe was typically 200–300 μM. The experiments were carried out in 50 mM phosphate buffer (pH 7.4) at 25 °C. The experimental thermograms were baseline corrected, and the peaks were integrated to determine the heat produced by each ligand injection. Heat of dilution of the peptides was measured in independent titrations with only buffer in the calorimeter's cell and subtracted from the heat of the titration of the protein. Finally, each heat was normalized per mole of injected ligand. The resulting binding isotherms were fitted using a binding model of independent and equivalent sites, allowing the determination of the binding constant, K_b; the binding enthalpy, ΔH_b; and the binding stoichiometry, n. From these values, the standard Gibbs energy and entropy of binding could be derived as $\Delta G_b = -RT \ln K_b$ and $\Delta S_b = (\Delta H_b - \Delta G_b)/T$. In some experiments, the binding model was corrected from the influence of protein self-association, as described in the Results.

2.6. HIV-1 Inhibition Assays

The inhibition of HIV-1 replication was determined using the conventional TZM-bl assay measured as a function of reductions in Tat-regulated Firefly luciferase (Luc) reporter gene expression [31]. The viruses used for TZM-bl cell infection were HIV SF162 strain (Tier 1, easy to neutralize subtype B strain) and CE1176 (Tier 2, difficult to neutralize subtype C strain). The IC50, the concentration (in nM) of inhibitor inducing a 50% decrease in relative luminometer units (RLU), corresponding to a 50% decrease in virus replication compared to the control, was calculated by non-linear regression using a sigmoidal Hill function, as implemented in Origin software (v.8.5, Originlab, Northampton, MA, USA).

3. Results

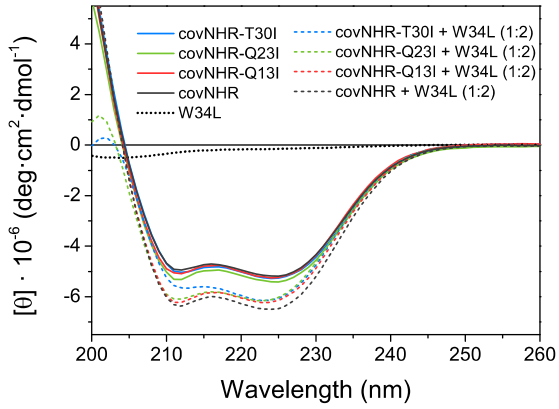
3.1. Structure and Stability of the Mutant Proteins

All the covNHR core mutants were expressed in *E. coli* with good yields and showed high solubility, similarly to the parent covNHR protein. CD spectra (Figure 2a) indicate highly α-helical structures with no significant differences between the four variants, indicating that the core mutations do not alter the coiled-coil structure.

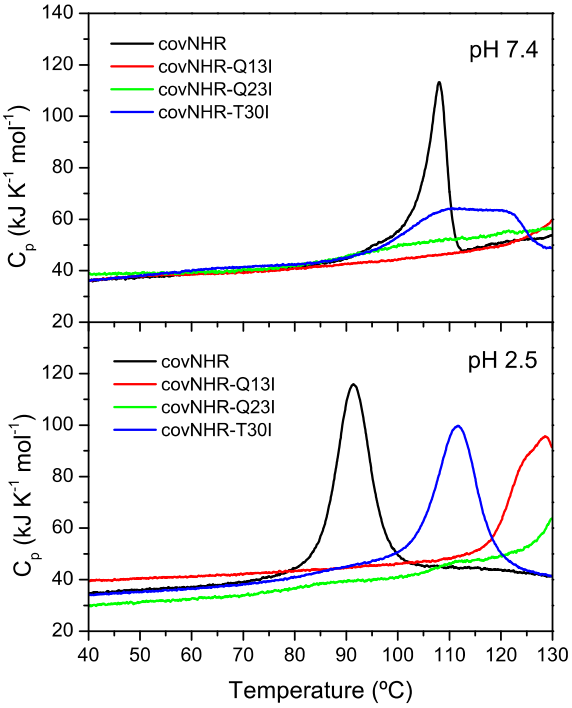
To characterize the effects of the mutations on the stability of the proteins, thermal scans were monitored with CD at pH 2.5 (50 mM glycine-HCl buffer) and pH 7.4 (50 mM sodium phosphate buffer) (Figure S1). At both pH values the three mutant proteins were highly stable and the denaturation transitions fall outside the temperature range of the CD scans (up to 98 °C).

DSC scans were carried out up to 130 °C (Figure 2b). CovNHR-T30I was significantly more stable than covNHR. Strikingly, covNHR-Q13I and especially covNHR-Q23I showed extreme thermostability, with denaturation temperatures near or well above 130 °C. This indicates that the stability penalty of burying polar side chains within the coiled-coil core is larger for Gln than for Thr. At pH 2.5, the DSC reheating indicated partial reversibility of the unfolding transitions, whereas at pH 7.4 the denaturation profiles were fully irreversible.

The DSC thermograms of covNHR and covNHR-T30I at acid pH could not be accurately fitted with the two-state unfolding model, but they could be fitted very well with a model of two sequential unfolding transitions (N ⇄ I ⇄ U), as described previously [22] (Figure S2). At pH 7.4 irreversible denaturation profiles were analyzed using a Lumry–Eyring model N ⇄ I → F) [23]. The thermodynamic parameters of the thermal denaturation processes are shown in Table S2. The results indicate that the Thr to Ile mutations reduce the unfolding cooperativity.



(**a**)



(**b**)

Figure 2. Structure and stability of the covNHR mutants. (**a**) Circular dichroism (CD) spectra of the free proteins and in complex with W34L peptide. The spectrum of the free peptide is also shown for reference. The spectra are normalized in ellipticity units per mole of protein. (**b**) Differential scanning calorimetry (DSC) thermograms of covNHR proteins at pH 2.5 and pH 7.4.

3.2. Binding of CHR Peptides

The three core mutants bind the W34L peptide (gp41 residues 117–150, more commonly known as C34), which acquires α-helical structure according to the CD spectra of the protein–peptide mixtures (Figure 2a). To elucidate the binding stoichiometry and determine the thermodynamic parameters, ITC titrations of the proteins were carried out with W34L. The binding isotherms (Figure 3) showed highly exothermic binding with very sharp sigmoidal shapes indicating high affinity and approximately 1:1 stoichiometry, similarly to that observed for the reference covNHR protein [23]. The data could be fitted using a binding model of n independent and identical sites. The binding parameters are given in Table 1. Compared to covNHR, the affinity of which for W34L was so high that it could not be assessed by ITC titration [23], the covNHR-Q13I mutant shows apparent reduction in binding enthalpy and affinity. However, the binding affinities of the Q23I and T30I variants remain very high and fall out of the upper limit of the values accessible by direct ITC titrations.

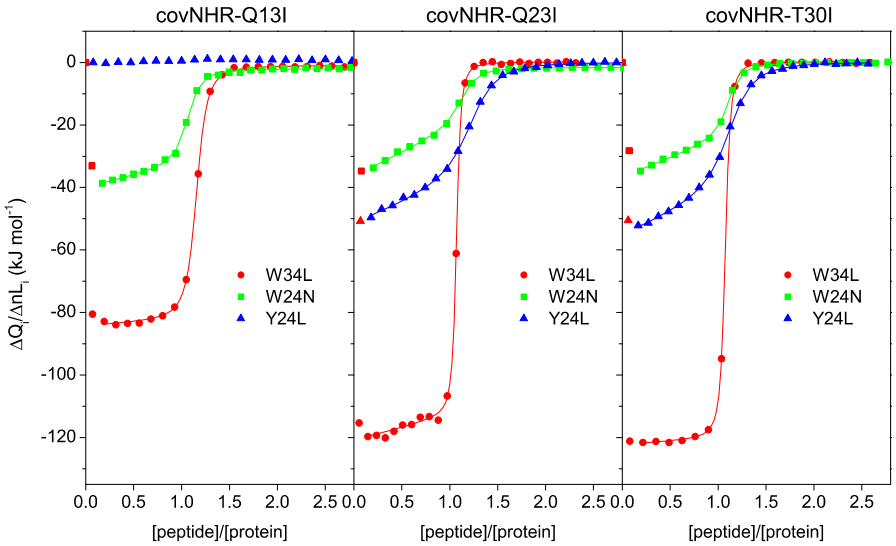


Figure 3. Binding of the covNHR mutants to C-terminal heptad repeat (CHR) peptides. Binding isotherms determined by isothermal titration calorimetry (ITC) for each protein to different peptides as indicated in the figure. The symbols represent the experimental heat normalized per mole of injected ligand, and the curves represent the best fittings obtained with the binding models as described in the text.

Therefore, to better estimate the binding affinities, we employed an approach based on DSC, as previously used for covNHR [23], in which we analyzed the protein–peptide mixtures at different molar ratios observing the influence of the thermally induced dissociation on the DSC thermograms. Since binding is highly exothermic, dissociation is endothermic and will be forced by heating. The resulting thermograms for the three mutants are shown in Figure 4. Interestingly, due to the extreme stability of the Q13I and Q23I mutants, the peak corresponding to peptide dissociation induced by heating becomes fully separated from the denaturation transitions of the proteins, which occur at much higher temperatures. In the case of the T30I mutant, the peaks corresponding to peptide dissociation and protein denaturation partially overlap. The curves were globally fitted with a 1:1 equilibrium binding model (N + L $\leftrightarrow$ NL) coupled to the Lumry–Eyring denaturation model (N $\leftrightarrow$ I $\rightarrow$ F), as described previously [23]. The binding enthalpy measured by ITC at 25 °C was fixed in the fittings. The resulting binding affinities are detailed in Table 1. Strikingly, the binding constants estimated with this procedure are considerably higher than those estimated by ITC, even up to five orders of magnitude, as it is the case of the Q23I mutant. These dramatic differences are unlikely to be due to experimental errors of the calorimetric experiments, but they likely reflect some type of anomalous binding behavior that is not

accounted for by the simple binding models used to analyze the data quantitatively, as discussed later in the paper.

Table 1. Thermodynamic parameters of binding of gp41 CHR peptides to the covNHR proteins determined by isothermal titration calorimetry (ITC) or differential scanning calorimetry (DSC).

Protein Variant	Peptide	Technique	K_b (M^{-1})	K_d (nM)	ΔH_b (kJ·mol^{-1})	n
covNHR [1]	W34L	ITC	Not measurable		−128.8 ± 0.3	1.01
		DSC	$(1.32 \pm 0.03) \times 10^{13}$	0.000076 ± 0.000002		
	W24N	ITC	$(1.25 \pm 0.11) \times 10^{7}$	80 ± 7	−41 ± 0.3	0.83
	Y24L	ITC	$(1.11 \pm 0.09) \times 10^{7}$	90 ± 7	−63 ± 1.0	0.83
covNHR-Q13I	W34L	ITC	$(4.4 \pm 0.3) \times 10^{7}$	22.7 ± 1.3	−81.1 ± 0.6	1.16
		DSC	$(2.62 \pm 0.05) \times 10^{9}$	0.382 ± 0.007		
	W24N	ITC	$(2.3 \pm 0.3) \times 10^{7}$	44 ± 5	−29.9 ± 0.6	1.07
	Y24L	ITC	Undetectable binding			
covNHR-Q23I	W34L	ITC	$(2.0 \pm 0.3) \times 10^{8}$	5.1 ± 0.7	−112.4 ± 0.8	1.07
		DSC	$(2.96 \pm 0.06) \times 10^{13}$	0.000034 ± 0.000001		
	W24N	ITC	$(1.61 \pm 0.23) \times 10^{7}$	62 ± 9	−17.9 ± 0.6	1.14
	Y24L	ITC	$(7.0 \pm 0.5) \times 10^{6}$	142 ± 9	−36.3 ± 0.8	1.24
covNHR-T30I	W34L	ITC	$(1.58 \pm 0.08) \times 10^{8}$	6.3 ± 0.3	−120.4 ± 0.5	1.08
		DSC	$(2.6 \pm 0.3) \times 10^{12}$	0.00038 ± 0.00004		
	W24N	ITC	$(2.48 \pm 0.24) \times 10^{7}$	40 ± 4	−23.0 ± 0.4	1.12
	Y24L	ITC	$(7.2 \pm 0.4) \times 10^{6}$	139 ± 8	−40.4 ± 0.7	1.14

[1] Data from refs. [23,24].

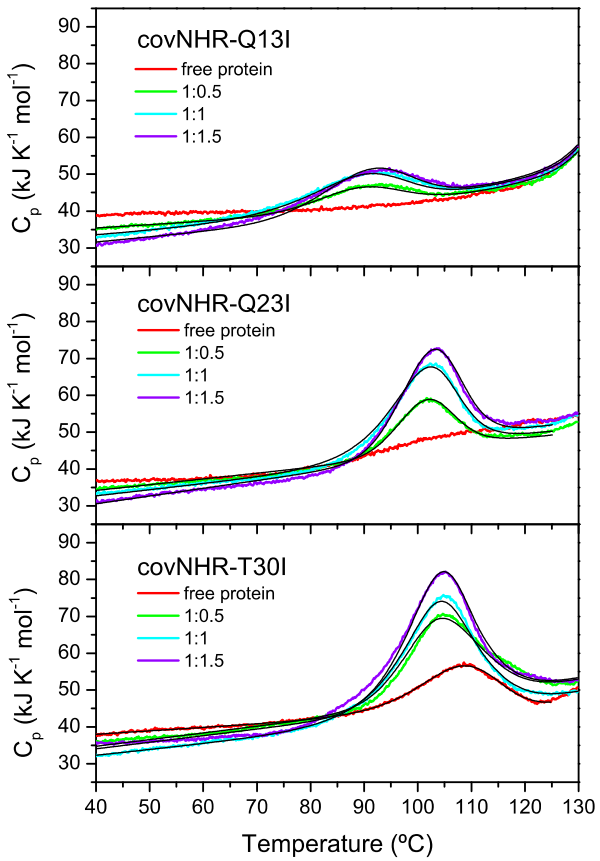


Figure 4. DSC analysis of thermally induced dissociation of the complexes of covNHR mutants with W34L. Protein–peptide mixtures were analyzed at the indicated molar ratios. The experimental data are represented in colors. The black solid lines represent the best fittings using the model of ligand binding coupled to a Lumry–Eyring denaturation.

We previously reported that the high affinity between the NHR and CHR regions is a result of cooperative contributions of several binding pockets along the NHR crevice [24]. To elucidate whether the observed changes in binding affinity produce regional effects at different pockets, we investigated the binding of shorter CHR peptides that encompass only part of the binding determinants. We carried out ITC titrations with peptide Y24L (gp41 residues 127–150), which binds to the N-terminal polar pocket (NTP) and a middle pocket (MP) of the NHR crevice, and peptide W24N (residues 117–140), which covers the MP and the deep hydrophobic pocket (HP). The binding isotherms (Figure 3) also indicate 1:1 stoichiometry for all mutants and much lower binding enthalpies than those observed for W34L, which contains the binding determinants of the three pockets. Interestingly, the Y24L peptide does not bind to the Q13I mutant, indicating that these buried glutamine side chains play an important role in the interactions at the NTP site. Despite the 1:1 stoichiometry, the ITC profiles do not match the typical sigmoidal shape expected for a simple 1:1 binding equilibrium and show a significant slope at the beginning of the titrations. These effects suggest additional processes coupled to the main binding event that alter the shape of the binding isotherm.

3.3. Oligomerization State

Since the core mutations increase the hydrophobicity of the mutant proteins, it is possible that they could also enhance their self-association propensity compared to the parent protein and this may have an influence on the binding to the peptides. To explore this possibility, we measured the apparent hydrodynamic radii (R_h) by DLS (Table S3). The three core mutants show R_h values slightly larger than the covNHR reference protein, which shows a value expected for a monomer. This suggests that the mutant proteins have a propensity for self-association. To confirm this hypothesis, we measured the scattering intensities as a function of the protein concentration and made the corresponding Debye plots (Figure 5). The three mutant proteins show approximately linear tendencies with intercepts in good agreement with $1/M_w$ of the dimer (Figure 5, upper panel). In contrast, the reference protein departs at low concentration from this tendency and shows an intercept in agreement with a monomeric state. The results confirm a significantly increased self-association as dimers of the core mutants compared to the reference protein, explaining the anomalous ITC profiles observed for the core mutants. The scattering measurements with the covNHR-Q23I mutant in complex with the W34L, at a 1:2 protein/peptide ratio, show and intercept in the Debye plot corresponding to a molecular weight of about 32 kDa, closer to the expected M_w of the complex (25 kD) than to that of the dimer (50 kDa). These results indicate a considerable reduction in molecular mass indicating that peptide binding shifts the dimerization equilibrium towards the monomeric state, in good consistency with the 1:1 binding stoichiometry observed by ITC (Figure 3). This implies that self-association perturbs peptide binding, probably by occluding or distorting the NHR binding crevice.

To account for the influence of self-oligomerization on the ITC binding isotherms of the W24N and Y24L peptides, we fitted the curves using a model of binding coupled to a dimerization equilibrium, assuming that the peptide can only bind to the monomer. This model described very satisfactorily the ITC isotherms (Figure 4) although, due to over parameterization, the dimerization constants and the association enthalpies show very strong dependency and could not be determined independently. We found good fittings using fixed dimerization constants of $1 \times 10^3\ M^{-1}$, corresponding to a mole fraction of dimers of about 0.5 under the conditions of the ITC experiments. Nevertheless, the peptide binding parameters have a very low dependency with the parameters of the dimerization, allowing reasonable estimates of the former from these analyses. In contrast, self-association has a very small influence on shape of the ITC isotherm for titrations with W34L and fit almost equally well to a simple 1:1 binding model.

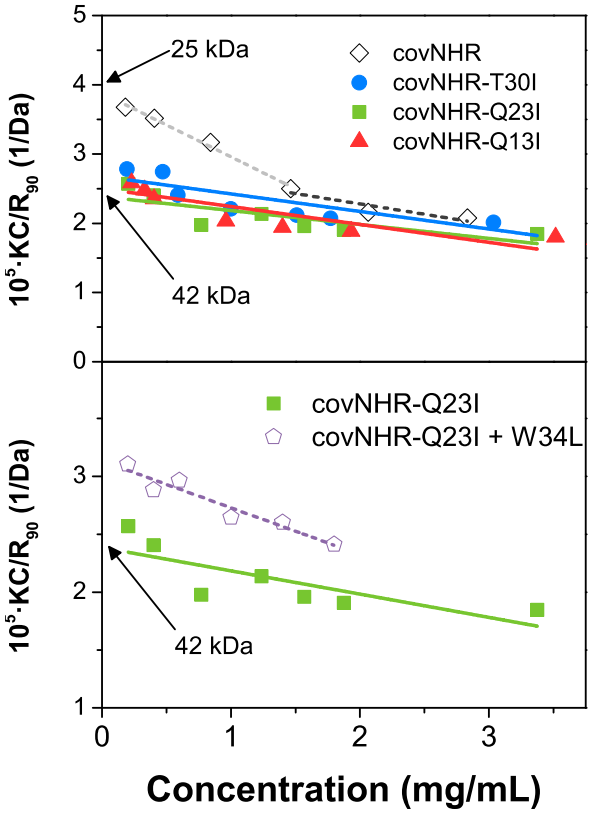


Figure 5. Molecular size of the covNHR proteins. Upper panel: Debye plots corresponding to the scattering intensities of protein solutions at different concentrations for the covNHR reference protein and the triple mutants. The intercepts of the linear tendencies indicate the weight averaged molar mass of the molecules. Bottom: Debye plots corresponding to the covNHR-Q23I mutant in its free form and in complex with the W34L peptide.

The parameters of Table 1 indicate that peptides Y24L and W24N, which only occupy two pockets on the NHR crevice, have a much lower affinity than W34L, which covers the three pockets. The core mutations generate significant decreases in the binding enthalpies compared to the parent covNHR protein but, in general, produce relatively small changes in the binding affinity for the short peptides, except in the case of the complete abolishment of binding of Y24L to covNHR-Q13I. The loss of favorable binding enthalpy is therefore compensated by a lower entropy cost of binding, as a result of the increase in hydrophobicity produced by the mutations.

A striking result is the large discrepancy between the affinity constants of W34L derived from direct ITC experiments and the dissociation constants obtained by DSC of protein–peptide mixtures. The difference is particularly outstanding for the Q23I mutant, for which there is a difference of 5 orders of magnitude. ITC measures the heat released upon complex formation between the free protein and peptide, whereas the DSC experiments monitor the heat of dissociation of a pre-formed complex at high temperature. It is possible that these processes are not just opposite steps of the same reversible binding equilibrium. Other factors, such as the observed dimerization propensity of the proteins and the existence of kinetic limits to the binding-dissociation processes may also influence the apparent binding constants and produce inconsistent values between the two methods.

To attempt to shed light on this inconsistency, we carried out an ITC displacement experiment, in which we placed covNHR-Q13I and W34L at 1:2 molar ratio in the calorimetric cell and titrated with the reference covNHR protein from the syringe. According to the large difference in affinity and binding enthalpy, if the proteins were competing for the peptide in simple reversible binding events, the binding profile of this experiment would yield two sigmoidal steps, one at 1:0.5 covNHR/W34L stoichiometry showing the binding

of covNHR to the excess free W34L peptide, and a second step at 1:1 covNHR/W34L stoichiometry corresponding to the displacement of the Q13I mutant from binding of the remaining W34L peptide, with heat corresponding to the net enthalpy balance of the processes. In contrast, the experiment did not show such profile and only a first binding step with an approximate 1:0.5 stoichiometry (Figure S3), indicating that covNHR cannot displace the Q13I mutant from the complex with W34L within the time frame of the ITC titration at 25 °C. This points to a slow dissociation of the peptide that limits the capacity of ITC to estimate these extremely high affinity constants.

3.4. HIV-1 Inhibition Activity

Finally, we aimed to investigate how the core mutations affected the capacity of the proteins to inhibit the HIV-1 infection in vitro. The four proteins were tested in TZM-bl neutralization assays using SF162 and CE1176 pseudoviruses (Figure S4). The IC50 values are given in Table 2. All proteins show a strong inhibitory activity in the low nanomolar range with insignificant differences between them. It is interesting that the Q13I mutant keeps a similar inhibitory potency to the other proteins, despite its considerably decreased affinity for the CHR region produced by perturbation of the interactions at the NTP pocket. The IC50 values are similar between the different proteins but much higher than the dissociation constants of the complexes with W34L derived from DSC analysis. There is no correlation between the inhibitory potency and the peptide binding affinity.

Table 2. HIV-1 inhibition activity of covNHR proteins measured with the standard TZM-bl assay using SF162 (subtype B) and CE1176 (subtype C) pseudoviruses. Values correspond to IC50 in nM units.

Protein Variant	SF162	CE1176
covNHR	8.0 ± 1.3	10.8 ± 0.9
covNHR-Q13I	8.9 ± 1.3	10.7 ± 0.8
covNHR-Q23I	7.1 ± 0.9	5.3 ± 1.1
covNHR-T30I	8.6 ± 0.5	7.4 ± 0.05

4. Discussion

Here we show how replacing buried polar Thr or Gln sidechains for Ile in three core positions of the covNHR protein leads to a very strong increase in thermal stability. The stability increase is much larger for Gln than Thr substitution. This is consistent with higher hydrogen bonding potential of the Gln side chain that cannot be fully satisfied in the core interior compared to the water environment. Although hydrogen bonds buried in the apolar interior are particularly strong [26], there is considerable energy penalty compared to the hydrophobic and van der Waals interactions between Ile side chains.

An immobilized water molecule was found hydrogen bonded to the Q23, Q89, and Q134 side chains in the crystallographic structure of the covNHR protein in the complex with W34L. These hydrogen bonds may add an additional enthalpic component to the stability of the covNHR protein. However, the large entropy cost of immobilizing this water molecule adds a considerable energy penalty that may explain why the Q23I variant is more stable than the Q13I one.

The Q13, Q99, and Q124 triad only establish hydrogen bonds with groups from opposite helices and do not hold any buried water molecule in the center of the NHR coiled coil. However, the Q13 side chain is implied in a hydrogen bond with a buried water molecule at the NHR–CHR interface (Figure 1), which is part of a hydrogen bond network at the NTP [23]. This has a strong implication for peptide binding as discussed below.

T30, T82, and T141 sidechains are fully buried in the crystallographic structure, but neither hydrogen bonding nor immobilized water are associated with these side chains. Thus, the increased stability observed for the Ile mutants of these residues should be ascribed to improved internal packing and enhanced hydrophobic interaction.

Lu and coworkers used a recombinant model of the SIV gp41 6HB conformation, designated N36(L6)C34, to investigate the role of the similar buried polar residues in SIV gp41 [27]. Specifically, they studied equivalent mutations to those described here, i.e., Q565I, Q575I, T582I, and T586I. They reported 15–20 °C increases in the unfolding temperatures, which are lower than the variations observed here for the Gln-Ile mutants. This difference may be caused by a lower symmetry of the core interactions of the covNHR proteins compared to the trimeric N36(L6)C34 constructs, allowing less optimal hydrogen bonding of the Gln side chains. Their Q575I variant (equivalent to covNHR-Q23I in this work) was insoluble in pure physiological buffer but soluble in presence of 1.5 M GdmHCl. Our results also show that the increased hydrophobic character produced by the mutations favors self-association of the proteins into dimeric species. These data support the hypothesis that buried polar interactions contribute to structural specificity and uniqueness at a cost of stability [28,32,33].

Nevertheless, the N36(L6)C34 constructs do not make it possible to discriminate between the effects of the mutations on the stability of the NHR trimeric coils and their influence on the interactions with the CHR region. Here we show that each set of mutations plays a different role in the NHR–CHR interaction.

Each of the triads of the mutated residues is a structural part of the regions harboring different pockets at the binding interface between covNHR and the W34L peptide (Figure 1). The Q13 triad participates in the NTP, the Q23 triad is part of the MP, and the T30 triad pertains to the HP [24]. Each set of mutations to Ile should therefore influence the conformational properties of each pocket and thereby the binding interactions with CHR peptides. The effects, however, are not restricted locally but extend to other pockets.

The strongest effects are observed for covNHR-Q13I, which shows undetectable affinity for the Y24L peptide and has a strongly decreased affinity for the W34L peptide. This is a local effect that produces also a considerable reduction in negative binding enthalpy by about one third, indicating a decrease in the favorable balance of interactions. This is justified by the disruption of the extensive network of interactions at the NTP, in which an interfacial water molecule bridges polar residues between NHR and CHR (Figure 1d). These conserved glutamine residues have a role not only in determining structural specificity of the NHR coiled coil but also in affinity and specificity of the NHR–CHR interaction. It has been shown that buried networks of hydrogen bonds in coiled coils are very important in determining the folding and oligomeric assembly in apolar environments [34] and can extend to outer shells of concentric helices, completely determining the coiled-coil topology [35].

The Q23I and T30I sets of mutations slightly reduce the binding affinity for the Y24L peptide. In contrast, the three triads of polar-to-Ile mutations produce small increments in the binding affinity for the W24N peptide. All these effects are accompanied by considerable reductions in favorable binding enthalpy, which are compensated by increased entropy of binding (Figure 6), resulting in strong enthalpy–entropy compensation. Since the mutated residues do not participate directly in the binding interface, except Q13, the only mechanism that can explaining these thermodynamic changes is a strong coupling between binding and structural tightening, as previously proposed [23].

We have recently reported that the NHR coiled coil is composed of two subdomains with different stability [36]. The N-terminal half, containing the NTP and MP pockets, is intrinsically unstable, whereas the C-terminal part harboring the HP pocket shows a much higher stability. It is interesting to notice how the highly stabilizing core mutations studied here enhance binding at the C-terminal part while reducing affinity at the N-terminal part, especially the Q13I mutations. Replacement of polar residues for isoleucine appears therefore to diminish the energetic linkage between CHR binding and stability of the N-terminal subdomain. A possible reason for these observations is that the gp41 NHR sequence may be finely tuned to balance the stability of the NHR trimer of helices in order to avoid exposure of a too stable binding groove that would expose persistently conserved epitopes susceptible to neutralization.

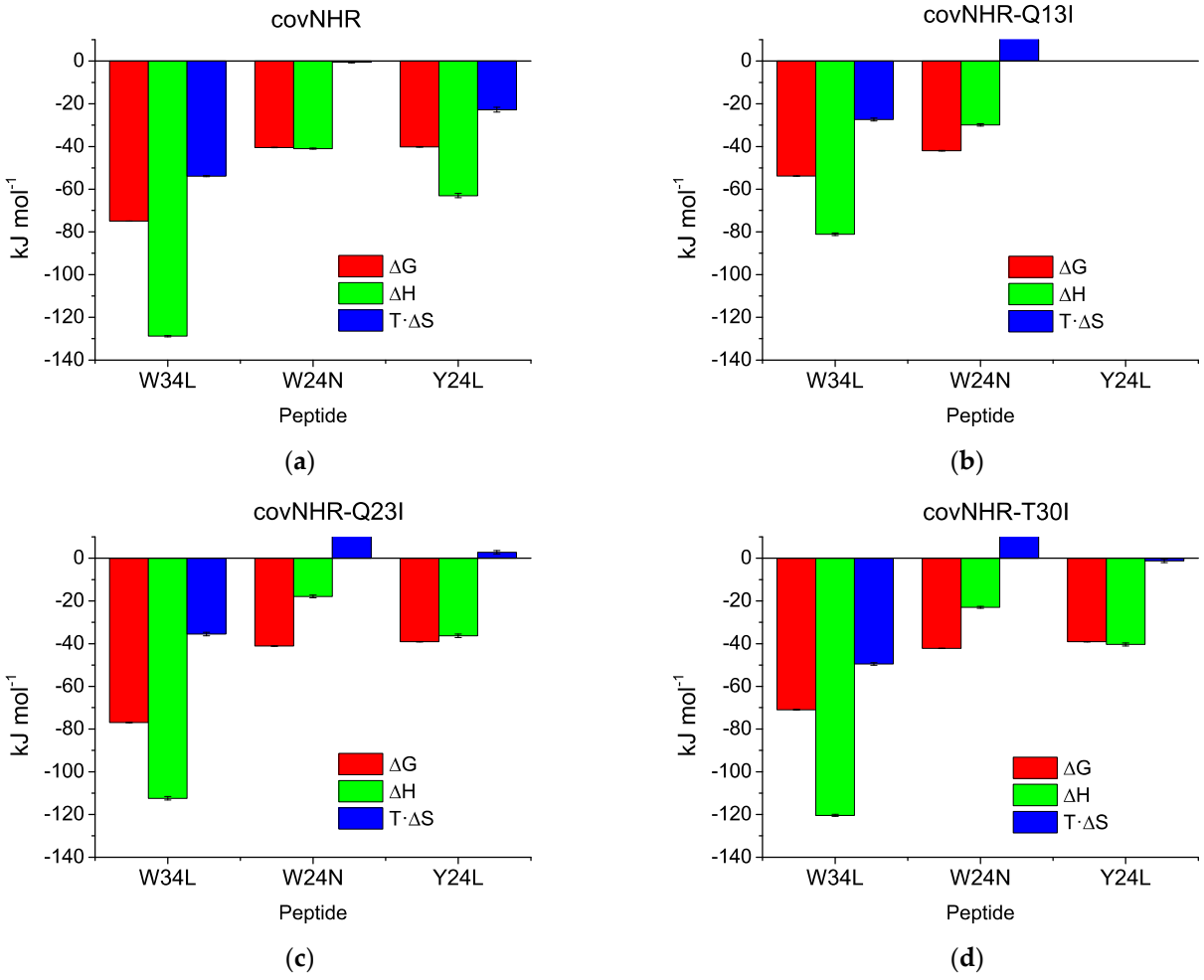


Figure 6. Thermodynamic signature of the binding of CHR peptides to the covNHR proteins. Enthalpy, entropy, and Gibbs energy of binding were obtained from parameters listed in Table 1. (**a**) Reference covNHR protein; (**b**) covNHR-Q13I; (**c**) covNHR-Q23I; (**d**) covNHR-T30I.

Despite the observed effects in stability and binding affinity for the CHR target peptides, the in vitro inhibitory activity of all the protein variants is quite similar. This result is in agreement with our previous study of a mutant covNHR protein that also showed abolished binding at the NTP but enhanced binding at the HP [23], similarly to what is observed here for the Q13I mutant. We have previously shown that strong inhibition activity can be achieved with a minimum of two NHR pockets (MP + HP or MP + NTP) [23,24]. However, the inhibitory activity of the covNHR proteins is quite uncorrelated with the overall binding affinity for CHR. It has been proposed that inhibitor access to gp41 CHR is kinetically limited by a slow and transient exposure of the CHR target to interaction with external NHR binders such as 5-helix [37], so that IC50 values in inhibition assays are actually inversely correlated with the association rates [38], which are similar for different mutant proteins that have binding affinities spanning several orders of magnitude. This explains why the protein variants studied here show a similar inhibitory activity.

This and previous work [22,23,36] show that the covNHR proteins inhibit a broad variety of HIV-1 strains with high potency, which highlights their strong potential to be developed as therapeutic tools against HIV-1. In the development of a peptide- or protein-based antiviral, it is important not only to achieve high inhibitory potency but also other properties such as high solubility, ease of production and purification, high stability, and structural uniqueness. These are key aspects in drug formulation, as well as for appropriate bioavailability and pharmacokinetics. For instance, T20 (enfuvirtide) treatment requires

a very high dosage due to its short half-life in the blood stream, which limits its clinical use [39]. In another example, T-1249, a promising enfuvirtide successor, was discontinued due to formulation difficulties [40]. Therefore, improving the conformational and stability properties of protein-based drug candidates is crucial. The core mutations described here produce strong stability increases without affecting significantly their high solubility and high production yield. However, the mutations enhance their propensity for self-association, which interferes with the potential to recognize their target. Additional studies in vivo will be necessary to establish whether or not the strategy of protein stabilization by mutating core polar residues results in better drug candidates against HIV-1.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/biom11040566/s1, Figure S1: Thermal scans monitored by CD; Figure S2: Fitting of DSC thermograms; Figure S3: ITC displacement experiment; Figure S4: In vitro inhibition assays; Table S1: Amino acid sequences of the covNHR proteins; Table S2: Thermodynamic and kinetic parameters of thermal unfolding; Table S3: Apparent hydrodynamic radii of the covNHR proteins.

Author Contributions: Conceptualization, F.C.-L. and B.M.; methodology, M.C.-M., B.M., C.M. and F.C.-L.; software, M.C.-M.; investigation, M.C.-M., S.C., J.L., B.M. and C.M.; data curation, F.C.-L., C.M. and M.C.-M.; writing—original draft preparation, M.C.-M. and F.C.-L.; writing—review and editing, F.C.-L.; supervision, F.C.-L.; project administration and funding acquisition, F.C.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants BIO2016-76640-R and PID2019.107515RB.C21 from the Spain's State Research Agency, SRA/10.13039/501100011033, co-funded by ERDF/ESF, "A way to make Europe"/"Investing in your future".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The experimental data presented in this study are openly available at the institutional repository of the University of Granada (Digibug) at http://hdl.handle.net/10481/67394, accessed on 11 April 2021.

Acknowledgments: M.C.-M. acknowledges a grant from Youth Employment Operative Program of the Andalusia Government and the European Social Fund (ESF). S.C. acknowledges an exchange studentship from the ERASMUS+ program of the European Union.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UNAIDS. HIV Data and Estimates. Available online: http://www.unaids.org/en/resources/documents/2020/unaids-data (accessed on 7 March 2021).
2. Stephenson, K.E.; Wagh, K.; Korber, B.; Barouch, D.H. Vaccines and Broadly Neutralizing Antibodies for HIV-1 Prevention. *Annu. Rev. Immunol.* **2020**, *38*, 673–703. [CrossRef] [PubMed]
3. Iyidogan, P.; Anderson, K.S. Current Perspectives on HIV-1 Antiretroviral Drug Resistance. *Viruses* **2014**, *6*, 4095–4139. [CrossRef]
4. Silva, B.F.; Peixoto, G.M.L.; da Luz, S.R.; de Moraes, S.M.F.; Peres, S.B. Adverse effects of chronic treatment with the Main subclasses of highly active antiretroviral therapy: A systematic review. *HIV Med.* **2019**, *20*, 429–438. [CrossRef]
5. Chan, D.C.; Kim, P.S. HIV entry and its inhibition. *Cell* **1998**, *93*, 681–684. [CrossRef]
6. Ward, A.B.; Wilson, I.A. The HIV-1 envelope glycoprotein structure: Nailing down a moving target. *Immunol. Rev.* **2017**, *275*, 21–32. [CrossRef]
7. Melikyan, G.B. Membrane Fusion Mediated by Human Immunodeficiency Virus Envelope Glycoprotein. In *Membrane Fusion*; Current Topics in Membranes; Chernomordik, L.V., Kozlov, M.M., Eds.; Elsevier: Amsterdam, The Netherlands, 2011; Volume 68, pp. 81–106.
8. Chan, D.C.; Fass, D.; Berger, J.M.; Kim, P.S. Core structure of gp41 from the HIV envelope glycoprotein. *Cell* **1997**, *89*, 263–273. [CrossRef]
9. Weissenhorn, W.; Dessen, A.; Harrison, S.C.; Skehel, J.J.; Wiley, D.C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **1997**, *387*, 426–430. [CrossRef] [PubMed]
10. Eckert, D.M.; Kim, P.S. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11187–11192. [CrossRef] [PubMed]

11. Yi, H.A.; Fochtman, B.C.; Rizzo, R.C.; Jacobs, A. Inhibition of HIV Entry by Targeting the Envelope Transmembrane Subunit gp41. *Curr. HIV Res.* **2016**, *14*, 283–294. [CrossRef]
12. He, Y.; Cheng, J.; Lu, H.; Li, J.; Hu, J.; Qi, Z.; Liu, Z.; Jiang, S.; Dai, Q. Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16332–16337. [CrossRef] [PubMed]
13. Welch, B.D.; VanDemark, A.P.; Heroux, A.; Hill, C.P.; Kay, M.S. Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16828–16833. [CrossRef]
14. Allen, W.J.; Rizzo, R.C. Computer-Aided Approaches for Targeting HIVgp41. *Biology* **2012**, *1*, 311–338. [CrossRef]
15. Lu, L.; Yu, F.; Cai, L.; Debnath, A.K.; Jiang, S. Development of Small-molecule HIV Entry Inhibitors Specifically Targeting gp120 or gp41. *Curr. Top. Med. Chem.* **2016**, *16*, 1074–1090. [CrossRef]
16. Pu, J.; Wang, Q.; Xu, W.; Lu, L.; Jiang, S.B. Development of Protein- and Peptide-Based HIV Entry Inhibitors Targeting gp120 or gp41. *Viruses* **2019**, *11*, 705. [CrossRef]
17. Kilby, J.M.; Hopkins, S.; Venetta, T.M.; DiMassimo, B.; Cloud, G.A.; Lee, J.Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D.; et al. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **1998**, *4*, 1302–1307. [CrossRef]
18. Yu, D.W.; Ding, X.H.; Liu, Z.X.; Wu, X.Y.; Zhu, Y.M.; Wei, H.M.; Chong, H.H.; Cui, S.; He, Y.X. Molecular mechanism of HIV-1 resistance to sifuvirtide, a clinical trial-approved membrane fusion inhibitor. *J. Biol. Chem.* **2018**, *293*, 12703–12718. [CrossRef]
19. Louis, J.M.; Bewley, C.A.; Clore, G.M. Design and properties of N(CCG)-gp41, a chimeric gp41 molecule with nanomolar HIV fusion inhibitory activity. *J. Biol. Chem.* **2001**, *276*, 29485–29489. [CrossRef]
20. Root, M.J.; Kay, M.S.; Kim, P.S. Protein design of an HIV-1 entry inhibitor. *Science* **2001**, *291*, 884–888. [CrossRef] [PubMed]
21. Chen, X.; Lu, L.; Qi, Z.; Lu, H.; Wang, J.; Yu, X.; Chen, Y.; Jiang, S. Novel Recombinant Engineered gp41 N-terminal Heptad Repeat Trimers and Their Potential as Anti-HIV-1 Therapeutics or Microbicides. *J. Biol. Chem.* **2010**, *285*, 25506–25515. [CrossRef]
22. Crespillo, S.; Camara-Artigas, A.; Casares, S.; Morel, B.; Cobos, E.S.; Mateo, P.L.; Mouz, N.; Martin, C.E.; Roger, M.G.; El Habib, R.; et al. Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti-HIV-1 drugs. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18207–18212. [CrossRef]
23. Jurado, S.; Cano-Munoz, M.; Morel, B.; Standoli, S.; Santarossa, E.; Moog, C.; Schmidt, S.; Laumond, G.; Camara-Artigas, A.; Conejero-Lara, F. Structural and Thermodynamic Analysis of HIV-1 Fusion Inhibition Using Small gp41 Mimetic Proteins. *J. Mol. Biol.* **2019**, *431*, 3091–3106. [CrossRef] [PubMed]
24. Jurado, S.; Cano-Munoz, M.; Polo-Megias, D.; Conejero-Lara, F.; Morel, B. Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects. *Arch. Biochem. Biophys.* **2020**, *688*. [CrossRef]
25. Mason, J.M.; Arndt, K.M. Coiled coil domains: Stability, specificity, and biological implications. *Chembiochem* **2004**, *5*, 170–176. [CrossRef]
26. Gao, J.; Bosco, D.A.; Powers, E.T.; Kelly, J.W. Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat. Struct. Mol. Biol.* **2009**, *16*, 684–690. [CrossRef] [PubMed]
27. Ji, H.; Bracken, C.; Lu, M. Buried polar interactions and conformational stability in the simian immunodeficiency virus (SIV) gp41 core. *Biochemistry* **2000**, *39*, 676–685. [CrossRef]
28. Woolfson, D.N. Coiled-Coil Design: Updated and Upgraded. In *Fibrous Proteins: Structures and Mechanisms*; Subcellular Biochemistry; Parry, D.A.D., Squire, J.M., Eds.; Springer International Publishing Ag: Cham, Switzerland, 2017; Volume 82, pp. 35–61.
29. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]
30. Lumry, R.; Eyring, H. Conformation changes of proteins. *J. Phys. Chem.* **1954**, *58*, 110–120. [CrossRef]
31. Sarzotti-Kelsoe, M.; Bailer, R.T.; Turk, E.; Lin, C.L.; Bilska, M.; Greene, K.M.; Gao, H.; Todd, C.A.; Ozaki, D.A.; Seaman, M.S.; et al. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods* **2014**, *409*, 131–146. [CrossRef] [PubMed]
32. Gonzalez, L.; Woolfson, D.N.; Alber, T. Buried polar residues and structural specificity in the GCN4 leucine zipper. *Nat. Struct. Biol.* **1996**, *3*, 1011–1018. [CrossRef]
33. Akey, D.L.; Malashkevich, V.N.; Kim, P.S. Buried polar residues in coiled-coil interfaces. *Biochemistry* **2001**, *40*, 6352–6360. [CrossRef]
34. Tatko, C.D.; Nanda, V.; Lear, J.D.; DeGrado, W.F. Polar networks control oligomeric assembly in membranes. *J. Am. Chem. Soc.* **2006**, *128*, 4170–4171. [CrossRef] [PubMed]
35. Boyken, S.E.; Chen, Z.B.; Groves, B.; Langan, R.A.; Oberdorfer, G.; Ford, A.; Gilmore, J.M.; Xu, C.F.; DiMaio, F.; Pereira, J.H.; et al. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science* **2016**, *352*, 680–687. [CrossRef] [PubMed]
36. Jurado, S.; Moog, C.; Cano-Munoz, M.; Schmidt, S.; Laumond, G.; Ruocco, V.; Standoli, S.; Polo-Megias, D.; Conejero-Lara, F.; Morel, B. Probing Vulnerability of the gp41 C-Terminal Heptad Repeat as Target for Miniprotein HIV Inhibitors. *J. Mol. Biol.* **2020**, *432*, 5577–5592. [CrossRef]
37. Steger, H.K.; Root, M.J. Kinetic dependence to HIV-1 entry inhibition. *J. Biol. Chem.* **2006**, *281*, 25813–25821. [CrossRef]
38. Kahle, K.M.; Steger, H.K.; Root, M.J. Asymmetric Deactivation of HIV-1 gp41 following Fusion Inhibitor Binding. *PLoS Pathog.* **2009**, *5*, e1000674. [CrossRef]

39. Poveda, E.; Briz, V.; Soriano, V. Enfuvirtide, the first fusion inhibitor to treat HIV infection. *Aids Rev.* **2005**, *7*, 139–147.
40. Martin-Carbonero, L. Discontinuation of the clinical development of fusion inhibitor T-1249. *Aids Rev.* **2004**, *6*, 61. [PubMed]

Supplementary Material to Publication II

Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity.

Supplementary Material to:

Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity

Mario Cano-Muñoz [1,*], Samuele Cesaro [1,†], Bertrand Morel [1,‡], Julie Lucas [2], Christiane Moog [2] and Francisco Conejero-Lara [1,*]

[1] Departamento de Química Física, Instituto de Biotecnología, Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain; mariocano@ugr.es (M.C-M.), bmorel@ugr.es (B.M.); conejero@ugr.es (F.C-L.)

[2] INSERM U1109, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France; julie.lucas@etu.unistra.fr (J.L.); c.moog@unistra.fr (C.M.)

* Correspondence: mariocano@ugr.es (M.C-M.), conejero@ugr.es (F.C-L.); Tel.: +34 958242371 (F.C-L.)

† Present address: Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Strada Le Grazie, 8, 37134, Verona, Italy; samuele.cesaro@univr.it (S.C.)

‡ Present address: Angany Innovation, 1 voie de l'innovation, Pharmaparc II, 27100 Val de Reuil, France; bmorel76@gmail.com

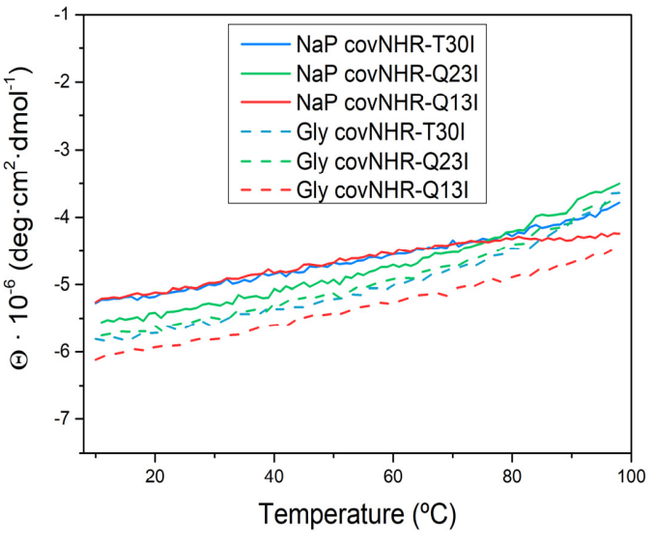


Figure S1: Thermal scans monitored by CD at 222 nm of the covNHR protein variants in 50 mM sodium phosphate buffer pH 7.4 (NaP) and 50 mM glycine/HCl buffer at pH 2.5 (Gly).

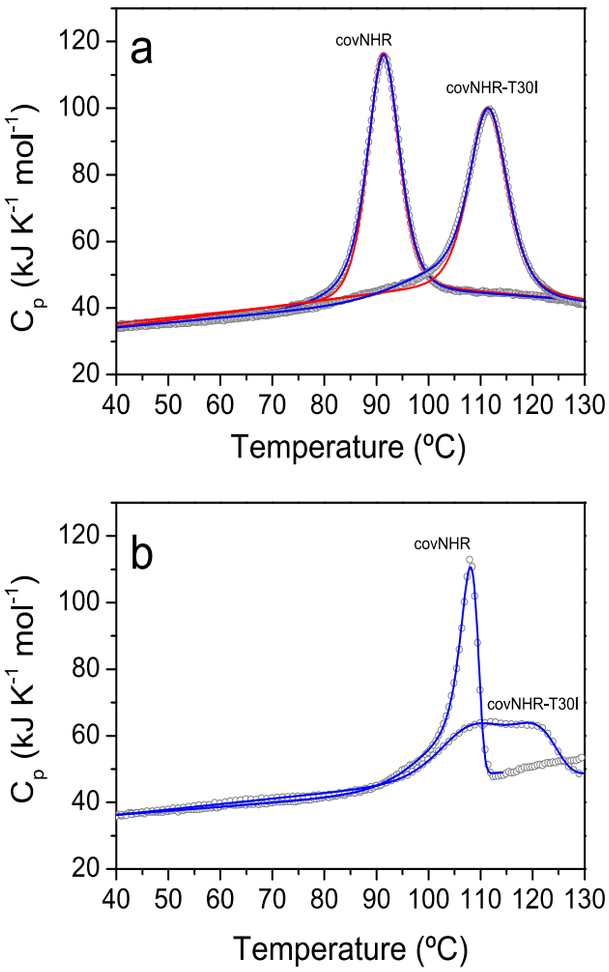


Figure S2: Fitting of DSC thermograms. DSC thermograms of covNHR and covNHR-T30I measured at pH 2.5 in 50 mM Glycine/HCl buffer (a) and pH 7.4 in 50 mM phosphate buffer (b). The experimental data are represented by symbols. The solid lines in (a) represent the best fits using a two-state unfolding model N ⇄ U (red) and a three-state sequential model N ⇄ I ⇄ U (blue). The blue solid lines in (b) represent the best fits using a Lumry-Eyring irreversible denaturation model N ⇄ I → F.

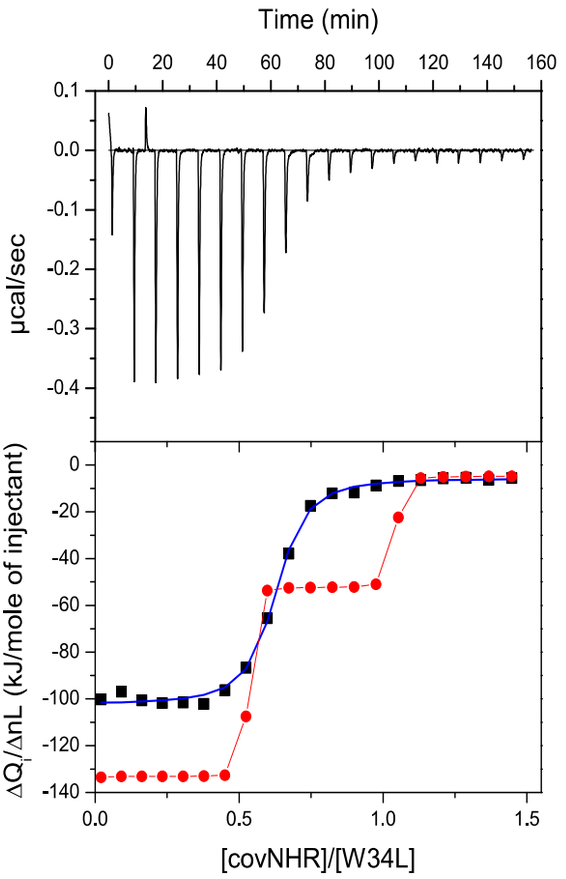


Figure S3: ITC displacement experiment in which a covNHR-Q13I:W34L mixture at a 1:2 molar ratio in the cell was titrated with the reference protein covNHR from the syringe. covNHR-Q13I was at 5 µM and W34L at 10 µM in the cell. (Upper panel) Experimental thermogram at 25°C; (lower panel) ITC isotherm showing the normalized heats per mole of injected protein. The black symbols represent the experimental heats and blue line represents the fit using a n-independent binding sites model. The predicted isotherm assuming rapid binding equilibria and using the parameters of Table 1 is represented in red.

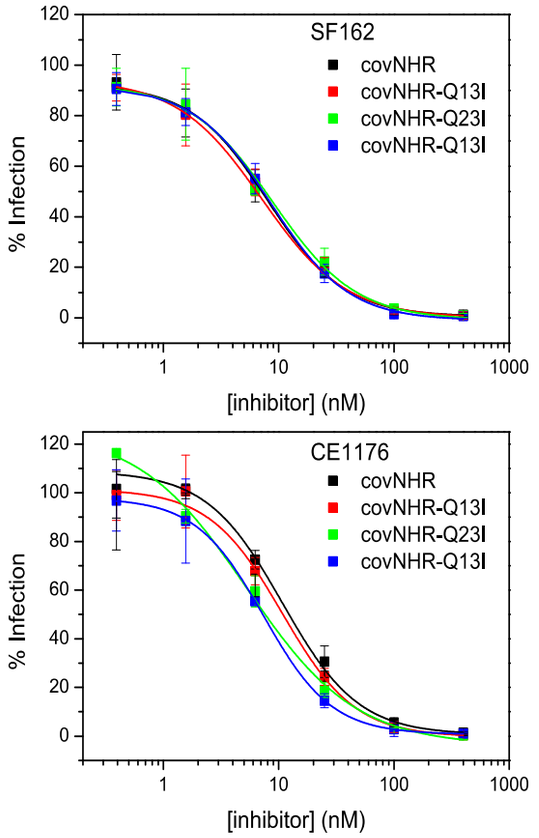


Figure S4: *In vitro* inhibition assays of infection of TZM-bl cells with SF-162 (upper panel) and CE1176 (lower panel) pseudoviruses by different concentrations of covNHR proteins. The data represent the mean values of three independent measurements, and the error bars correspond to the SDs. The solid lines show the fittings using a sigmoidal Hill function.

Table S1: Amino acid sequences of the covNHR proteins and gp41 CHR peptides studied in this work. Residues in bold correspond to the Polar-to-Isoleucine mutations. Loops connecting the helices are identical for all proteins. All constructs were expressed in E. coli with a N-terminal methionine and a C-terminal His-tag sequence GGGGSHHHHHH.

WT gp41	**NHR sequence 30-81**
	ARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAVERYLKDQQL
	Helix 1
covNHR	ARQELSGIVQK**Q**NNLLRQIEA**Q**QHLLQL**T**VSKIKQLQARILAVERYLKDQQL
covNHR-Q13	ARQELSGIVQK**I**NNLLRQIEAQQHLLQLTVSKIKQLQARILAVERYLKDQQL
covNHR-Q23	ARQELSGIVQK**Q**NNLLRQIEA**I**QHLLQL**T**VSKIKQLQARILAVERYLKDQQL
covNHR-T30	ARQELSGIVQK**Q**NNLLRQIEA**Q**QHLLQL**I**VSKIKQLQARILAVERYLKDQQL
	Loop1
	GKGNQ
	Helix 2
covNHR	PQQDKLYREVALIRAQLQKIESE**T**LQLLHQ**Q**AEIERELNN**Q**EQEIGSLKQR
covNHR-Q13	PQQDKLYREVALIRAQLQKIESETLQLLHQQAEIERELNN**I**EQEIGSLKQR
covNHR-Q23	PQQDKLYREVALIRAQLQKIESETLQLLHQ**I**AEIERELNNQEQEIGSLKQR
covNHR-T30	PQQDKLYREVALIRAQLQKIESE**I**LQLLHQQAEIERELNNQEQEIGSLKQR
	Loop2
	GLIDG
	Helix 3
covNHR	PLLSGIDQQ**Q**NNLKRAIEA**Q**KHLLQL**T**VWGIKQLQARILAVERYLKDQQL
covNHR-Q13	PLLSGIDQQ**I**NNLKRAIEAQKHLLQLTVWGIKQLQARILAVERYLKDQQL
covNHR-Q23	PLLSGIDQQQNNLKRAIEA**I**KHLLQLTVWGIKQLQARILAVERYLKDQQL
covNHR-T30	PLLSGIDQQQNNLKRAIEAQKHLLQL**I**VWGIKQLQARILAVERYLKDQQL
	CHR peptides
W34L	WMEWDREINNYTSLIHSLIEESQNQQEKNEQELL
W34N	WMEWDREINNYTSLIHSLIEES
Y24L	YTSLIHSLIEESQNQQEKNEQELL

Table S2: Thermodynamic and kinetic parameters of thermal unfolding of covNHR and covNHR-T30I derived from the fittings shown in Figure S2.

pH 2.5; model N ⇄ I ⇄ U					
Protein	**ΔH_{I-N} (kJ·mol^{-1})**	**$T_{m,1}$ (°C)**	**ΔH_{U-I} (kJ·mol^{-1})**	**$T_{m,2}$ (°C)**	**$\Delta C_{p,U-N}$ (kJ·K^{-1}·mol^{-1})**
covNHR	188 ± 1	89.1 ± 0.1	467 ± 1	90.41 ± 0.02	6.41 ± 0.06
covNHR-T30I	185.6 ± 0.6	102.5 ± 0.1	475.8 ± 0.7	111.4 ± 0.01	2.27 ± 0.06

pH 7.4; model N ⇄ I → F						
Protein	**ΔH_{I-N} (kJ·mol^{-1})**	**$T_{m,1}$ (°C)**	**ΔH_{F-N} (kJ·mol^{-1})**	**E_a [(1)] (kJ·mol^{-1})**	**T^* [(2)] (°C)**	**$\Delta C_{p,F-N}$ (kJ·K^{-1}·mol^{-1})**
covNHR	237 ± 2	102.8 ± 0.2	423.0 ± 1.4	685 ± 3	107.9 ± 0.01	3.85 ± 0.13
covNHR-T30I	269.8 ± 0.7	109.25 ± 0.05	408 ± 2	270 ± 3	125.6 ± 0.09	-2.34 ± 0.13

[(1)] Arrhenius activation energy for the irreversible denaturation process.
[(2)] Temperature at which the rate of denaturation, k_F, is equal to 1 min^{-1}.

Table S3: Apparent hydrodynamic radii of the covNHR proteins measured by dynamic light scattering at a protein concentration of 20 μM.

Protein	**R_h (nm)**
covNHR	2.7
covNHR-Q13I	3.0
covNHR-Q23I	2.8
covNHR-T30I	3.0
covNHR-Q23I + W34L (1:2)	2.7

Publication III

Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion.

International Journal of
Molecular Sciences

MDPI

Article

Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion

Mario Cano-Muñoz [1,2,*], Julie Lucas [2], Li-Yun Lin [2], Samuele Cesaro [1,†], Christiane Moog [2] and Francisco Conejero-Lara [1,*]

[1] Departamento de Química Física, Instituto de Biotecnología y Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain; samuele.cesaro91@gmail.com
[2] INSERM U1109, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, 67084 Strasbourg, France; julie.lucas@etu.unistra.fr (J.L.); li-yun.lin@etu.unistra.fr (L.-Y.L.); c.moog@unistra.fr (C.M.)
* Correspondence: mariocano@ugr.es (M.C.-M.); conejero@ugr.es (F.C.-L.); Tel.: +34-958242371 (F.C.-L.)
† Present address: Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Strada Le Grazie 8, 37134 Verona, Italy.

Abstract: Inhibition of the HIV-1 fusion process constitutes a promising strategy to neutralize the virus at an early stage before it enters the cell. In this process, the envelope glycoprotein (Env) plays a central role by promoting membrane fusion. We previously identified a vulnerability at the flexible C-terminal end of the gp41 C-terminal heptad repeat (CHR) region to inhibition by a single-chain miniprotein (named covNHR-N) that mimics the first half of the gp41 N-terminal heptad repeat (NHR). The miniprotein exhibited low stability, moderate binding to its complementary CHR region, both as an isolated peptide and in native trimeric Envs, and low inhibitory activity against a panel of pseudoviruses. The addition of a disulfide bond stabilizing the miniprotein increased its inhibitory activity, without altering the binding affinity. Here, to further study the effect of conformational stability on binding and inhibitory potency, we additionally stabilized these miniproteins by engineering a second disulfide bond stapling their N-terminal end, The new disulfide-bond strongly stabilizes the protein, increases binding affinity for the CHR target and strongly improves inhibitory activity against several HIV-1 strains. Moreover, high inhibitory activity could be achieved without targeting the preserved hydrophobic pocket motif of gp41. These results may have implications in the discovery of new strategies to inhibit HIV targeting the gp41 CHR region.

Keywords: fusion inhibitor; calorimetry; coiled-coil; envelope glycoprotein; N-terminal domain; antiviral therapy; gp41

Citation: Cano-Muñoz, M.; Lucas, J.; Lin, L.-Y.; Cesaro, S.; Moog, C.; Conejero-Lara, F. Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion. *Int. J. Mol. Sci.* **2022**, *23*, 2794. https://doi.org/10.3390/ijms23052794

Academic Editor: Vladimir N. Uversky

Received: 3 February 2022
Accepted: 28 February 2022
Published: 3 March 2022

1. Introduction

The HIV/AIDS pandemic is still very active and continues to be one of the world's largest pandemics to date with more than 40 million people currently living with HIV still representing a worldwide health issue [1,2]. What is more striking is the upsurge in HIV infections over different populations around the world, such as the outbreak in China's students, where the number of newly diagnosed college students has seen an annual growth rate ranging from 30 to 50% over the past several years [3]. All this together with the fact that HIV newly infects 1.8 million people each year, makes the development of an HIV vaccine a global health priority [4]. However, almost 40 years after the discovery of HIV as the causative agent of AIDS we still do not have a licensed vaccine. Progress has been hindered by the extensive genetic variability of HIV and our limited understanding of the immune responses required to protect against HIV acquisition [5].

This emphasizes the importance of therapeutics to treat the infection. Despite modern Highly Active Antiretroviral Therapy (HAART) having helped to reduce the number of deaths, the absence of an effective vaccine combined with the growing emergence of multi-resistant HIV variants to several of these drugs urges the development of some new anti-HIV compounds directed against the different stages of the virus life cycle, and in particular against the entry of HIV into the cell [1,6].

In order to enter the target human cell, the virus must fuse its membrane with that of the cell. This fusion process is mediated by the Envelope glycoprotein (Env), a non-covalently associated trimer of heterodimers composed of two glycoprotein subunits, gp120 and gp41 [7]. CD4 receptor and co-receptor (CCR5 or CXCR4) binding to gp120 triggers a series of conformational changes that ultimately cause the adoption of a more energetically favorable conformation called the 6 helix bundle (6HB) formed by the N-terminal heptad repeat (NHR) and C-terminal heptad repeat (CHR) regions of gp41. In this 6HB structure, three CHR regions associate externally over an inner helical coiled-coil NHR trimer in an antiparallel fashion. This energetically favorable interaction between NHR and CHR brings viral and host-cell membranes into close proximity promoting fusion and eventually causing infection. Consequently, compounds that interact with either CHR or NHR interfere with this key process and thereby constitute HIV fusion inhibitors [8–10]. For this reason, gp41 has become a very attractive target for the development of potential HIV-1 inhibitors.

Different kinds of fusion inhibitors have been described and classified into two major categories regarding whether they interact with the NHR or CHR regions in gp41. Class-1 inhibitors target the exposed hydrophobic grooves of the NHR helical trimer and comprise a variety of molecules including CHR peptide mimetics, artificial D-peptides, natural products and small-molecule compounds and antibodies [11–15]. On the other hand, class-2 inhibitors target the CHR region and generally encompass NHR peptide mimetics that have traditionally been regarded as low activity anti-HIV compounds. This limited potency may be due to the low solubility and the tendency of NHR peptides to aggregate in solution. These problems can be alleviated by engineered stabilized protein constructs that mimic exposed trimeric NHR grooves. However, they have certain advantages, such as their activity against strains resistant to CHR inhibitors [16,17]. Despite these promising therapeutic approaches, the only FDA-approved fusion inhibitor of AIDS/HIV is T20 (enfuvirtide), a CHR-derived peptide whose clinical use has been limited by its short half-life [18] (proteolysis-sensitive and rapid renal filtration) requiring, therefore, high dosage injections at least twice a day. Moreover, the continuous and expensive treatment generates the appearance of T20-resistant viruses. Nevertheless, compounds that are able to interfere with the formation of the gp41 6HB continue to be very attractive targets for drug design strategies [19,20].

Recently, we have developed several protein molecules called covNHR which consist of a single polypeptide chain with three helical regions that fold as an antiparallel trimeric bundle with a structure highly similar to the NHR gp41 region [21,22]. These proteins can be produced recombinantly by expression in *E. coli* with high yields, without any post-translational modification, are easy to purify, very stable and highly soluble [23]. The NHR binding surface has been described as composed of four different hotspots, namely an N-terminal polar pocket (NTP), a shallow middle pocket (MP), a deep and prominent hydrophobic pocket (HP), which has been widely used as a drug discovery target [14], and finally a C-terminal pocket (CTP) adjacent to the HP [23] (Figure 1C). In a recent study, we designed, produced and characterized two single-chain covNHR miniproteins each encompassing only two consecutive pockets out of the four pockets of the NHR groove [24]. Each miniprotein mimics the N- and the C-terminal half of NHR, respectively, and they were called covNHR-N (harboring only NTP and MP) and covNHR-C (exposing only the HP and the CTP). These miniproteins folded autonomously and represent subdomains of NHR, with very different intrinsic stability. Although both covNHR miniproteins could bind their respective complementary CHR peptides with similar affinity, the covNHR-C protein could not bind its target in soluble prefusion Env spikes and did not show any HIV-1 inhibitory activity in vitro. This is probably due to the HP and CTP binding motifs

being engaged in a tryptophan clasp involving the side chains of Trp623 (CTP motif), Trp628 and Trp631 (HP motif) that locks Env in its pre-fusion conformation [25,26]. On the other hand, covNHR-N could bind its target in soluble prefusion Env spikes and showed moderate HIV-1 inhibitory activity in vitro. However, it proved to be quite unstable and required to be stabilized by engineering a disulfide bond connecting one of its two loops to its C-terminal end. The stabilized versions of covNHR-N with one disulfide bond (called covNHR-N-SS) showed a similar affinity towards its complementary CHR peptide and similar capability to bind to soluble prefusion Env spikes. Strikingly, these stabilized variants showed improved inhibitory potency against different HIV strains.

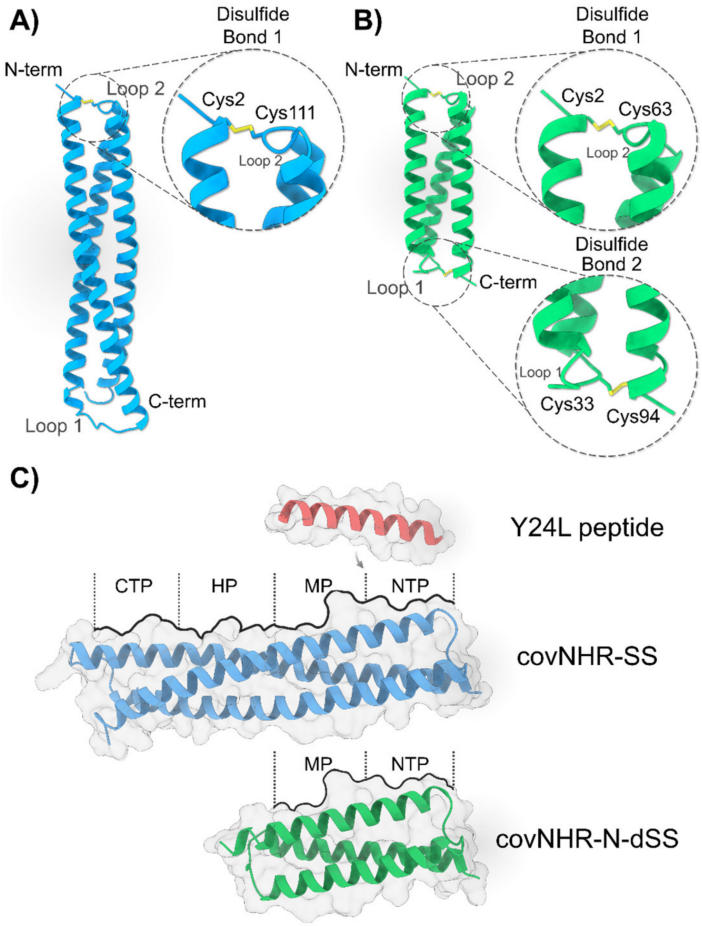


Figure 1. Design of covNHR miniproteins. (**A**,**B**) Ribbon models of covNHR-SS (**A**, blue) and covNHR-N-dSS (**B**, green) showing the location of the residues chosen for mutations to form disulfide bonds. Cysteines are shown in sticks and colored in yellow. (**C**) Model of the structure of the proteins and peptides involved in this study depicted in molecular surface and colored green (covNHR-N-dSS), blue (covNHR-SS) and red (Y24L peptide, gp41 residues 638–661). CovNHR is a highly accurate mimic of the full trimeric gp41 NHR coiled-coil, as such, its binding surface is also composed of four different hotspots: CTP (C-terminal Pocket), HP (Hydrophobic Pocket), MP (Middle Pocket) and NTP (N-terminal Pocket), see details in the text.

Here, we have furtherly increased the stability of covNHR-N by engineering an additional internal disulfide bond stapling the other end of the molecule to test the hypothesis that an increase in stability can lead to a substantial improvement of the affinity to its complementary CHR peptide accompanied by a subsequent improvement in anti-HIV inhibitory potency. This stabilizing strategy has also been implemented in the complete covNHR parent molecule. Therefore, both protein molecules, called covNHR-N-dSS and covNHR-SS, respectively, (Figure 1) were designed, biophysically characterized and tested

for direct Env binding upon several variants. Their capacity to inhibit different HIV-1 pseudoviruses and primary isolates, including T20-resistant strains, was also assessed. The results provide valuable knowledge to the development of protein-based antivirals and reveal new ways to exploit the vulnerability of the gp41 CHR region.

2. Results

2.1. Design and Stabilization of CovNHR Miniproteins

The initial designs of the parent molecules in this study, covNHR and covNHR-N, contain, respectively, the four and two (NTP and MP) of the pockets described in gp41 (Figure 1 and Figure S1). In order to design the new proteins covNHR-SS and covNHR-N-dSS, we used the crystallographic structure of covNHR in complex with C34 peptide (PDB ID: 6R2G, previously determined by our group [22]) as a template. In our previous study, we had achieved a considerable stabilization of over +20 °C in covNHR-N ($T_m \approx 41$ °C) by engineering a disulfide bridge that connected the first loop of the miniprotein with its C-terminal end (G33C/R94C mutations) resulting in a new variant called covNHR-N-SS [24]. This "staple" strategy was recreated in this study in order to achieve even further stabilization of the molecules. To accomplish that, Disulfide by Design [27], a web-based tool for disulfide engineering in proteins, was used to predict pairs of residues that will likely form a disulfide bond if mutated to cysteines. We found one possibility for disulfide bond creation connecting different structural elements of the protein by X-Cys mutations. For covNHR-N-dSS, residues Ala2 at the N-terminus and Leu64 in loop 2 fulfilled the strict geometric constraints that disulfide bonds usually require and were also mutated to Cys. Accordingly, covNHR-N-dSS contains mutations A2C/L63C and G33C/R94C. The equivalent disulfide bond connecting the N-terminus and loop 2 was also engineered for covNHR-SS, which in this case contains mutations A2C/L111C (Figure 1).

In order to validate and assess in silico the stability and dynamic behavior of the newly designed miniproteins we conducted all-atom explicit-solvent molecular dynamics (MD) simulations of the free molecules. Root mean square fluctuation (RMSF) values, which are a measurement of the average atomic mobility of the residues in the protein, showed an increase in the stability (i.e., lower values) in the sites where X-Cys mutations were placed (Figure 2A,B). CovNHR-SS showed decreased RMSF values in the vicinity of the mutated sites compared to covNHR (Figure 2A), which helped to decrease the overall residue mobility of the N-terminal end and the second loop, two hotspots for protein instability. A similar scenario occurred with covNHR-N-dSS if we compare its RMSF values with those of the covNHR-N parent molecule (Figure 2B), in this case, the four hotspots in the protein chain, the N- and C-term as well as the first and second loops were stabilized by the addition of the two disulfide bonds. Moreover, the mean RMSF values for each new protein were lower than their respective parent molecule (Figure 2A,B).

Figure 2C,D shows the time evolution of backbone root mean square deviations (RMSD) of the miniproteins. All the molecules reached equilibrium within the first 5 ns of MD simulation. However, the disulfide-stabilized miniproteins reached a more stable plateau and maintained it throughout the entire simulation time while their parent molecules showed a less stable profile. This is also supported by the fact that the mean RMSD values for the newly engineered miniproteins were also lower than those from their parent molecules. These results indicate that the engineered disulfide bonds reduce the overall conformational fluctuations of the new proteins compared to their parent molecules.

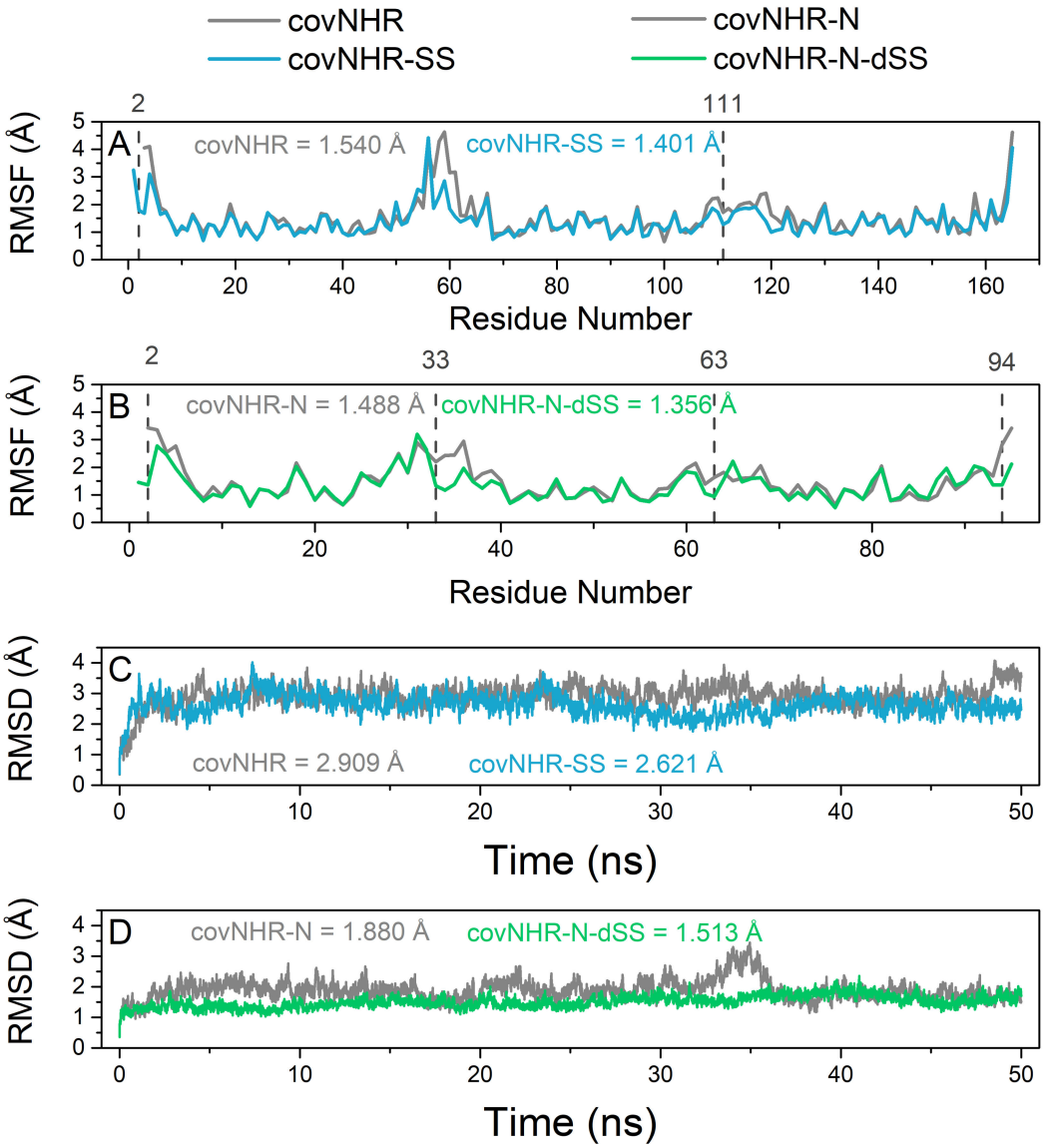


Figure 2. Molecular dynamics simulations analysis of the two miniproteins: covNHR-SS (blue) and covNHR-N-dSS (green) compared with their parent molecule covNHR and covNHR-N, respectively (grey). (**A**) Root mean square fluctuation (RMSF) per residue for covNHR and covNHR-SS. (**B**) RMSF per residue for covNHR-N and covNHR-N-dSS. (**C**) Evolution of mean backbone root mean square deviation (RMSD) for covNHR and covNHR-SS. (**D**) Evolution of RMSD for covNHR-N and covNHR-N-dSS. The locations of residues forming the disulfide bonds are highlighted with grey dashed lines and with their residue number shown above in A and B. Mean values for the parameters throughout the entire simulation time are also displayed in each panel with the same color code used above.

2.2. Biophysical Characterization of CovNHR Variants

Both disulfide-bonded mutants covNHR-N-dSS and covNHR-SS could be expressed and produced recombinantly in *E. coli* with good yields even higher than those of their respective parent molecule. All purification steps were made in the presence of 10 mM β-mercaptoethanol, and a final oxidation step was carried out by extensive dialysis with buffer without a reducing agent. The formation of the disulfide bonds was confirmed in both mutants using Ellman's assay (Thermo Fisher, Waltham, MA, USA). The protein purity was assessed by SDS-PAGE, and the identity of each protein variant was confirmed by mass spectrometry analysis.

The structure and stability of the miniproteins were characterized using various biophysical techniques (Figure 3). The two miniproteins showed far-UV circular dichroism (CD) spectra typical of a mostly α-helical structure (Figure 3A,B); covNHR-SS has a similar α-helical structure content to its parent molecule covNHR [22]. On the other hand, covNHR-N-dSS has higher negative ellipticity values than covNHR-N and covNHR-N-SS, its parent molecules [24], showing a higher α-helical structure percentage (72.4% against 64% and 67.1%, respectively). This indicates that the newly formed disulfide bridge is stabilizing the antiparallel helical bundle. At the same time, mixtures between each protein and the Y24L peptide, containing the NTP and MP binding motifs (Figure 1C), showed an increase in negative ellipticity relative to the theoretical ellipticities of the spectra calculated as the sum of the spectra of the free molecules (Figure 3A,B). This indicates the acquisition of the helical conformation for the Y24L peptide as a consequence of binding onto the NHR groove of both miniproteins. The ellipticity increase was similar in both complexes indicating a comparable acquisition of helical structure. This was not the case for covNHR-N, which showed a higher relative ellipticity increase when binding to Y24L peptide, as a consequence of the protein acquiring a more ordered and helical structure when bound to Y24L [24].

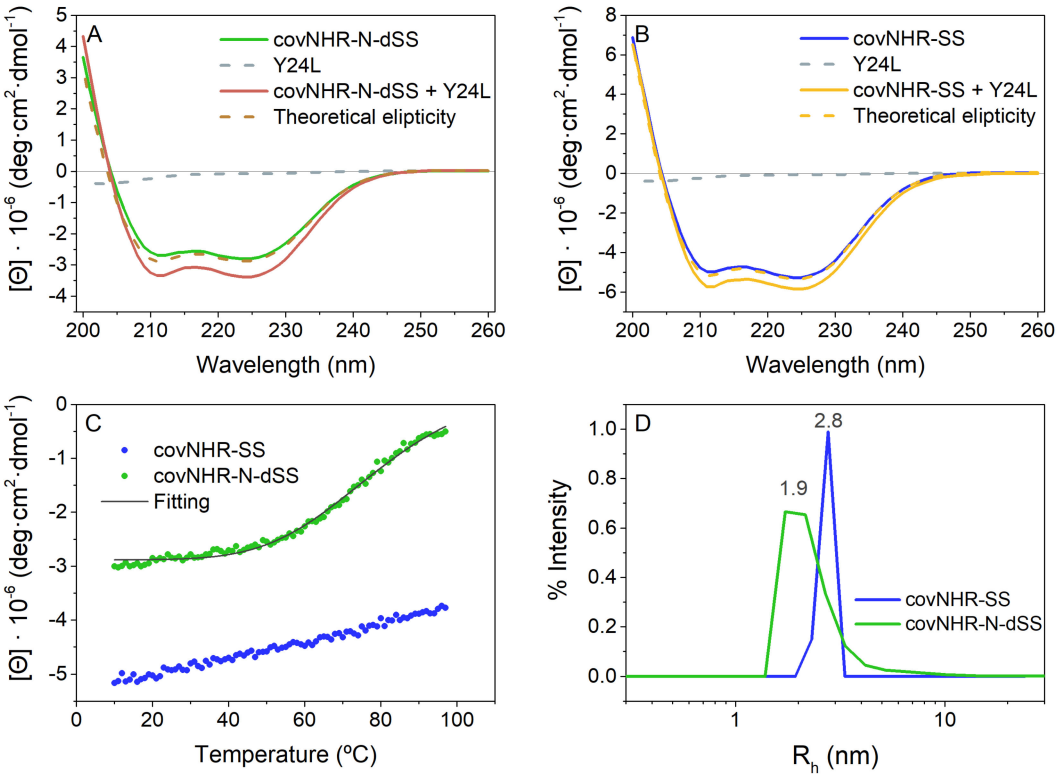


Figure 3. Secondary structure and thermal stability of covNHR-N-dSS (green) and covNHR-SS (blue). (**A**) Far UV CD spectra of free covNHR-N-dSS (green, solid line) and (**B**) covNHR-SS (blue, solid line) and in a (1:2) mixture with their complementary CHR peptide, Y24L (red and yellow lines, respectively). Y24L peptide alone is shown in dashed grey lines in both panels. The red and yellow dashed lines represent the theoretical sum of the spectra of the free molecules. (**C**) Thermal unfolding of covNHR-N-dSS (green symbols) and covNHR-SS (blue symbols) followed by monitoring the CD signal at 222 nm. The grey solid line corresponds to the best fitting carried out using a two-states unfolding model. (**D**) Particle size distributions measured by dynamic light scattering with solutions of covNHR-N-dSS (green) and covNHR-SS (blue). All experiments were carried out at pH 7.4 in 50 mm sodium phosphate.

Thermal denaturation experiments of covNHR-SS and covNHR-N-dSS indicated strong stabilization of the proteins by the disulfide bonds. The melting temperature (T_m)

of covNHR-N-dSS was −79 °C (Figure 3C), a strong increase of +38 °C compared with covNHR-N without any disulfide bridge and an increase of +18 °C compared with the single disulfide bonded covNHR-N-SS [24]. Regarding covNHR-SS, no unfolding transition was observed even after heating up to 98 °C, confirming its extremely high thermostability. The thermal stability of covNHR-SS was estimated by differential scanning calorimetry (DSC) −124 °C (Figure S2), compared to the 105 °C of the covNHR protein under the same conditions [22]. The denaturation peak showed a sharp drop on the high-temperature side of the peak suggesting thermally induced aggregation similar to what happens to covNHR parent molecule. The thermal stability is thus increased by about 18–20 °C with each disulfide bond engineered in both miniproteins.

Both protein variants are highly soluble at physiological pH and the particle sizes of both proteins were assessed by dynamic light scattering (DLS) (Figure 3D). CovNHR-N-dSS showed an apparent hydrodynamic radius (R_h) of 1.9 nm and covNHR-SS exhibits a R_h of 2.8 nm similar to their respective parent molecules and to their theoretical R_h. This demonstrates that both variants are monomeric at physiological pH in 50 mM sodium phosphate buffer.

2.3. Binding of the CHR Peptide to CovNHR Miniproteins

To characterize in detail the thermodynamics of binding of the miniproteins to the complementary CHR peptide, we performed isothermal titration calorimetry (ITC) analysis (Figure 4 and Figure S3 and Table 1) by direct titration of the protein solutions with peptide Y24L, corresponding to gp41 residues 638–661, at the second half of CHR. Previously, we determined the interaction between Y24L and covNHR, which happened to be moderately tight (K_d = 90 ± 7 nM at 25 °C) [23]. CovNHR-SS exhibited a similar K_d of 116 ± 11 nM at the same temperature. The binding enthalpies and heat capacities are also very similar for these two protein variants (−1.6 vs. −1.7 kJ·K^{-1}·mol^{-1}, Table 1) [23]. This indicates that the presence of the disulfide bond at the N-terminus of the complete covNHR protein has a small influence on the binding thermodynamics to the CHR peptide and suggests that the presence of a continuous NHR coiled-coil structure already provides strong conformational stability to the N-domain for a competent binding capability at the NTP and MP pockets.

Table 1. Thermodynamic parameters of binding of gp41 CHR peptide Y24L to covNHR miniproteins measured by ITC.

Protein	Peptide	Temperature (°C)	K_d (nM)	ΔH_b (kJ·mol^{-1})	n	ΔC_{pb} (kJ·K^{-1}·mol^{-1})
covNHR-SS	Y24L	10	80 ± 8	−30.4 ± 1.0	0.75	−1.70 ± 0.21
		15	85 ± 9	−43.5 ± 0.8	0.77	
		20	91 ± 4	−52.0 ± 1.2	0.8	
		25	116 ± 11	−57.7 ± 0.6	0.83	
covNHR-N-dSS	Y24L	10	61 ± 8	−60.7 ± 0.4	0.76	−2.26 ± 0.11
		15	94.3 ± 2.4	−72.1 ± 2.1	0.77	
		20	149 ± 6	−84.0 ± 1.3	0.78	
		25	243 ± 10	−90 ± 4	0.89	

Uncertainties in the parameters correspond to standard errors of the fittings.

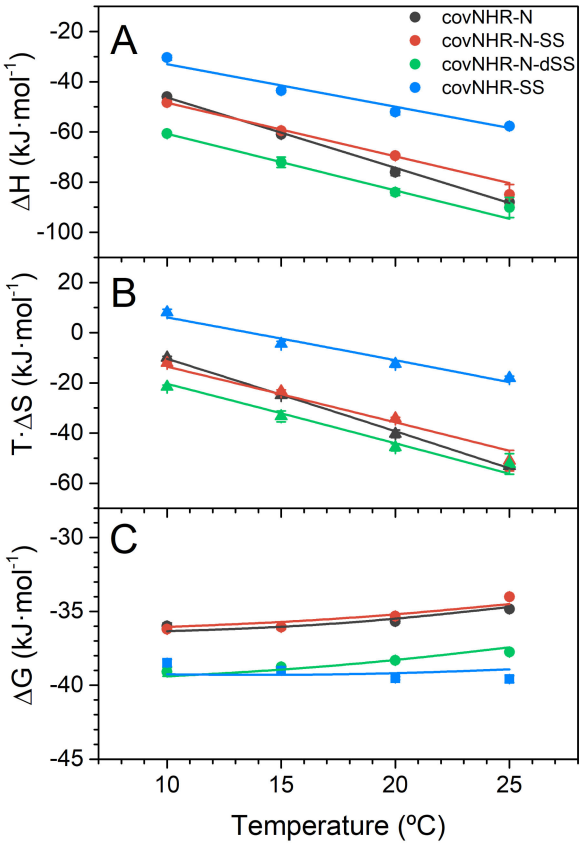


Figure 4. Thermodynamic parameters of Y24L peptide binding to the covNHR proteins. Data have been calculated from the parameters of Table 1 and data from [24], measured by isothermal titration calorimetry (ITC) (**A**) Binding enthalpies; (**B**) binding entropies and (**C**) binding Gibb's energies. The symbols correspond to the values derived from experimental data and the lines represent the temperature dependencies of each parameter according to the binding heat capacity changes.

On the other hand, the affinity of covNHR-N for Y24L peptide was determined to be much lower (K_d = 790 ± 20 nM) and with more negative binding enthalpy and heat capacity as a result of a considerable entropy penalty associated to structural stabilization of covNHR-N upon interaction with the CHR peptide [24]. This binding entropy cost was not reduced in the singly disulfide-bonded covNHR-N-SS [24], which showed very similar binding parameters although slightly reduced binding heat capacity (Figure 4). In marked contrast, the disulfide bond at the N-terminus in covNHR-N-dSS confers a remarkably higher affinity to Y24L (Table 1 and Figure 4), about a 4.5-fold increase. The thermodynamic magnitudes (Table 1, Figure 4A,B) showed significantly more negative binding enthalpy, partially compensated by a higher entropy cost. These magnitudes indicate a tighter interaction in the complex produced by the new disulfide bond. In fact, the binding affinities and Gibb's energies are close to those measured for the complete covNHR proteins at low temperatures, although they decrease rapidly with temperature due to a more negative heat capacity change.

2.4. Binding to Envelope Proteins

We investigated the influence of stabilization by disulfide bonding on the interaction of covNHR-SS and covNHR-N-dSS with their target gp41 CHR region in various Env proteins, both in uncleaved and cleaved trimeric pre-fusion conformations (Table S1). ELISA experiments showed that the two miniproteins bind efficiently to all Envs (Figure 5), comparable to each parent molecule [24], except for a slightly enhanced binding of covNHR-N-dSS to gp140 CN54 Env compared to covNHR-N-SS.

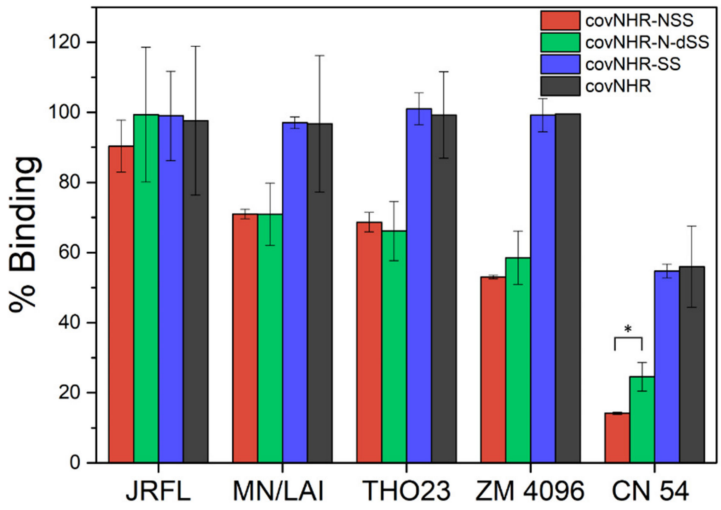


Figure 5. Binding of covNHR miniproteins to different soluble Envs (Table S1) measured by ELISA. Miniproteins binding to Env was detected using anti-6X Histag Ab as primary antibody. Background binding was measured without Env and subtracted from the data; 100% positive control was measured with wells directly coated with a His-tagged Env. Data correspond to mean ± S.D. values of three independent measurements. * $p < 0.05$, difference between covNHR-N-dSS and covNHR-N-SS.

Gp41 mimetics encompassing the four NHR pockets, i.e., covNHR and covNHR-SS, displayed higher levels of binding than the miniproteins harboring only two pockets, covNHR-N-SS and covNHR-N-dSS, except for the structurally folded trimeric JRFL Env, in good agreement with their higher affinity for CHR. These and our previous results [24] demonstrate that the C-terminal part of CHR in a prefusion-like Env conformation is accessible to interaction by the covNHR miniproteins encompassing the MP and NTP. However, stabilization by disulfide bonds has a small influence on the capability of each protein to bind the CHR region in pre-fusion stabilized Envs.

2.5. HIV-1 Inhibition

The inhibitory activities of the miniproteins against HIV-1 infection in vitro were analyzed using the conventional TZM-bl assay (Figure 6). In these studies, we used different HIV-1 pseudovirus strains: two easy to neutralize pseudoviruses (SF162 and MW956.26) [28], two pseudoviruses with pNL4.3 backbone, (pNL4.3 XCS and pNL4.3 DIM) displaying mutations conferring resistance to T20 [22], and one difficult to neutralize primary isolate (CE1176) [28].

The IC_{50} values were compared with those of the parent covNHR proteins (Table 2). The two proteins containing the four NHR binding pockets, covNHR and covNHR-SS, display similar IC_{50} values in the low nanomolar range meaning that the addition of the N-terminal disulfide bond did not significantly modify the inhibitory capacity of the entire NHR groove mimetic. This is consistent with the fact that both proteins show very similar binding to the CHR target, as shown above.

The inhibitory activity of covNHR-N (containing only two pockets) was very poor as a result of its low affinity for its CHR target [24]. However, conformational stabilization by one disulfide bond in covNHR-N-SS and two disulfide bonds in covNHR-N-dSS increased strongly and progressively the inhibitory activity (Figure 6 and Table 2). These observed activity increments do not appear to be a result of an increase in target binding affinity but rather a consequence of the improved conformational stability of the miniproteins. In fact, the first disulfide bond in covNHR-N-SS did not alter the affinity for the Y24L peptide and the three miniproteins show a similar capacity to bind CHR of prefusion trimeric Envs in ELISA experiments.

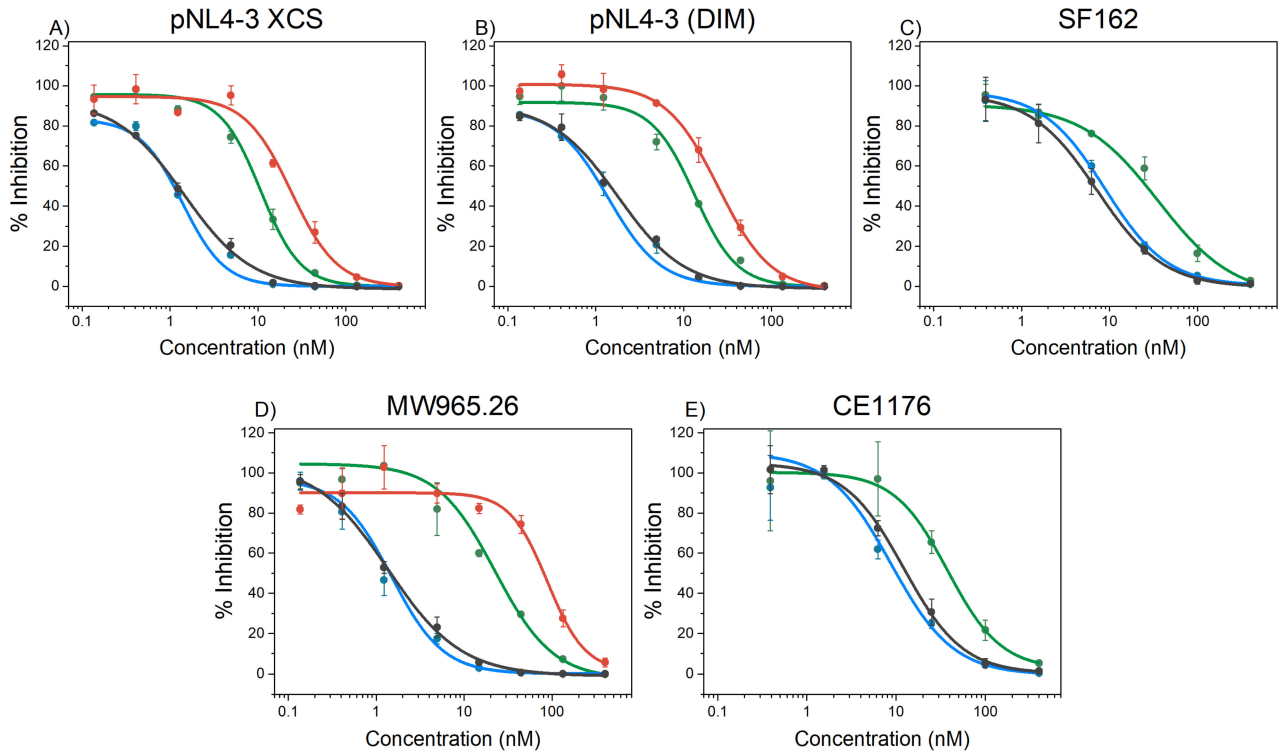


Figure 6. HIV-1 inhibitory activity of the miniproteins implemented in this study on different HIV-1 strains. In vitro inhibition of different HIV-1 strains infection of TZM-bL cells by fusion inhibitors, covNHR (black), covNHR-SS (blue), covNHR-N-SS (red) and covNHR-N-dSS (green), added at different concentrations. The different HIV-1 strains are (**A**) pNL4-3 XCS (pseudovirus designed for resistance to T20); (**B**) pNL4-3 (DIM) (pseudovirus designed for resistance to T20); (**C**) SF162 (pseudovirus strain); (**D**) MW965.26 (pseudovirus strain) and (**E**) CE1176 (primary isolate). Data are the mean ± S.D. of three independent measurements. Continuous lines correspond to non-linear regression curves using a sigmoidal Hill function as implemented in Origin software (Originlab, Northampton, MA, USA).

Table 2. In vitro HIV-1 inhibition by covNHR miniproteins.

Pseudovirus	covNHR-N-SS	covNHR-N-dSS	covNHR	covNHR-SS
pNL4-3 XCS [a]	25.6 ± 2.4	11 ± 4 **	1.3 ± 0.2	1.4 ± 0.1
pNL4-3 (DIM) [a]	24.8 ± 2.0	13 ± 1.3 **	1.6 ± 0.1	1.4 ± 0.2
SF162	n.d.	36 ± 12	8.0 ± 1.3	8.9 ± 0.9
MW965.26	96 ± 12	23 ± 3 ***	1.5 ± 0.3	2.0 ± 0.5
CE1176 [b]	n.d	37.9 ± 1.1	10.8 ± 1.3	8.8 ± 1.6

Inhibitory activity (IC_{50} nM ± S.D. of triplicates) was measured with the standard TZM-bl assay using different pseudoviruses; [a] T20-resistant strains. [b] Primary isolate. ** $p < 0.01$; *** $p < 0.001$, differences between covNHR-N-dSS and covNHR-N-SS.

Strikingly, covNHR-N-dSS exhibits IC_{50} values of 1–3 tens of nanomolar, only 4- to 10-fold higher than covNHR and covNHR-SS, despite the fact that the former miniprotein does not harbor the HP and CTP pockets and has a consequently a much lower affinity for the CHR region.

Compared to T20, the stabilized miniproteins show consistently higher and broader inhibitory activity against the same virus strains, including T20-resistant strains [24]. Additionally, no cytotoxicity, monitored by microscopic examination, was detected even with the highest concentration of the miniproteins used in the assay.

3. Discussion

The NHR region has traditionally been considered as a low stability domain in gp41 with a strong tendency to aggregate, and therefore, NHR-based peptides are poor HIV-1 inhibitors. However, this highly preserved region remains to be an attractive target for fusion inhibition and vaccine design. Different approaches to stabilize an exposed NHR coiled-coil trimer have generally involved inter-helical tethering [29], fusion to foldon domains [9], or partial association with CHR sequences [30]. In these studies, a substantial correlation between coiled-coil stability and inhibitory activity has been repeatedly reported but no clear explanation for this correlation has been provided. We have previously demonstrated that using an innovative design and engineering approach, highly stable mimics of a fully exposed NHR region can be produced in a single-chain form [21–24,31]. Within the NHR coiled coil, its N-terminal region has also been described as an intrinsically unstable sub-domain compared to that of the C-terminal domain or the entire NHR [24,32]. Here we show how the versatility of our single-chain design allows for simple and effective conformational stabilization by disulfide bond engineering. Each disulfide bond is formed spontaneously in the correct configuration and thermally stabilizes the proteins by about 20 °C, consistently with a strong reduction in the conformational entropy of folding [33].

Despite the additive global stabilizing effect of each disulfide bond, only the addition of the second disulfide bond at the N-terminus has a significant influence on the binding affinity to the CHR peptide. The thermodynamics of binding suggest tighter interactions of the doubly disulfide bonded variant with the CHR peptide, possibly due to a reduced conformational dynamics at the peptide-protein complex interface, as a result of a local stabilization effect produced by the disulfide bond as suggested by the MD simulations. Nevertheless, both disulfide bonds contribute to marked increases in HIV-1 inhibitory activity. Disulfide bond stabilization also produced strong increases in anti-HIV-1 activity in other NHR coiled-coil trimer mimetics [10,34]. However, this high affinity, stability and inhibitory activity has, to the best of our knowledge, never been seen before for a NHR gp41 construct lacking the HP, which has been in the last decades the main focus of attention to interfere with the NHR-CHR interaction of gp41 and thereby inhibit fusion and infection [35–37]. For instance, stabilized NHR constructs IZN17, IZN23 and IZN36, all containing the HP but differing in the sequence extension of NHR towards the N-terminus, were reported to have similar inhibitory activities, suggesting that the antiviral activity of this class of chimeric NHR peptides is recapitulated in the HP region [10].

In marked contrast, we show here that the HP is not necessary to achieve a potent and broad inhibitory activity in the tens of nanomolar range for an NHR based construct, such as covNHR-N-dSS. The inhibitory activity is relatively close (only 4- to 10-fold lower) to the activity of the complete NHR mimetic covNHR, as well as other NHR mimetics, with IC_{50} values in the low nM range, even for a difficult to neutralize primary virus isolate (Figure 6). This is surprising because the binding interface of the complete covNHR with its complementary CHR region is much larger and can bind long CHR peptides that also include the HP binding motif, such as C34, with sub-pM affinity [22], whereas the affinity of covNHR-N-dSS for its complementary CHR region, encompassing only the MP and NTP, is only in the high nM range. This suggests that not every CHR binding motif is equally accessible to covNHR for inhibition. It has also been reported that inhibition potency of 5-helix constructs is kinetically restricted by the rate of association of the inhibitor to its target in CHR, which is only transiently exposed during fusion [38,39]. However, all NHR-based inhibitors used in previous studies contain the HP and interact with the HP binding motif, which is engaged in a tryptophan clamp stabilizing the prefusion Env conformation, and therefore, needs Env activation to become accessible [25,26]. On the other hand, accessibility of covNHR-N-dSS to its targeted CHR region may be facilitated by its small size reducing steric impediments and by the high flexibility of the CHR C-terminus [40]. In fact, a recent cryo-EM structure of full-length Env localizes the connection between the CHR end and the membrane-proximal external region (MPER) at a flexible and disordered polar segment composed of residues E^{654}KNEQE^{659} [41]. This segment is actually part of the binding

motif of the NTP and is engaged in a water-mediated network of hydrogen bonds in the gp41 post-fusion 6HB structure [22]. Both the NTP and the NTP-binding motif are as highly preserved as the HP and HP-binding motif (Figure S4), suggesting chief importance of this interaction in gp41 function. It is, therefore, possible that this flexible segment connecting CHR and MPER constitutes the primary target of covNHR-N-dSS in the virion context. This would explain why a stabilization of the miniprotein favoring this conformationally constrained interaction acts so strongly to increase inhibition potency.

Our results suggest two different and probably complementary modes of fusion inhibition targeting gp41 CHR. First, inhibitors containing the HP, which mainly act by targeting the HP binding motif and require its release and exposure by Env activation. Second, inhibitors targeting the C-terminal part of CHR, immediately upstream of the MPER, such as the covNHR-N miniproteins described here, which due to their small size and the higher accessibility and flexibility of their target, are less sterically restricted, and therefore, can achieve potent activities with less stringent requirements in binding affinity.

Our covNHR miniproteins have advantages over peptide-based fusion inhibitors, such as T20. First, they can be produced in recombinant form by *E. coli* expression with high yields. They spontaneously fold with the correct mimetic structure without any posttranslational modification. They are monomeric, highly soluble and stable, and can be lyophilized and reconstituted in standard buffers without structure or activity loss. All these features facilitate the scaling up of production and storage. Second, as potential antivirals, due to their polypeptidic nature, such as T20, they would need to be administered by intravenous injection but, because of their folded structure and high stability, it is highly likely that they will have a higher resistance to proteolytic degradation and longer life in the bloodstream, allowing a considerable reduction of the dosage and/or the frequency of injection, which are among the main drawbacks of T20 treatment.

These results shed light on the design of new inhibitors encompassing the N-terminal subdomain of NHR gp41 traditionally less investigated, proving the potency of the stabilized gp41's NHR mimetics and opening up new ways of inhibiting HIV-1 by engineering new modifications increasing the stability of this region, as well as by improving the already high binding affinity for its target by adding, for instance, new motifs targeting the nearby MPER region.

4. Materials and Methods

4.1. Molecular Dynamics Simulations

All-atom molecular dynamics simulations were performed using YASARA Structure (v.17.12.24) [42] with explicit solvent (TIP3P water, the solvent density was equilibrated to a final value of 0.997 g/mL) in a periodic box with a size 10 Å larger than the protein in every dimension. In order to describe long-range electrostatics, the Particle Mesh Ewald (PME) [43] method was used with a cutoff distance of 8 Å at physiological conditions (0.9% NaCl, pH 7.4), constant temperature (298 K) using a weakly-coupled Berendsen thermostat and constant pressure (1 bar). Ewald summation was used to assign amino acids charge according to their predicted side chain pK_a and was neutralized by adding counterions (NaCl) [44]. The AMBER14 [45] force field was used together with multiple time step integration where intra-molecular forces were calculated every 2 fs and inter-molecular forces every 2.5 fs. The structures were initially energy-minimized using first steepest descent without electrostatics to remove steric clashes and conformational stress and subsequently relaxed by steepest descent minimization and simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, i.e., the energy improved by less than 0.05 kJ mol^{-1} per atom during 200 steps. The minimized system was slowly heated up during an equilibration phase until the target temperature and density was reached. Every system was simulated for a minimum of 50 ns and coordinates were saved every 10 ps, yielding 5000 time points for each trajectory.

4.2. Protein and Peptide Samples

The NHR and CHR gp41 sequences used in this work are described in Figure S1. The reference gp41 sequence was taken from the full gp160 precursor glycoprotein of the HIV-1 BRU isolate (Swiss-Prot entry sp|P03377|ENV_HV1BR). CovNHR miniproteins were computationally designed using YASARA software. The DNA encoding the protein sequences were synthesized and cloned into pET303 expression vectors (Thermo Fisher Scientific, Waltham, USA). To facilitate purification by Ni- Sepharose affinity chromatography, the protein sequences were histidine tagged at the C terminus with the sequence GGGGSHHH-HHH. The covNHR proteins were produced and purified following the protocol previously described [21]. Synthetic CHR peptides, both N-acetylated and C- amidated, were acquired from Genecust (Luxembourg), with a purity >95%. Protein and peptide concentrations were determined by UV absorption measurements at 280 nm using the extinction coefficients calculated according to their respective amino acid sequences with the ExPasy ProtParam server (https://web.expasy.org/protparam/ accessed on 2 February 2022) [46].

4.3. Circular Dichroism

CD spectra were recorded in a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a Peltier thermostatic cell holder. Measurements of the far-UV CD spectra (260–200 nm) were made with a 1-mm path-length quartz cuvette at a protein concentration of ~15 μM. Spectra were recorded at a scan rate of 100 nm/min, 1-nm step resolution, 1-s response, and 1-nm bandwidth. The resulting spectra were usually the average of five scans and the percentage of the α-helical structure was estimated from the far-UV CD spectra as described elsewhere [47]. In thermal melting experiments, the CD signal was monitored as a function of temperature at 222 nm. Each spectrum was corrected by baseline subtraction using the blank spectrum obtained with the buffer and finally, the CD signal was normalized to molar ellipticity ([θ], in $deg \cdot dmol^{-1} \cdot cm^2$). The interaction experiments with CHR peptides were carried out at a 1:2 molar ratio between the proteins and the corresponding peptide.

4.4. Dynamic Light Scattering

The particle sizes of the covNHR proteins were assessed by DLS measurements using a DynaPro MS-X instrument (Wyatt, Santa Barbara, CA, USA). Dynamics software (Wyatt Technology Corporation, Santa Barbara, CA, USA) was used in data collection and processing. Sets of DLS data were measured at 25 °C with an average number of 50 acquisitions and an acquisition time of 10 s.

4.5. Isothermal Titration Calorimetry

ITC measurements were carried out in a Microcal VP-ITC calorimeter (Malvern Instruments, Worcestershire, UK). The protein solutions were typically titrated with 25 injections of 5 μL of the peptide solution at 480 s intervals. Protein concentration in the cell was ~20 μM, while the ligands in the syringe were typically at ~300 μM. The experiments were carried out in 50 mM sodium phosphate buffer, pH 7.4. As a blank, an independent experiment with only buffer in the calorimeter's cell was performed with the same peptide solution to determine the corresponding heats of dilution. The experimental thermograms were baseline corrected and the peaks were integrated to determine the heats produced by each ligand injection. Finally, each heat was normalized per mole of added ligand. The resulting binding isotherms were fitted using a binding model of identical and independent sites, allowing the determination of the binding constant, K_b, the binding enthalpy, ΔH_b, and the binding stoichiometry, n, for each interaction. From these values, the Gibbs energy and entropy of binding could be derived as $\Delta G_b = -RT \cdot \ln K_b$ and $T \cdot \Delta S_b = \Delta H_b - \Delta G_b$. Binding heat capacities were determined from the slope of the dependences of the binding enthalpies measured at different temperatures (ranging from 10 to 25 °C).

4.6. Binding to HIV-1 Envelope Spikes

The capacity of the covNHR proteins to bind soluble HIV-1 envelope proteins (Env) was determined by ELISA. Briefly, 96-well ELISA plates (Maxisorp, Nunc) were coated at 4 °C overnight with various Envs (Table S1) in 0.1 M bicarbonate buffer (pH 9.6). After saturation with 2% BSA, 0.05% Tween in PBS for 1.5 h at 25 °C, 0.01 µM of covNHR molecules, corresponding to the protein concentrations that allowed detecting optical density changes within a linear range, (100 µL diluted in 1% BSA 0.05% Tween solution) were added and incubated for 2 h at room temperature. The plate was then washed five times and covNHR binding was detected with 100 µL anti-6X His-tag antibody conjugated to horseradish peroxidase (HRP) (Abcam, Cambridge, UK) at 1/dilution incubated for 1 h at room temperature. Antibody binding was then revealed with tetramethylbenzidine (TMB) substrate buffer, the reaction was stopped with 1 M H_2SO_4 and optical density was read at 450 nm with a Molecular Device Plate Reader equipped with SoftMax Pro 6 program. Background binding was measured in plates without Env and subtracted from the data. The percentage of binding was calculated using the readings with wells coated with His-tagged Env incubated with PBS buffer instead of covNHR molecules as a control for 100% binding.

4.7. HIV-1 Inhibitory Assays

The inhibition of HIV replication was determined using the conventional TZM-bl assay measured as a function of reductions in Tat-regulated Firefly luciferase (Luc) reporter gene expression [48]. Pseudoviruses expressing different Env were tested for HIV inhibitory potential [28]. The IC_{50}, the concentration (in nM) of inhibitor inducing a 50% decrease in relative luminometer units (RLU), corresponding to a 50% decrease in virus replication was calculated by non-linear regression using a sigmoidal Hill function, as implemented in Origin software (Originlab, Northampton, MA, USA).

4.8. Statistical Analysis

All statistical analyses were performed using the Prism 6 scientific software. Data were expressed as the mean ± SD of 3 experiments per group. An unpaired Student's *t*-test was used to compare the differences between two experimental groups. A *p*-value of 0.05 or less was considered to be statistically significant.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23052794/s1.

Author Contributions: Conceptualization, M.C.-M. and F.C.-L.; methodology, M.C.-M., C.M. and F.C.-L.; software, M.C.-M. and F.C.-L.; formal analysis, M.C.-M. and F.C.-L.; investigation, M.C.-M., J.L., L.-Y.L., S.C., C.M. and F.C.-L.; data curation, M.C.-M. and F.C.-L.; writing—original draft preparation, M.C.-M. and F.C.-L.; writing—review and editing, M.C.-M. and F.C.-L.; supervision, F.C.-L., C.M.; project administration and funding acquisition, F.C.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants BIO2016-76640-R and PID2019.107515RB.C21 from the Spain's State Research Agency, SRA/10.13039/501100011033, co-funded by ERDF/ESF, "A way to make Europe"/"Investing in your future".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: M.C.-M. acknowledges a grant from Youth Employment Operative Program of the Andalusia Government and the European Social Fund (ESF). S.C. acknowledges an exchange studentship from the ERASMUS+ program of the European Union. The results shown are included as part of M.C.-M. doctoral thesis.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McLaren, P.J.; Fellay, J. HIV-1 and human genetic variation. *Nat. Rev. Genet.* **2021**, *22*, 645–657. [CrossRef] [PubMed]
2. Cano-Muñoz, M.; Jurado, S.; Morel, B.; Conejero-Lara, F. Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors. *Int. J. Biol. Macromol.* **2021**, *192*, 90–99. [CrossRef] [PubMed]
3. Li, G.; Jiang, Y.; Zhang, L. HIV upsurge in China's students. *Science* **2019**, *364*, 711. [CrossRef] [PubMed]
4. Steichen, J.M.; Lin, Y.C.; Havenar-Daughton, C.; Pecetta, S.; Ozorowski, G.; Willis, J.R.; Toy, L.; Sok, D.; Liguori, A.; Kratochvil, S.; et al. A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science* **2019**, *366*. [CrossRef] [PubMed]
5. Ng'uni, T.; Chasara, C.; Ndhlovu, Z.M. Major Scientific Hurdles in HIV Vaccine Development: Historical Perspective and Future Directions. *Front. Immunol.* **2020**, *11*, 590780. [CrossRef] [PubMed]
6. Deeks, S.G.; Overbaugh, J.; Phillips, A.; Buchbinder, S. HIV infection. *Nat. Rev. Dis. Prim.* **2015**, *1*, 15035. [CrossRef]
7. Ward, A.B.; Wilson, I.A. The HIV-1 envelope glycoprotein structure: Nailing down a moving target. *Immunol. Rev.* **2017**, *275*, 21–32. [CrossRef]
8. Yi, H.A.; Fochtman, B.C.; Rizzo, R.C.; Jacobs, A. Inhibition of HIV Entry by Targeting the Envelope Transmembrane Subunit gp41. *Curr. HIV Res.* **2016**, *14*, 283–294. [CrossRef]
9. Eckert, D.M.; Malashkevich, V.N.; Hong, L.H.; Carr, P.A.; Kim, P.S. Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **1999**, *99*, 103–115. [CrossRef]
10. Eckert, D.M.; Kim, P.S. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 11187–11192. [CrossRef]
11. Chan, D.C.; Fass, D.; Berger, J.M.; Kim, P.S. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell* **1997**, *89*, 263–273. [CrossRef]
12. Weissenhorn, W.; Dessen, A.; Harrison, S.C.; Skehel, J.J.; Wiley, D.C.; Weissenhorn, W.; Dessen, A.; Harrison, S.C.; Skehel, J.J.; Wiley, D.C. Atomic structure of the ectodomain from HIV-1 gp41. *Nature* **1997**, *387*, 426–430. [CrossRef]
13. Kilby, J.M.; Hopkins, S.; Venetta, T.M.; DiMassimo, B.; Cloud, G.A.; Lee, J.Y.; Alldredge, L.; Hunter, E.; Lambert, D.; Bolognesi, D.; et al. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **1998**, *4*, 1302–1307. [CrossRef]
14. He, Y.; Cheng, J.; Lu, H.; Li, J.; Hu, J.; Qi, Z.; Liu, Z.; Jiang, S.; Dai, Q. Potent HIV fusion inhibitors against Enfuvirtide-resistant HIV-1 strains. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 16332–16337. [CrossRef]
15. Medina-O'donnell, M.; Rivas, F.; Reyes-Zurita, F.J.; Cano-Muñoz, M.; Martinez, A.; Lupiañez, J.A.; Parra, A. Oleanolic Acid Derivatives as Potential Inhibitors of HIV-1 Protease. *J. Nat. Prod.* **2019**, *82*, 2886–2896. [CrossRef]
16. Welch, B.D.; VanDemark, A.P.; Heroux, A.; Hill, C.P.; Kay, M.S. Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 16828–16833. [CrossRef]
17. Frey, G.; Rits-Volloch, S.; Zhang, X.Q.; Schooley, R.T.; Chen, B.; Harrison, S.C. Small molecules that bind the inner core of gp41 and inhibit HIV envelope-mediated fusion. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 13938–13943. [CrossRef]
18. Lu, L.; Yu, F.; Cai, L.; Debnath, A.K.; Jiang, S. Development of Small-molecule HIV Entry Inhibitors Specifically Targeting gp120 or gp41. *Curr. Top. Med. Chem.* **2015**, *16*, 1074–1090. [CrossRef]
19. Miller, M.D.; Geleziunas, R.; Bianchi, E.; Lennard, S.; Hrin, R.; Zhang, H.; Lu, M.; An, Z.; Ingallinella, P.; Finotto, M.; et al. A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 14759–14764. [CrossRef]
20. He, Y. Synthesized Peptide Inhibitors of HIV-1 gp41-dependent Membrane Fusion. *Curr. Pharm. Des.* **2013**, *19*, 1800–1809. [CrossRef]
21. Crespillo, S.; Cámara-Artigas, A.; Casares, S.; Morel, B.; Cobos, E.S.; Mateo, P.L.; Mouz, N.; Martin, C.E.; Roger, M.G.; El Habib, R.; et al. Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti–HIV-1 drugs. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18207–18212. [CrossRef] [PubMed]
22. Jurado, S.; Cano-Muñoz, M.; Morel, B.; Standoli, S.; Santarossa, E.; Moog, C.; Schmidt, S.; Laumond, G.; Cámara-Artigas, A.; Conejero-Lara, F. Structural and Thermodynamic Analysis of HIV-1 Fusion Inhibition Using Small gp41 Mimetic Proteins. *J. Mol. Biol.* **2019**, *431*, 3091–3106. [CrossRef] [PubMed]
23. Jurado, S.; Cano-Muñoz, M.; Polo-Megías, D.; Conejero-Lara, F.; Morel, B. Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects. *Arch. Biochem. Biophys.* **2020**, *688*, 108401. [CrossRef] [PubMed]
24. Jurado, S.; Moog, C.; Cano-Muñoz, M.; Schmidt, S.; Laumond, G.; Ruocco, V.; Standoli, S.; Polo-Megías, D.; Conejero-Lara, F.; Morel, B. Probing Vulnerability of the gp41 C-Terminal Heptad Repeat as Target for Miniprotein HIV Inhibitors. *J. Mol. Biol.* **2020**, *432*, 5577–5592. [CrossRef]
25. Pancera, M.; Zhou, T.; Druz, A.; Georgiev, I.S.; Soto, C.; Gorman, J.; Huang, J.; Acharya, P.; Chuang, G.; Ofek, G.; et al. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **2014**, *514*, 455–461. [CrossRef]
26. Lee, J.H.; Ozorowski, G.; Ward, A.B. Cryo-EM structure of a native, fully glycosylated, cleaved HIV-1 envelope trimer. *Science* **2016**, *351*, 1043–1048. [CrossRef]
27. Craig, D.B.; Dombkowski, A.A. Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins. *BMC Bioinform.* **2013**, *14*, 346. [CrossRef]

28. DeCamp, A.; Hraber, P.; Bailer, R.T.; Seaman, M.S.; Ochsenbauer, C.; Kappes, J.; Gottardo, R.; Edlefsen, P.; Self, S.; Tang, H.; et al. Global Panel of HIV-1 Env Reference Strains for Standardized Assessments of Vaccine-Elicited Neutralizing Antibodies. *J. Virol.* **2014**, *88*, 2489–2507. [CrossRef]
29. Bianchi, E.; Finotto, M.; Ingallinella, P.; Hrin, R.; Carella, A.V.; Hou, X.S.; Schleif, W.A.; Miller, M.D.; Geleziunas, R.; Pessi, A. Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12903–12908. [CrossRef]
30. Root, M.J.; Kay, M.S.; Kim, P.S. Protein Design of an HIV-1 Entry Inhibitor. *Science* **2001**, *291*, 884–888. [CrossRef]
31. Cano-muñoz, M.; Cesaro, S.; Morel, B.; Lucas, J.; Moog, C.; Conejero-lara, F. Extremely thermostabilizing core mutations in coiled-coil mimetic proteins of HIV-1 gp41 produce diverse effects on target binding but do not affect their inhibitory activity. *Biomolecules* **2021**, *11*, 566. [CrossRef]
32. Dwyer, J.J.; Hasan, A.; Wilson, K.L.; White, J.M.; Matthews, T.J.; Delmedico, M.K. The hydrophobic pocket contributes to the structural stability of the N-terminal coiled coil of HIV gp41 but is not required for six-helix bundle formation. *Biochemistry* **2003**, *42*, 4945–4953. [CrossRef]
33. Dombkowski, A.A.; Sultana, K.Z.; Craig, D.B. Protein disulfide engineering. *FEBS Lett.* **2014**, *588*, 206–212. [CrossRef]
34. Tong, P.; Lu, Z.; Chen, X.; Wang, Q.; Yu, F.; Zou, P.; Yu, X.; Li, Y.; Lu, L.; Chen, Y.H.; et al. An engineered HIV-1 gp41 trimeric coiled coil with increased stability and anti-HIV-1 activity: Implication for developing anti-HIV microbicides. *J. Antimicrob. Chemother.* **2013**, *68*, 2533–2544. [CrossRef]
35. Chan, D.C.; Chutkowski, C.T.; Kim, P.S. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Med. Sci.* **1998**, *95*, 15613–15617. [CrossRef]
36. Sofiyev, V.; Kaur, H.; Snyder, B.A.; Hogan, P.A.; Ptak, R.G.; Hwang, P.; Gochin, M. Enhanced potency of bivalent small molecule gp41 inhibitors. *Bioorg. Med. Chem.* **2017**, *25*, 408–420. [CrossRef]
37. Gochin, M.; Cai, L. The Role of Amphiphilicity and Negative Charge in Glycoprotein 41 Interactions in the Hydrophobic Pocket. *J. Med. Chem.* **2009**, *52*, 4338–4344. [CrossRef]
38. Steger, H.K.; Root, M.J. Kinetic dependence to HIV-1 entry inhibition. *J. Biol. Chem.* **2006**, *281*, 25813–25821. [CrossRef]
39. Kahle, K.M.; Steger, H.K.; Root, M.J. Asymmetric deactivation of HIV-1 gp41 following fusion inhibitor binding. *PLoS Pathog.* **2009**, *5*, e1000674. [CrossRef]
40. Guttman, M.; Garcia, N.K.; Cupo, A.; Matsui, T.; Julien, J.P.; Sanders, R.W.; Wilson, I.A.; Moore, J.P.; Lee, K.K. CD4-induced activation in a soluble HIV-1 Env trimer. *Structure* **2014**, *22*, 974–984. [CrossRef]
41. Pan, J.; Peng, H.; Chen, B.; Harrison, S.C. Cryo-EM Structure of Full-length HIV-1 Env Bound with the Fab of Antibody PG16. *J. Mol. Biol.* **2020**, *432*, 1158–1168. [CrossRef]
42. Krieger, E.; Vriend, G. YASARA View—Molecular graphics for all devices—From smartphones to workstations. *Bioinformatics* **2014**, *30*, 2981–2982. [CrossRef]
43. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]
44. Krieger, E.; Nielsen, J.E.; Spronk, C.A.E.M.; Vriend, G. Fast empirical pKa prediction by Ewald summation. *J. Mol. Graph. Model.* **2006**, *25*, 481–486. [CrossRef]
45. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65*, 712–725. [CrossRef]
46. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R.D.; Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **2003**, *31*, 3784–3788. [CrossRef]
47. Luo, P.; Baldwin, R.L. Mechanism of helix induction by trifluoroethanol: A framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **1997**, *36*, 8413–8421. [CrossRef]
48. Sarzotti-Kelsoe, M.; Bailer, R.T.; Turk, E.; Lin, C.; Bilska, M.; Greene, K.M.; Gao, H.; Todd, C.A.; Ozaki, D.A.; Seaman, M.S.; et al. Optimization and validation of the TZM-bl assay for standardized assessments of neutralizing antibodies against HIV-1. *J. Immunol. Methods* **2014**, *409*, 131–146. [CrossRef]

Supplementary Material to Publication III

Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion.

International Journal of *Molecular Sciences*

MDPI

Supplementary information to:

Conformational stabilization of gp41-mimetic miniproteins opens up new ways of inhibiting HIV-1 fusion

Mario Cano-Muñoz[1,2,*], Julie Lucas[2], Lin Li-Yun[2], Samuele Cesaro[1,†], Christiane Moog[2] and Francisco Conejero-Lara[1,*]

[1] Departamento de Química Física, Instituto de Biotecnología y Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain; mariocano@ugr.es, conejero@ugr.es

[2] INSERM U1109, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, 67084 Strasbourg, France; julie.lucas@etu.unistra.fr (J.L.); li-yun.lin@etu.unistra.fr (L.L.-Y.); c.moog@unistra.fr (C.M.)

* Correspondence: mariocano@ugr.es (M.C.-M.); conejero@ugr.es (F.C.-L.); Tel.: +34-958242371 (F.C.-L.)

† Present address: Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biological Chemistry, University of Verona, Strada Le Grazie 8, 37134 Verona, Italy.

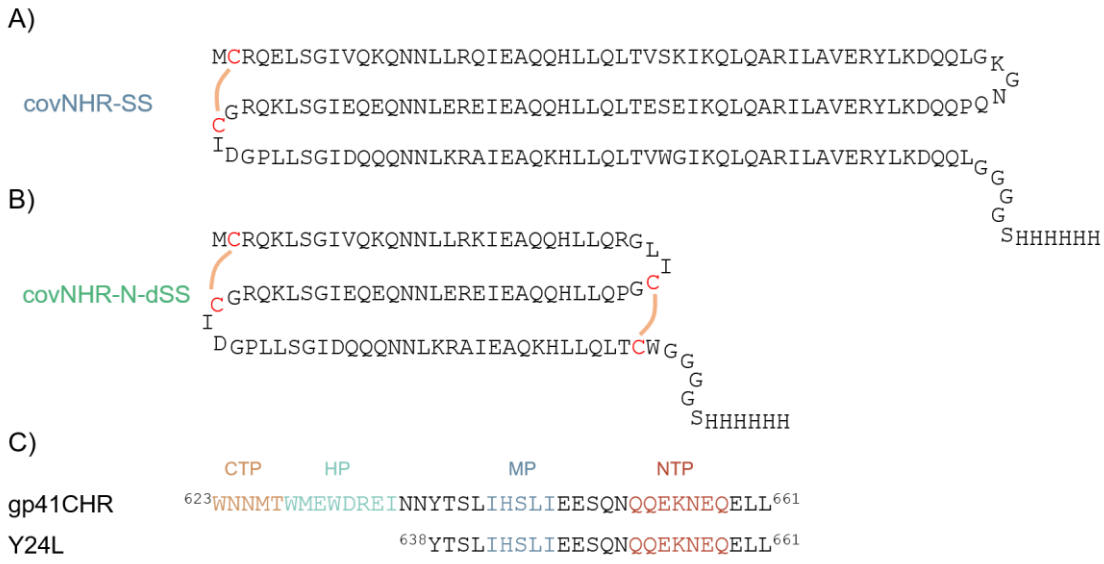


Figure S1. Sequences and topology of the covNHR miniproteins and peptides. Mutations in red were engineered in this work. A) covNHR-SS sequence. B) covNHR-N-dSS sequence C) CHR C34 and Y24L peptide sequences. Residues corresponding to the different binding pocket motifs are colored as follows: CTP in orange, HP in cyan, MP in blue and NTP in red.

Table S1. Description of envelope glycoproteins used in the ELISA binding experiments.

Abbreviation	Type	Description	Reference
JRFL	Gp140	A soluble uncleaved gp140 Env stabilized trimer derived from HIV-1 JRFL containing a C-terminal foldon sequence.	[1]
MN/LAI	Gp160	A hybrid oligomeric gp160 Env with gp120 derived from HIV-1 MN and gp41 derived from HIV-1 LAI.	[2]
THO23/LAI	Gp160	Recombinant Env with gp120 from HIV-1 92TH023 linked to gp41 from LAI, with a deletion in the immunodominant region.	[3]
ZM 4096	Gp140	Synthetic construct derived from gp140 sequence of the codon-optimized HIV-1 96ZM651.	[4]
CN54	Gp140	Recombinant Env trimer containing gp120+gp41 ectodomain from HIV-1 CN54.	[5]

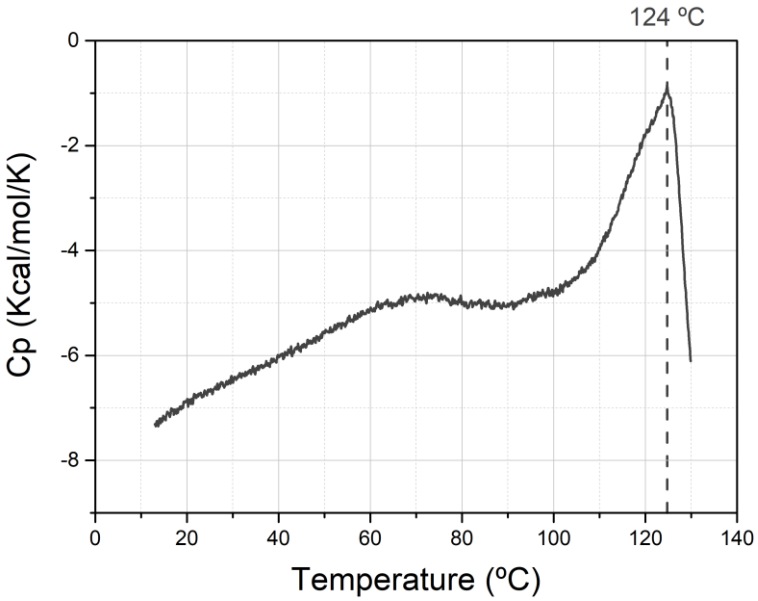


Figure S2. Differential scanning calorimetry thermogram of the thermally induced denaturation of covNHR-SS. The DSC thermogram was recorded at a scan rate of 2°C·min^{-1} at a concentration of 30 μM.

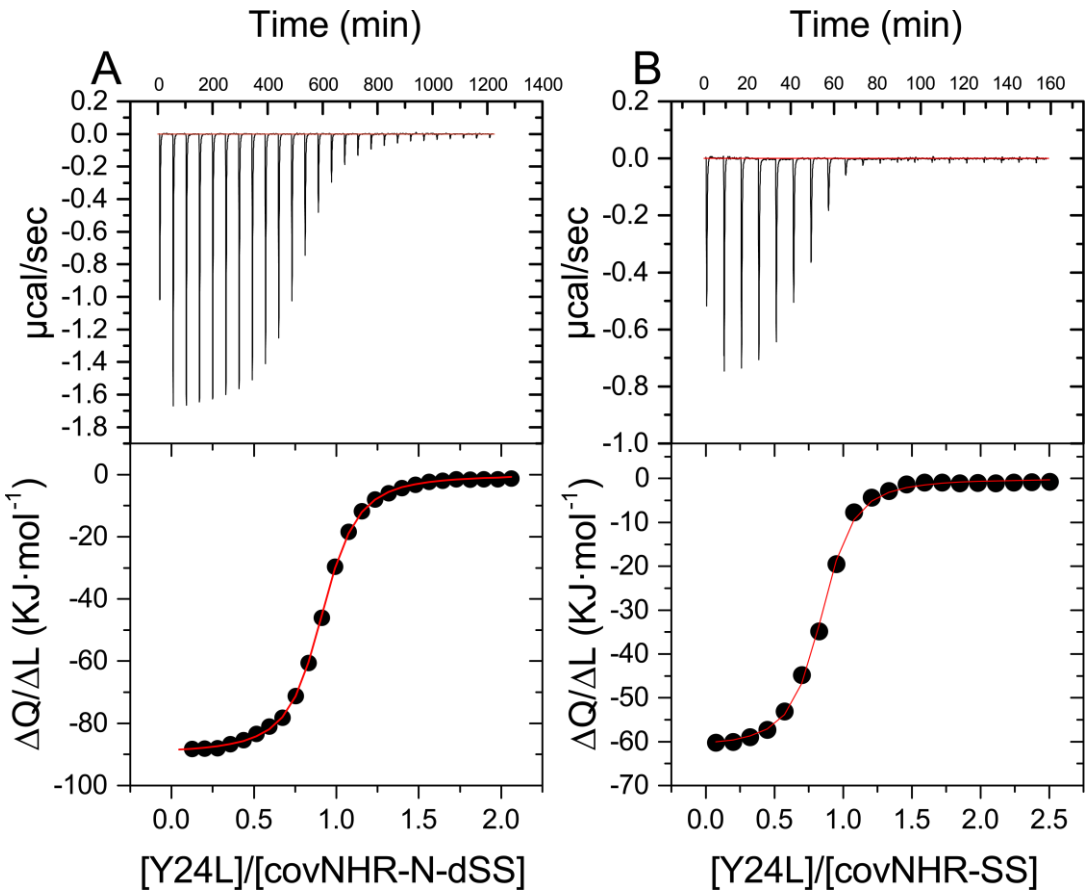


Figure S3. Isothermal titration calorimetry experiments of Y24L peptide binding to the covNHR proteins. **A)** Y24L binding to covNHR-N-dSS and B) to covNHR-SS. The experiments were measured at 25 °C by titration of 10 μM of each miniprotein in the cell with ~300 μM of Y24L peptide from the syringe. The upper panels show the experimental ITC thermograms and the lower panels the normalized binding isotherms. The symbols in the lower panels correspond to the

experimental heats normalized per mole of injected peptide and the lines represent the best fittings using a binding model of n identical and independent sites.

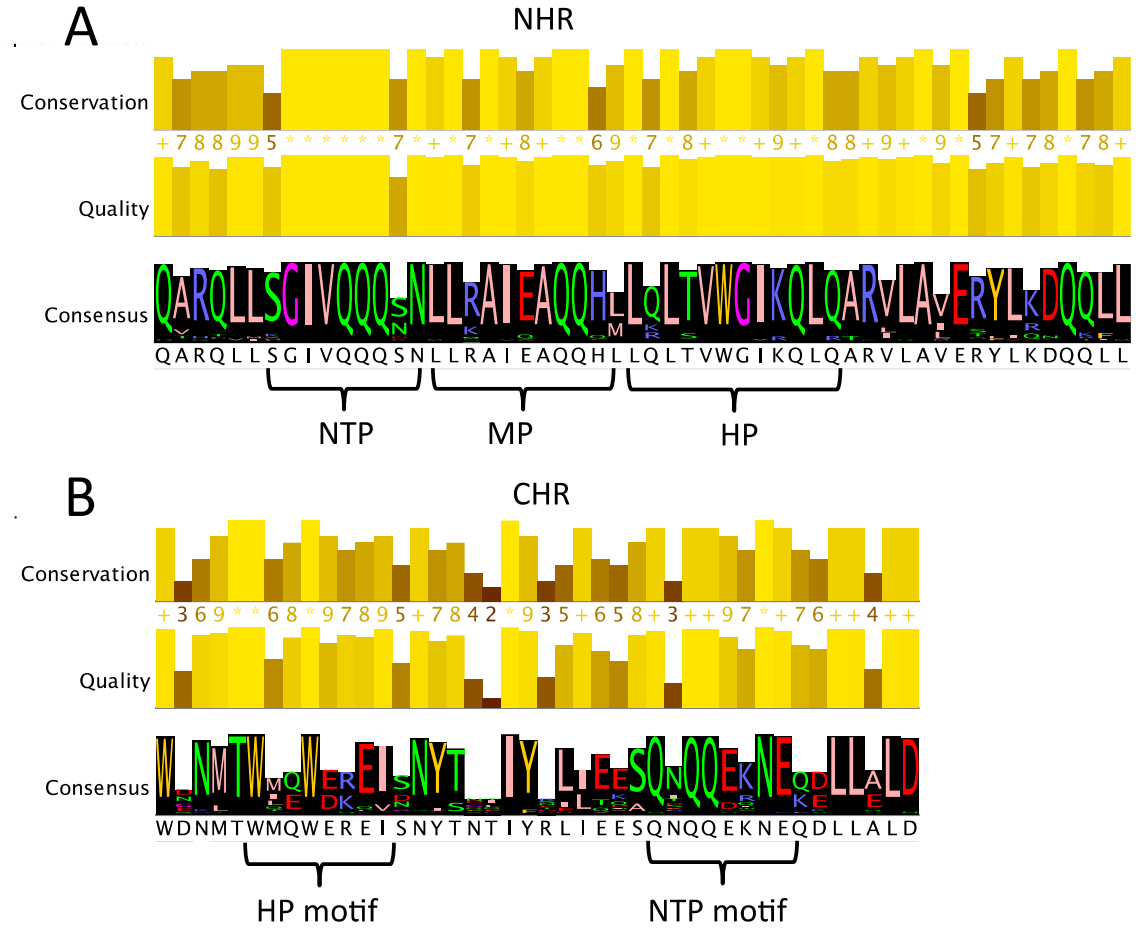


Figure S4. Sequence conservation and consensus sequence of NHR and CHR in gp41. The plots have been made with Jalview [6] using the 2018 Compendium sequence alignment [7] from Los Alamos Sequence Database (https://www.hiv.lanl.gov).

References:

1. Chakrabarti, B. K.; Feng, Y.; Sharma, S. K.; McKee, K.; Karlsson Hedestam, G. B.; LaBranche, C. C.; Montefiori, D. C.; Mascola, J. R.; Wyatt, R. T. Robust Neutralizing Antibodies Elicited by HIV-1 JRFL Envelope Glycoprotein Trimers in Nonhuman Primates. *J. Virol.* **2013**, *87*, 13239–13251.
2. Pialoux, G.; Excler, J. L.; Rivière, Y.; Gonzalez-Canali, G.; Feuillie, V.; Coulaud, P.; Gluckman, J. C.; Matthews, T. J.; Meignier, B.; Kieny, M. P.; et al. A Prime-Boost Approach to HIV Preventive Vaccine Using a Recombinant Canarypox Virus Expressing Glycoprotein 160 (MN) followed by a Recombinant Glycoprotein 160 (MN/LAI). *AIDS Res. Hum. Retroviruses* **1995**, *11*, 373–381.
3. Rerks-Ngarm, S.; Pitisuttithum, P.; Excler, J. L.; Nitayaphan, S.; Kaewkungwal, J.; Premsri, N.; Kunasol, P.; Karasavvas, N.; Schuetz, A.; Ngauy, V.; et al. Randomized, double-blind evaluation of late boost strategies for HIV-uninfected vaccine recipients in the RV144 HIV vaccine efficacy trial. *J. Infect. Dis.* **2017**, *215*, 1255–1263.
4. Zurawski, G.; Shen, X.; Zurawski, S.; Tomaras, G. D.; Montefiori, D. C.; Roederer, M.; Ferrari, G.; Lacabaratz, C.; Klucar, P.; Wang, Z.; et al. Superiority in Rhesus Macaques of Targeting HIV-1 Env gp140 to CD40 versus LOX-1 in Combination with Replication-Competent NYVAC-KC for Induction of Env-Specific Antibody and T Cell Responses. *J. Virol.* **2017**, *91*, 1–20.
5. Cranage, M. P.; Fraser, C. A.; Cope, A.; McKay, P. F.; Seaman, M. S.; Cole, T.; Mahmoud, A. N.; Hall, J.; Giles, E.; Voss, G.; et al. Antibody responses after intravaginal immunisation with trimeric HIV-1CN54 clade C gp140 in Carbopol gel are augmented by systemic priming or boosting with an adjuvanted formulation. *Vaccine* **2011**, *29*, 1421–1430.
6. Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191.

7. HIV Sequence Compendium 2018. Brian Foley, Thomas Leitner, Cristian Apetrei, Beatrice Hahn, Ilene Mizrachi, James Mullins, Andrew Rambaut, Steven Wolinsky, and Bette Korber editors. 2018. Publisher: Los Alamos National Laboratory, Theoretical Biology and Biophysics, Los Alamos, New Mexico. LA-UR-18-25673.

Publication IV

Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity

PNAS

Main Manuscript for

Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity.

Mario Cano-Muñoz[1], Daniel Polo-Megías[1], Ana Cámara-Artigas[2], José A. Gavira[3], María J. López-Rodríguez[1], Géraldine Laumond[4], Sylvie Schmidt[4], Julien Demiselle[5], Seiamak Bahram[4], Christiane Moog[4,6,*] and Francisco Conejero-Lara[1,*]

[1]Departamento de Química Física, Instituto de Biotecnología y Unidad de Excelencia de Química Aplicada a Biomedicina y Medioambiente (UEQ), Facultad de Ciencias, Universidad de Granada, 18071 Granada, Spain.

[2]Department of Chemistry and Physics, Agrifood Campus of International Excellence (ceiA3) and CIAMBITAL, University of Almeria, Carretera de Sacramento s/n, 04120 Almeria, Spain.

[3]Laboratory of Crystallographic Studies, IACT, (CSIC-UGR), Avenida de las Palmeras 4, 18100 Armilla, Granada, Spain.

[4]Laboratoire d'ImmunoRhumatologie Moléculaire, Institut National de la Santé et de la Recherche Médicale (INSERM) UMR_S 1109, Institut thématique interdisciplinaire (ITI) de Médecine de Précision de Strasbourg, Transplantex NG, Faculté de Médecine, Fédération Hospitalo-Universitaire OMICARE, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Université de Strasbourg, Strasbourg, France.

[5]UMR_S1260, NanoRegMed, University of Strasbourg, Strasbourg, France.

[6]Vaccine Research Institute (VRI), Créteil, France.

*Correspondence to Christiane Moog and Francisco Conejero-Lara

Email: c.moog@unistra.fr; conejero@ugr.es

Author Contributions: F.C-L. and C.M. conceived the project, planned and supervised the research, analyzed and interpreted the results. F.C-L. and M.C-M. designed the chimeric proteins. M.C-M., D.P-M. and M.J.L-R. produced, purified and characterized biophysically the proteins. M.C-M. analyzed the binding of the proteins and the HR2 peptides. M.C-M. carried out the binding experiments with trimeric spike and antibody sera. D.P-M. carried out the

crystallization tests. J.A.G. and A.C-A. supervised the crystallization of the proteins and acquired the X-ray diffraction data. A.C-A. solved the protein-peptide complex structure. G.L. and S.S. performed the inhibitory activity assays of the proteins on SARS-CoV-2 strains, J.D. collected the sera from a cohort of COVID-19 patients, S.B. supervised the research. F.C-L. wrote the article with input from all authors.

Competing Interest Statement: The University of Granada, the University of Strasbourg and the Institut National de la Santé et de la Recherche Médicale (INSERM) have filed a patent application (P202230188) on the chimeric proteins described in this article. Inventors: F.C-L., M.C-M. and C.M.

Classification: Applied Biological Sciences.

Keywords: COVID-19, protein design, immunity, fusion inhibitor, vaccines.

This PDF file includes:

Main Text
Figures 1 to 4

Abstract

SARS-CoV-2 spike (S) protein mediates virus attachment to the cells and fusion between viral and cell membranes. Membrane fusion is driven by mutual interaction between the highly conserved heptad-repeat regions 1 and 2 (HR1 and HR2) of the S2 subunit of the spike. For this reason, these S2 regions are interesting therapeutic targets for COVID-19. Although HR1 and HR2 have been described as transiently exposed during the fusion process, no significant antibody responses against these S2 regions have been reported. Here we designed chimeric proteins that imitate highly stable HR1 helical trimers and strongly bind to HR2. The proteins have broad inhibitory activity against WT B.1 and BA.1 viruses. Sera from COVID-19 convalescent donors showed significant levels of reactive antibodies (IgG and IgA) against the HR1 mimetic proteins, whereas these antibody responses were absent in sera from uninfected donors. Moreover, both inhibitory activity and antigenicity of the proteins correlate positively with their structural stability but not with the number of amino acid changes in their HR1 sequences, indicating a conformational and conserved nature of the involved epitopes. Our results reveal previously undetected spike epitopes that may guide the design of new robust COVID-19 vaccines and therapies.

Significance Statement

Inhibition of SARS-CoV-2 fusion to the host cell is a promising therapeutic approach against SARS-CoV-2 infection. Here we show how new chimeric proteins can be designed and engineered to imitate a highly conserved region of the spike S2 subunit of SARS-CoV-2 and interfere with a key step during the viral-cell fusion process. The proteins inhibit SARS-CoV-2 replication in vitro, including the original WT D612G variant and the Omicron variant. Moreover, the proteins capture antibody responses present in sera from COVID-19 convalescent patients

and therefore can imitate spike epitopes of immunological relevance. These chimeric molecules are very easy to produce in recombinant form and may have a strong therapeutic potential.

Main Text

Introduction

Since SARS-CoV-2 emerged in late 2019, a huge amount of research developed has increased dramatically our understanding of the molecular basis of the disease and has provided novel strategies to fight it. Vaccines based in immunization with the spike (S) protein, either using mRNA or viral vectors, have been very successful to decrease viral transmission and severity of the disease (1). However, the immunity of vaccinated people appears to decline after few months from vaccination and new variants of the virus that may escape the protection of current vaccines appear continuously (2, 3). Moreover, only a limited number of antivirals have been approved to treat the infection, making it essential to continue the development of new treatments and immunization strategies.

As in other coronaviruses, SARS-CoV-2 spike (S) protein decorates the virus surface and promotes its entry into the host cells. Like other Class-I fusion proteins, the spike protein is a trimer of heterodimers composed of S1 and S2 subunits (4, 5). The S1 subunit consists of the N-terminal domain (NTD), the receptor-binding domain (RBD) and two C-terminal domains (CTD). Three S1 subunits cover the S2 trimer and lock it in its prefusion conformation. S2 contains in a fusion peptide (FP) and two heptad-repeat regions (HR1 and HR2) that are essential to promote membrane fusion. The S2 prefusion structure is organized around a coiled-coil trimer formed by its central helix (CH) and the connector domain (CD), located between the HR1 and HR2 regions. The protein is embedded in the viral membrane by a transmembrane (TM) segment that is followed by an internal short C-terminal tail (CT).

Cell infection by SARS-CoV-2 starts with S1 binding to the angiotensin-converting enzyme 2 receptor (ACE2) using the receptor-binding domain (RBD). Then, proteolysis of the S2 subunit at the S2' site mediated by a host proteases (TMPRSS2 in plasma membrane or Cathepsins in the endosomes (6)) triggers a conformational transition in which HR1 becomes extended to continue the CH trimeric coiled-coil and produces the insertion of FP into the cell membrane (7). Then, S1 becomes shed and S2 folds on itself forming six-helix bundle (6-HB) structure, in which three HR2 pack externally against the hydrophobic grooves of a central trimeric HR1 helical bundle (5, 8). This process brings into close proximity the viral and cell membranes promoting their fusion and subsequent insertion of viral content inside the cell.

Because of their importance in viral fusion, HR1 and HR2 are potential targets for coronavirus treatment (8, 9). Moreover, these regions have particularly high sequence conservation in SARS-CoV-2 (10), as well as between different coronaviruses (9, 11). There is however a lack of structural information about HR2 in the prefusion structure of the spike. Current high-resolution structures of the pre-fusion spike do not resolve the stalk region connecting the S head with the transmembrane region (4). In situ cryo-electron tomography combined with molecular dynamics has characterized the S2 stalk as highly flexible (12), with two coiled-coil regions separated by flexible hinges (13). Interestingly, the SARS-CoV-2 spike is highly glycosylated (14, 15) and HR2 contains two N-glycosilation sites at residues Asn1173 and Asn1194, which may protect this

highly preserved region from antibody access although its flexibility could make it vulnerable to other types of molecules and small compounds.

S2-mediated fusion of SARS-CoV-2, as well as other coronaviruses, is inhibited by peptides derived from HR2 (8, 9, 16, 17), similarly to that observed for gp41 mediated HIV-1 fusion (18, 19). HR1-based peptides are much less potent inhibitors (16, 17) but stabilized mimics of a trimeric helical bundle based on the HIV-1 gp41 HR1 that target HR2 have shown potent inhibitory activity of HIV-1 fusion (20-22) and have also been found to inhibit infection by human coronaviruses (23). Moreover, a 5-helix construct based on the S2 6-HB structure but lacking one HR2 region has been reported very recently to inhibit several SARS-CoV-2 variants (24). All this evidence indicates that both HR1 and HR2 are at least transiently exposed during coronavirus fusion and susceptible to inhibition, as it has been reported for HIV-1. Vaccination with stabilized gp41 HR1 trimers can elicit neutralizing antibodies against HIV-1 (25) and HIV-1 infected patients elicit neutralizing antibodies that target exposed HR1 epitopes (26, 27). Moreover, human mAbs directed against HR1 and HR2 regions of SARS-CoV show broad neutralizing activity (28). Given the similarities between the class-I proteins mediating viral fusion mechanisms, it is conceivable that similar antibody responses may be elicited during SARS-CoV-2 infection. However, this type of antibody responses has so far passed unnoticed.

Antibody responses produced by SARS-CoV-2 infection are mostly induced against the nucleocapsid protein and the spike protein and, within this, against the S1 subunit (29, 30). Neutralizing Ab responses in COVID-19 convalescent patients are mainly directed to the RBD but mutations in emerging SARS-CoV-2 variants drastically reduce their neutralization potential (3, 31, 32). Neutralizing responses targeting S2 are scarcer than against S1 but they are more interesting (33) because S2 is more conserved than S1. Moreover, S2-targetting neutralizing antibody responses that cross-react with other related coronaviruses have been reported, even in uninfected individuals (34). This type of memory B cell immunity may confer durable broad coronavirus protection (35).

Epitope mapping has so far detected only a few immunogenic epitopes in S2 (36) mostly located at sequence region 765-835, comprising the FP and the S2 cleavage site, and residues 1140-1160 at the S2 stem region immediately upstream of HR2. Both epitopes have been described as neutralizing (37). Also, a neutralizing mAb targeting the 1140-1160 linear helical epitope has been recently described (38). There is however a lack of highly immunogenic epitopes within HR1 and HR2 described in the literature. A possible reason is that HR1 and HR2 are highly protected from the immune system due to their importance in the conformational changes driving membrane fusion. It is also possible that conformational variability of these regions could make their sensitive epitopes to be only transiently exposed and poorly immunogenic. Though, immune responses to conformational epitopes may have gone unnoticed in epitope mapping studies using linear peptides.

We previously designed HR1 mimetic proteins targeting HIV gp41 (22, 39-43). These chimeric proteins consist in single polypeptide chains that stably and spontaneously fold as trimeric helical bundles with helix-loop-helix-loop-helix topology. These proteins imitate accurately an exposed HR1 groove and tightly bind to gp41 HR2 peptides with identical HR1-HR2 interactions to the post-fusion gp41 structure (41, 42). Moreover, the proteins have a potent and broad inhibitory activity against a variety of HIV-1 strains (22, 39, 41, 43). Because of the structural

similarities between the post-fusion structures of gp41 and S2, we hypothesized that similar single-chain chimeric proteins could be designed to mimic an exposed SARS-CoV-2 HR1 and that these proteins could also have antiviral activity.

Results

Design of SARS-Cov-2 HR1 mimetic proteins.

We designed several single-chain proteins imitating the coiled-coil structure of the SARS-CoV-2 HR1 trimer using a similar strategy as previously employed with HIV-1 gp41 (22). The chimeric proteins, subsequently named CoVS-HR1, were designed as antiparallel helical trimers with a helix-loop-helix-loop-helix topology, in which the second helix was reversed in both sequence and spatial orientation. The helices were tethered with manually built, four-residue loops. Then, three different variants (codenamed L3A, L3B and L3C) were created carrying out different sets of amino acid substitutions in order to stabilize their trimeric coiled-coil structure (See SI Methods, Figs. S1-S2 and Table S1 for details). The CoVS-HR1 proteins were produced in recombinant form by overexpression in *E. coli* with high yields and could be easily purified by two-step standard chromatographic methods. All the proteins were highly soluble (≥10 mg mL^{-1}), mainly monomeric and acquired spontaneously the predicted α-helical structure according to their circular dichroism (CD) spectra (Fig. 1a-b). The three proteins were very stable against thermal denaturation monitored by differential scanning calorimetry (DSC), showing melting temperatures above 80°C (Fig 1c) but L3B and L3C showed larger and sharper denaturation peaks than L3A, indicating more stable and cooperative structures in the two former variants. Among them, L3C was more thermostable than L3B by about 13°C, as a result of higher α-helical propensity of its sequence produced by substitution of glycine residues.

CoVS-HR1 proteins bind the S2 HR2 region.

Binding of the CoVS-HR1 proteins to an HR2-derived peptide V39E (spike sequence 1164-1202) slightly increased the α-helix structure as observed in the CD spectra (Fig. 1a). In presence of the HR2 peptide, the denaturation peaks in the DSC thermograms showed an increase of their area, as a result of peptide dissociation accompanying unfolding at high temperature (Fig 1c). Titration of the proteins with the V39E peptide by isothermal titration calorimetry (ITC) showed a 1:1 binding stoichiometry, dissociation constants in the low nM range and considerably negative binding enthalpy (Fig. 1d and Table S2). Interestingly, although the binding affinities to the HR2 peptide were similar for the three protein variants, the favorable binding enthalpy decreased in magnitude in the order L3A > L3B > L3C, whereas the binding entropy changed from negative to positive with the same order, indicating a strong enthalpy-entropy compensation (Fig 1e). This suggests that complex formation involves an entropy cost contribution due to considerable conformational tightening of the HR1 helical bundle, as previously observed for homologous HIV-1 gp41-HR1 mimetic proteins (39, 42). This entropy penalty decreases, however, as a result of a higher intrinsic stability of the HR1 helical bundle in the most stable variant L3C.

To investigate the structural details of the interaction between the CoVS-HR1 proteins and their HR2 target we produced crystals of the L3B and L3C proteins in complex with the V39E peptide.

Crystals of good quality for X-ray diffraction could be only obtained for the L3B-V39E complex. Diffraction data were collected to 1.45 Å resolution (Table S3) allowing a very detailed structure characterization (Fig 2). The structure of the L3B protein is highly similar to the designed model (Fig 2a). The conformation of the V39E peptide in the complex and the interactions at the protein-peptide interface were virtually identical to those observed between HR1 and HR2 in the post-fusion S2 structure (8) (Fig. 2c). The high resolution of the structure allowed the visualization of fine details of the HR1-HR2 interface (Fig S3). The interaction is the result of a large number of hydrophobic HR2 residues that insert their side chains along the narrow HR1 crevice, flanked by a considerable number of hydrogen bonds, some of them involving interfacial water molecules. Moreover, the HR1 coiled-coil contains several water-filled internal cavities that establish hydrogen-bond networks with HR1 buried polar residues (Fig S4).

The three protein variants also bound strongly to recombinant trimeric spike in ELISA assays (Fig. 3a). This binding was specific of the HR2 target region because it was competitively inhibited by addition of the V39E peptide. Moreover, chimeric proteins that mimic HIV-1 gp41 HR1 (41, 44) did not show significant binding. Residual binding in presence of V39E varied in the order L3C > L3B > L3A consistently with their stability.

CoVS-HR1 proteins inhibit SARS-CoV-2 infection.

Because the CoVS-HR1 proteins can make strong interactions with the HR2 region structurally similar to those driving membrane fusion, we tested their capacity to inhibit cell infection by SARS-CoV-2. Vero 76 cells were infected with SARS-CoV-2 primary viruses (B1 D614G and BA.1 Omicron genotypes) in presence and absence of CoVS-HR1 proteins (Fig. 3b-c). The three proteins showed inhibitory activities for both viruses, indicating that the proteins recognize the conserved HR2 target sequence in the two strains. This is consistent with their capacity to bind HR2 and thereby to block the conformational transition of S2 promoting membrane fusion and infection. Strikingly, despite a similar binding affinity for the HR2 peptide, L3C showed much higher inhibitory activity for both virus variants than L3A and L3B. This highlights the importance of conformational stability of the HR1 helical bundle in these designs.

CoVS-HR1 proteins detect Ab responses against SARS-CoV-2 infection.

Since the CoVS-HR1 proteins faithfully mimic a fully functional HR1 structure of S2, we asked whether these chimeric proteins could also reproduce immunogenic epitopes in S2 with relevance during SARS-CoV-2 infection. Thus, we tested the IgG and IgA reactivity against the CoVS-HR1 proteins in sera from convalescent patients collected three months after the onset of COVID-19. At that time, most patients' sera showed a sustained SARS-CoV-2-specific IgG response against RBD (Fig. 4a) and nucleocapside protein and also displayed significant neutralizing activity (45). Interestingly, we also detected significant IgG responses against the three protein variants. This response was not detectable in the sera from uninfected donors, indicating a development of SARS-CoV-2-specific immune responses against HR1. Although the mean Ab response against L3C was slightly lower than against RBD, some patients displayed high levels of anti-L3C Abs, similar or higher than the anti-RBD Abs levels. Moreover, the mean IgG binding clearly correlated positively with the relative conformational stability and inhibitory activity of the proteins, with the most stable protein L3C showing the highest IgG binding, which

strongly suggests that these epitopes are of conformational nature. This may explain why no immunogenic HR1 epitopes had been detected so far by peptide-based epitope scans.

Interestingly, the IgA reactivity in the sera against the L3C variant was even higher than against RBD (Fig 4b). These results clearly indicate that the CoVS-HR1 mimetic proteins are antigenic and imitate functionally relevant HR1 epitopes in S2 that elicit strong Ab responses during the course of the infection. Furthermore, the respective antibody response directed against L3C and RBD was extremely different between IgG and IgA (Fig 4c) further suggesting an intrinsically different B cell maturation process against these two antigens.

Discussion

Peptides derived from S2 HR1 region had been previously dismissed as potential fusion inhibitors (46) due to their intrinsic instability and high propensity to aggregate, similarly to that observed for gp41 HR1 in HIV-1. However, the stabilization of the HR1 trimeric coiled-coil structure has resulted in potent anti-HIV fusion inhibitors targeting HR2 (20, 22). We demonstrated here that a similar strategy also works for SARS-CoV-2 and, likely, for other related coronaviruses. A trimeric HR1 trimer design based of HIV-1 gp41-derived sequences has also shown a moderate but significant pan-coronavirus activity, suggesting relatively low specificity requisites for significant inhibition (23). Our mimetic proteins specifically based on the SARS-CoV-2 S2 HR1 sequence constitute a novel approach to design this type of HR1-based fusion inhibitors targeting the highly conserved HR2 region. This approach is further supported by a recently reported 5-helix construct with potent and broad inhibitory activity (24). Our strategy consisted in engineering a single-chain polypeptide that can fold autonomously to mimic a highly stable HR1 surface, without any chemical modification or addition of external trimerization motifs. This makes these proteins very easy to produce and handle, strongly facilitating their development as potential drugs or vaccines.

Recent results point to an Omicron spike evolved beyond immune evasion toward a more compact architecture with a well-regulated fusion machinery, altered dynamics and easy to release fusion peptide (47). Noteworthy, the CoVS-HR1 proteins showed similar inhibitory potency for BA.1 Omicron compared to the ancestral B.1 virus. Since the Omicron variant does not have any mutation at the HR2, it is likely that this preserved region is similarly exposed for both variants. On the other hand, BA.1 Omicron has three significant mutations at HR1, namely Q954H, N969K and L981F. Inspection of the equivalent amino acids in the L3B-V39E complex suggests that these amino acid substitutions, due to their conservative nature, should not have a strong impact on the HR1-HR2 interaction (Fig. S5). Targeting these highly conserved fusion domains is therefore of strong interest.

HR1 becomes transiently exposed after spike activation, since HR2 peptides are potent fusion inhibitors (46). Nonetheless, it was previously unknown whether or not this exposure was sufficient to elicit a robust Ab response during infection. In fact, no significant Ab responses against HR1 had been reported so far. The strong Ab binding activity against the L3C CoVS-HR1 protein in the sera of convalescent patients indicates that HR1 epitopes are in fact immunogenic during the course of the disease, suggesting a frequent and stable epitope exposure. Also, the highest Ab response measured for L3C compared to the other two variants may be attributed to

an improved Ab affinity for more stable HR1 epitopes displayed by L3C. This result, together with the fact that L3C contains the highest number of amino acid substitutions of the three variants, strongly suggests that the detected Ab responses are directed against HR1 conformational epitopes.

Elicitation of cryptic or transient conformational epitopes is a challenge for vaccine design. Protein structural scaffolds have been previously used to transplant complex epitopes that improve immunogenic responses (48-50). A desirable property of these designs is a high structural stability that strongly correlates with antibody binding and with strong antigenicity. Here we show that our single-chain chimeric proteins can accurately mimic immunologically relevant and highly stable HR1 epitopes. The structural stability of these epitopes are clearly correlated with their inhibitory activity of cell infection by interacting with HR2, as well as with their antigenicity against antibodies developed during COVID-19 disease. These results make the CoVS-HR1 proteins potential candidates as novel potent therapeutics or new vaccine compositions against COVID-19.

Materials and Methods

Protein models were built and analyzed using SwissPDBviewer (51) and YASARA structure (52) using as template the six-helical bundle formed by HR1 and HR2 in the S2 post-fusion structure (PDB id. 6LXT (8)). The DNA encoding the protein sequences, containing a C-terminal histidine tag, were synthesized and cloned into pET303 expression vectors by Thermo Fisher Scientific (Waltham, MA, USA). The three different protein constructs were produced by overexpression in *E. coli* cells (BL21(DE3)), purified by a two-step chromatographic method and biophysically characterized by CD, DSC and scattering measurements as described in SI Appendix. The thermodynamic parameters of binding between the CoVS-HR1 constructs and the synthetic peptide V39E (Genecust, Luxembourg), encompassing the S2 HR2 region (residues 1164-1202), were investigated by ITC. The complex between the CoVS-HR1-L3B protein and the V39E peptide was crystallized from solutions at 1:2 protein:peptide ratio using the sitting-drop vapor-diffusion method. Diffraction data were collected to 1.45 Å resolution at the beamline XALOC at the ALBA synchrotron (Barcelona, Spain), from which the structure was solved as detailed in SI Appendix. The complex coordinates were deposited at the Protein Data Bank under the accession code 7ZR2. The binding of the proteins to recombinant trimeric spike (ProSci Inc., Poway, CA) was assessed by ELISA. The *in vitro* inhibitory activity against infection of Vero 76 cells by SARS-CoV-2 viruses was analyzed using WT B1 (UK D614G genotype) or BA.1 (Omicron) strains. Binding of CoVS-HR1 proteins to IgG and IgA in sera of infected patients after 3 months of infection was determined by ELISA and compared with sera from healthy patients. Comprehensive experimental details are given in SI Appendix.

Acknowledgments

This work was supported by grants CV20.26565 from the Consejería de Economía y Conocimiento, Junta de Andalucía (Spain), PID2019.107515RB.C21 from the Spain's State Research Agency (SRA/10.13039/501100011033), and co-funded by ERDF/ESF, "A way to make Europe"/"Investing in your future". The work performed in C.M.'s laboratory was supported by grants from ANRS (Agence Nationale de Recherches sur le SIDA et les hépatites virales), the

Investissements d'Avenir program managed by the ANR under reference ANR-10-LABX-77 and EHVA (N°681032, Horizon 2020). Work in S.B.'s laboratory was supported by grants from the Agence Nationale de la Recherche (ANR) (ANR-11-LABX-0070_TRANSPLANTEX), the INSERM (UMR_S 1109), the Institut Universitaire de France (IUF), all the University of Strasbourg (IDEX UNISTRA), the European regional development fund (European Union) INTERREG V program (project n°3.2 TRIDIAG) and MSD-Avenir grant AUTOGEN. We are grateful to the Spanish Radiation Synchrotron Source (ALBA), Barcelona, Spain and the European Synchrotron Radiation Facility (ESRF), Grenoble, France, for the provision of time and staff assistance at XALOC (ALBA) and ID30B and ID23-2 (ESRF) beamlines during diffraction data collection. We thank María Carmen Salinas-García for her assistance in carrying out the crystallization screenings. We also thank Pilar González-García for her help with the statistical analysis.

Data availability

The crystallographic structure of the CoVS-HR1-L3B protein in complex with the V39E peptide has been deposited at the Protein Data Bank with accession code 7ZR2. The data that support all the findings of this study are available from the corresponding authors on reasonable request; see author contributions for specific data sets.

References

1. H. Fathizadeh *et al.*, SARS-CoV-2 (Covid-19) vaccines structure, mechanisms and effectiveness: A review. *Int. J. Biol. Macromol.* **188**, 740-750 (2021).
2. R. K. Gupta, Will SARS-CoV-2 variants of concern affect the promise of vaccines? *Nat. Rev. Immunol.* **21**, 340-341 (2021).
3. S. S. A. Karim, Q. A. Karim, Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet* **398**, 2126-2128 (2021).
4. D. Wrapp *et al.*, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
5. Y. F. Cai *et al.*, Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592 (2020).
6. M. Hoffmann *et al.*, SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 (2020).
7. M. A. Tortorici, D. Veesler, "Structural insights into coronavirus entry" in Complementary Strategies to Understand Virus Structure and Function, F. A. Rey, Ed. (2019), vol. 105, pp. 93-116.
8. S. Xia *et al.*, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **30**, 343-355 (2020).
9. S. Xia *et al.*, A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. *Sci. Adv.* **5** (2019).
10. S. Laha *et al.*, Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect. Genet. Evol.* **85**, 104445 (2020).

11. S. W. Liu *et al.*, Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* **363**, 938-947 (2004).
12. P. V. Raghuvamsi *et al.*, SARS-CoV-2 S protein:ACE2 interaction reveals novel allosteric targets. *Elife* **10**, e63646 (2021).
13. B. Turonova *et al.*, In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203-208 (2020).
14. L. Casalino *et al.*, Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 spike Protein. *ACS Central Sci.* **6**, 1722-1734 (2020).
15. Y. Watanabe *et al.*, Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat. Commun.* **11**, 2688 (2020).
16. B. J. Bosch *et al.*, Severe acute respiratory syndrome coroavirus (SARS-CoV) infection inhibition using spike protein heptad repeat-derived peptides. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8455-8460 (2004).
17. D. Schuetz *et al.*, Peptide and peptide-based inhibitors of SARS-CoV-2 entry. *Adv. Drug Deliv. Rev.* **167**, 47-65 (2020).
18. J. M. Kilby *et al.*, Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nature medicine* **4**, 1302-1307 (1998).
19. Y. He *et al.*, Design and evaluation of sifuvirtide, a novel HIV-1 fusion inhibitor. *J. Biol. Chem.* **283**, 11126-11134 (2008).
20. D. M. Eckert, P. S. Kim, Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11187-11192 (2001).
21. M. J. Root, M. S. Kay, P. S. Kim, Protein design of an HIV-1 entry inhibitor. *Science* **291**, 884-888 (2001).
22. S. Crespillo *et al.*, Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti-HIV-1 drugs. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18207-18212 (2014).
23. C. Wang *et al.*, Supercoiling Structure-Based Design of a Trimeric Coiled-Coil Peptide with High Potency against HIV-1 and Human beta-Coronavirus Infection. *J. Med. Chem.* **65**, 2809-2819 (2021).
24. L. Xing *et al.*, A Five-Helix-Based SARS-CoV-2 Fusion Inhibitor Targeting Heptad Repeat 2 Domain against SARS-CoV-2 and Its Variants of Concern. *Viruses-Basel* **14**, 597 (2022).
25. E. Bianchi *et al.*, Vaccination with peptide mimetics of the gp41 prehairpin fusion intermediate yields neutralizing antisera against HIV-1 isolates. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 10655-10660 (2010).
26. D. Corti *et al.*, Analysis of Memory B Cell Responses and Isolation of Novel Monoclonal Antibodies with Neutralizing Breadth from HIV-1-Infected Individuals. *PLoS One* **5**, e8805 (2010).
27. C. Sabin *et al.*, Crystal Structure and Size-Dependent Neutralization Properties of HK20, a Human Monoclonal Antibody Binding to the Highly Conserved Heptad Repeat 1 of gp41. *PLoS Pathog.* **6**, e1001195 (2010).
28. H. A. Elshabrawy, M. M. Coughlin, S. C. Baker, B. S. Prabhakar, Human Monoclonal Antibodies against Highly Conserved HR1 and HR2 Domains of the SARS-CoV spike Protein Are More Broadly Neutralizing. *PLoS One* **7**, e50366 (2012).
29. H. W. Jiang *et al.*, SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat. Commun.* **11**, 3581 (2020).

30. S. Ghotloo *et al.*, Epitope mapping of neutralising anti-SARS-CoV-2 monoclonal antibodies: Implications for immunotherapy and vaccine design. *Rev. Med. Virol.* 10.1002/rmv.2347, e2347-e2347 (2022).
31. M. Sakharkar *et al.*, Prolonged evolution of the human B cell response to SARS-CoV-2 infection. *Sci. Immunol.* **6** (2021).
32. S. Lusvarghi *et al.*, SARS-CoV-2 BA.1 variant is neutralized by vaccine booster-elicited serum, but evades most convalescent serum and therapeutic antibodies. *Sci. Transl. Med.* 10.1126/scitranslmed.abn8543, eabn8543-eabn8543 (2022).
33. P. Shah, G. A. Canziani, E. P. Carter, I. Chaiken, The Case for S2: The Potential Benefits of the S2 Subunit of the SARS-CoV-2 spike Protein as an Immunogen in Fighting the COVID-19 Pandemic. *Front. Immunol.* **12**, 637651 (2021).
34. K. W. Ng *et al.*, Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* **370**, 1339-1343 (2020).
35. N.-C. Phuong *et al.*, S Protein-Reactive IgG and Memory B Cell Production after Human SARS-CoV-2 Infection Includes Broad Reactivity to the S2 Subunit. *Mbio* **11**, e01991-01920 (2020).
36. L. Farrera-Soler, J.-P. Daguer, S. Barluenga, N. Winssinger, Experimental Identification of Immuno- dominant B-cell Epitopes from SARS-CoV-2. *Chimia* **75**, 276-284 (2021).
37. Y. Li *et al.*, Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 1095-1097 (2020).
38. P. Zhou *et al.*, A protective broadly cross-reactive human antibody defines a conserved site of vulnerability on beta-coronavirus spikes. *bioRxiv* 10.1101/2021.03.30.437769, 2021.2003.2030.437769 (2021).
39. M. Cano-Munoz *et al.*, Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity. *Biomolecules* **11**, 566 (2021).
40. M. Cano-Munoz, S. Jurado, B. Morel, F. Conejero-Lara, Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors. *Int. J. Biol. Macromol.* **192**, 90-99 (2021).
41. S. Jurado *et al.*, Structural and Thermodynamic Analysis of HIV-1 Fusion Inhibition Using Small gp41 Mimetic Proteins. *J. Mol. Biol.* **431**, 3091-3106 (2019).
42. S. Jurado, M. Cano-Munoz, D. Polo-Megias, F. Conejero-Lara, B. Morel, Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects. *Arch. Biochem. Biophys.* **688**, 108401 (2020).
43. S. Jurado *et al.*, Probing Vulnerability of the gp41 C-Terminal Heptad Repeat as Target for Miniprotein HIV Inhibitors. *J. Mol. Biol.* **432**, 5577-5592 (2020).
44. M. Cano-Munoz *et al.*, Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion. *Int. J. Mol. Sci.* **23**, 2794 (2022).
45. R. Carapito *et al.*, Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort. *Sci. Transl. Med.* 10.1126/scitranslmed.abj7521, eabj7521-eabj7521 (2021).
46. S. Xia *et al.*, Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell. Mol. Immunol.* **17**, 765-767 (2020).
47. S. M. C. Gobeil *et al.*, Structural diversity of the SARS-CoV-2 Omicron spike. *bioRxiv* 10.1101/2022.01.25.477784 (2022).

48. G. Ofek *et al.*, Elicitation of structure-specific antibodies by epitope scaffolds. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17880-17887 (2010).
49. M. L. Azoitei *et al.*, Computation-Guided Backbone Grafting of a Discontinuous Motif onto a Protein Scaffold. *Science* **334**, 373-376 (2011).
50. F. Sesterhenn *et al.*, De novo protein design enables the precise induction of RSV-neutralizing antibodies. *Science* **368**, eaay5051 (2020).
51. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).
52. E. Krieger, G. Vriend, YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981-2982 (2014).

Figures

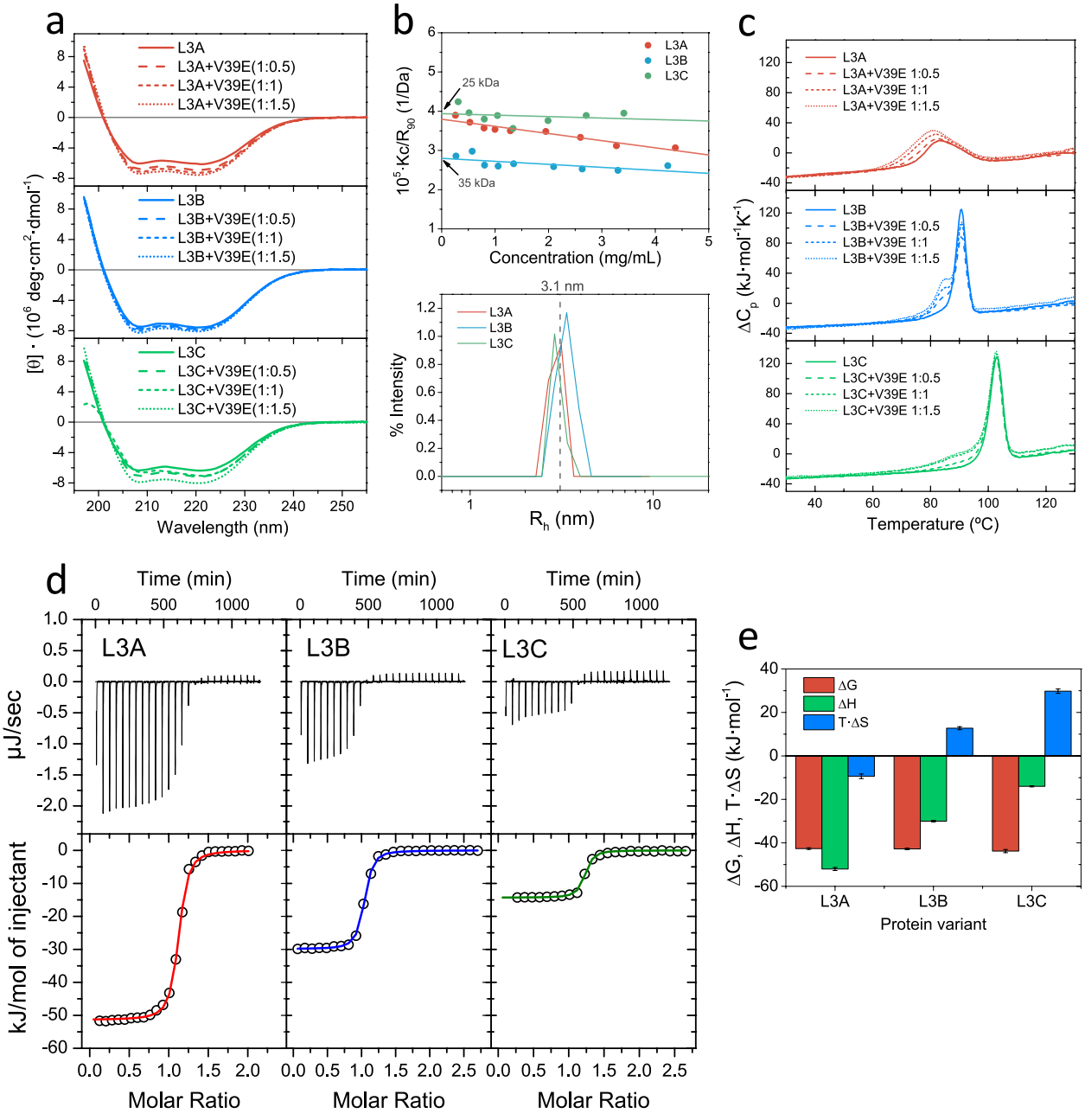


Figure 1. Biophysical characterization of CoVS-HR1 mimetic proteins. a) Far-UV Circular Dichroism (CD) spectra showing highly α-helical structures for the three protein variants. Addition of HR2-derived V39E peptide increases slightly the α-helical structure content due to its binding to the proteins in a helical conformation. **b)** Particle size analysis by light scattering. Upper panel shows the Debye plot made from scattering intensity measurements as a function

of concentration. The intercepts indicate the mean-weighted molar masses. Lower panel shows the hydrodynamic radius distributions measured by dynamic light scattering. **c)** Thermal stability analysis by differential scanning calorimetry (DSC). The denaturation peaks indicate high thermal stability for the three protein variants. The thermograms measured in presence of V39E peptide show increased heats of denaturation due to the dissociation of the peptide from the complex. **d)** Isothermal titration calorimetry (ITC) analysis of the binding of the HR2-derived V39E peptide to the different CoVS-HR1 proteins. The upper panels show the experimental thermograms with negative heats of binding. The lower panels show the normalized binding isotherms fitted using a model of independent binding sites. The binding stoichiometry is 1:1 (see Table S2). **e)** Thermodynamic magnitudes of binding of the HR2 peptide to each protein variant calculated from the ITC data. Error bars correspond to 95% confidence intervals of the parameters estimated from the fits in panel d.

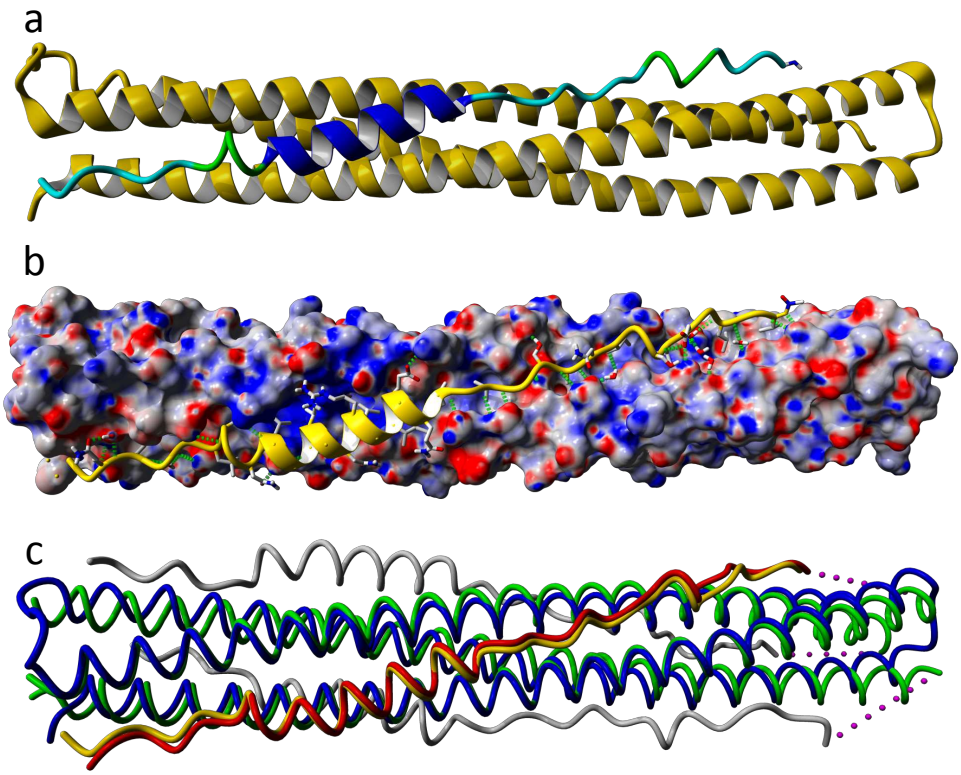


Figure 2. Crystal structure of the CoVS-HR1-L3B protein in complex with the HR2 V39E peptide. **a)** Ribbon representation showing the backbone topology and secondary structure of the complex. The L3B protein is colored in yellow and the V39E peptide in color gradation from blue to green according to helical structure. **b)** Molecular surface of the L3B protein colored according to the electrostatic potential with red indicating negative, white neutral and blue positive. The V39E peptide is represented with a yellow ribbon. The side chains of the peptide and interfacial water molecules that interact with the protein have been highlighted with sticks colored in CPK scheme. The protein-peptide interaction is highly hydrophobic along the binding crevice, flanked by polar and electrostatic interactions at along the edges. Interfacial hydrogen bonds are represented with green dashed lines. **c)** Structure alignment between the backbones of the CoVS-HR1-V39E complex (blue and yellow) and the HR1-HR2 post-fusion complex structure (green and red) (PDB id. 6LXT (8)). Alignment has been made using 142 matching residues corresponding to the two parallel HR1 helices (helix 1 and 3) of the CoVS-HR1 protein and the HR2 V39E peptide.

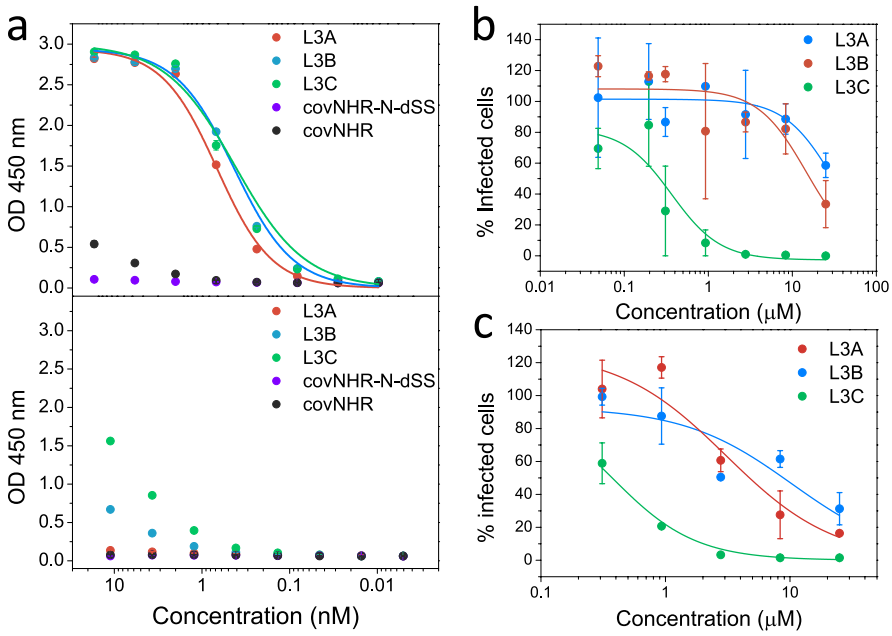


Figure 3. SARS-CoV-2 inhibition by interaction of CoVS-HR1 proteins with HR2. a) Upper panel: Strong binding of the CoVS-HR1 proteins to recombinant trimeric spike. ELISA plates were coated with trimeric spike and tested with 0.3-8.0 nM CoVS-HR1 proteins or with homologous HIV-1 gp41-HR1 mimetic proteins (39, 42). Binding was detected with Anti-6X Histag antibody conjugated to horseradish peroxidase. Binding of gp41-HR1 mimetic proteins was negligible. Lower panel: Similar ELISA experiment was done in presence of HR2 V39E peptide in a 2:1 peptide:protein ratio. The binding was strongly inhibited for all proteins. **b**, **c)** Inhibition of SARS-CoV-2 virus infection of Vero 76 cells by the CoVS-HR1 proteins. Vero 76 cells were infected by primary viruses (B1 D614G genotype, panel b, and Omicron BA.1, panel c) in the presence of CoVS-HR1 proteins at different concentrations. The percentage of infected cells was recorded after measurement of the number of infected cells (by anti-nucleocapside antibody) and the total number of cells (by SYTOX Green staining) using an imaging cytometer. Error bars correspond to standard deviations from duplicates.

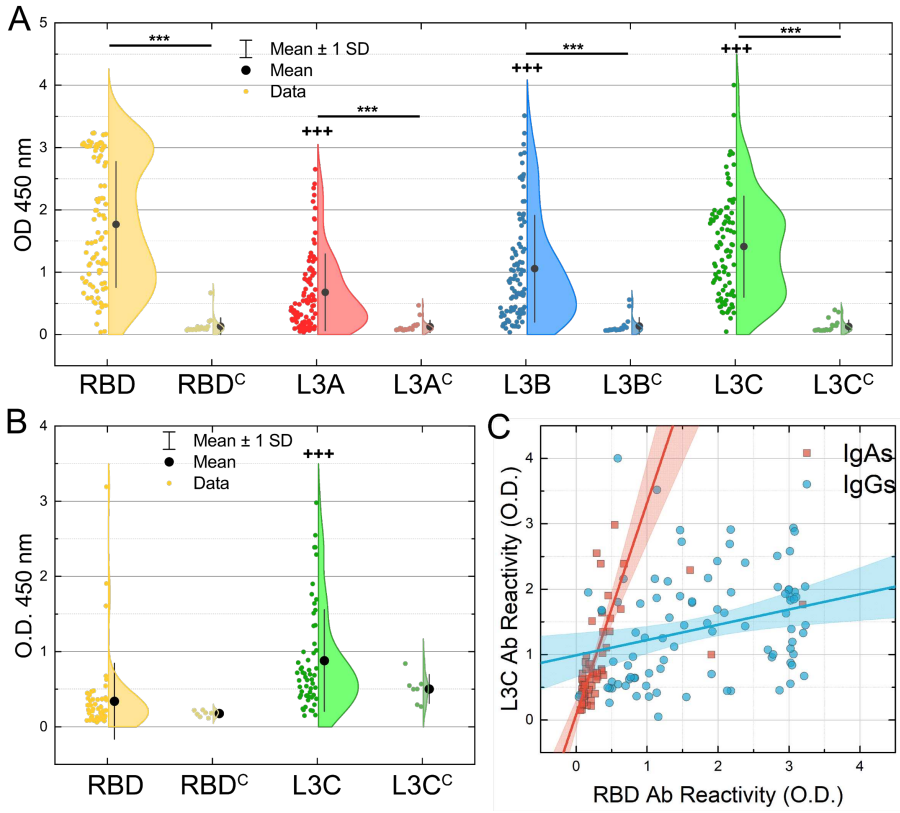


Figure 4. Antigenicity of the CoVS-HR1 proteins against COVID-19 convalescent patient's sera. **a)** Reactivity levels of IgG against CoVS-HR1 proteins in infected patient sera 3 months after infection compared to reactivity against RBD. ELISA plates were coated with CoVS-HR1 proteins or His-tagged RDB and tested with 1/1000 diluted sera of patients for 30 min. Control sera from healthy patients was also tested. Antibody binding was measured with goat anti-human IgG antibody conjugated to horseradish peroxidase. **b)** Recognition levels of IgAs present in infected patient sera 3 months after infection towards CoVS-HR1-L3C compared to recognition to RBD. **c)** IgG and IgA reactivity against L3C protein vs RBD in convalescent plasma of SARS-CoV-2 patients. Linear regression and Anova test were performed with the following values: IgAs vs RDB: Pearson's r = 0.75; $p < 0.0001$; n = 53. IgGs vs RBD: Pearson's r = 0.29; $p > 0.05$; n = 90.

Supplementary Material to Publication IV

Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity

PNAS

Supplementary Information for

Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity.

Mario Cano-Muñoz[1], Daniel Polo-Megías[1], Ana Cámara-Artigas[2], José A. Gavira[3], María J. López-Rodríguez[1], Géraldine Laumond[4], Sylvie Schmidt[4], Julien Demiselle[5], Seiamak Bahram[4], Christiane Moog[4,6,*] and Francisco Conejero-Lara[1,*]

*Correspondence to Christiane Moog and Francisco Conejero-Lara
Email: c.moog@unistra.fr; conejero@ugr.es

This PDF file includes:

SI Methods
Figures S1 to S5
Tables S1 to S3

SI Methods

Protein design and engineering

Modelling was carried out using SwissPDBviewer (1) and YASARA structure (2). As template, we used the published X-ray crystal structure of the six-helical bundle formed by HR1 and HR2 in the S2 post-fusion structure (PDB id. 6LXT (3)). To model an antiparallel trimer of helices, the HR2 chains were deleted from the model and one of the HR1 helices (Spike residues 914-988) was upturned and aligned to the original one. Due to the antiparallel orientation of the reversed helix, the side chain CA-CB bonds have different spatial orientation compared the native one. This could produce improper side chain packing with the other helices in the coiled-coil structure. To compensate this, the alignment was made trying to superimpose the CB atoms of the core residue side chains. To preserve the correct core coiled-coil packing, the amino acid sequence of the upturned helix was also reversed. Then, side chain clashes were removed by energy minimization. Four-residue loops were manually built to connect each pair of helix termini to create a helix-loop-helix-loop-helix topology. The amino acid composition of the loops was selected using the built model with RossetaDesign web server http://rosettadesign.med.unc.edu (4). To enhance the stability of the antiparallel trimeric bundle and reduce exposed hydrophobicity, several surface-exposed residues at e and g positions of the coiled-coil heptad repeats were replaced by charged or polar amino acids to engineer proper salt bridges and hydrogen bonds between the antiparallel helix and the other two helices. Additional mutations at solvent exposed positions were made to increase net positive charge. No mutations were carried out to modify the hydrophobic groove between the two parallel HR1 helices to preserve the HR2 binding potential. Two variants code named L3A and L3B were engineered using different sets of amino acid substitutions. L3A substitutions were selected manually, whereas L3B substitutions were selected with the aid of RossetaDesign. A third variant, code named L3C, was engineered

from the L3B variant by substitution of glycine residues (Fig. S2) in the middle of α-helical regions by polar side chains, selected with RossetaDesign to increase α-helix propensity. Finally, all the models were subjected to energy minimization. The amino acid sequences are collected in Table S1.

Protein production and HR2 peptides

The DNA encoding the protein sequences were synthesized and cloned into pET303 expression vectors by Thermo Fisher Scientific (Waltham, MA, USA). The sequences included a N-terminal methionine and a C-terminal histidine tag with sequence GGGGSHHHHHH. *E. coli* bacteria (BL21(DE3)) were transformed with the plasmids and cultured at 37°C in presence of 30 mg·mL^{-1} Ampicillin. Protein expression was induced with 0.5 mM IPTG and the cells were cultured overnight at 37°C. Cells were collected by centrifugation and resuspended in lysis buffer (50 mM Tris/HCl, 500 mM NaCl, 1 mM EDTA, 1 mM β-mercaptoethanol) containing a cocktail of protease inhibitors (Sigma-Aldrich, St. Louis, MO). The cells were then lysed with three 30 s ultrasonication cycles on ice and the soluble and insoluble fractions were separated by 30 min ultracentrifugation at 4°C at 30000 rpm. The proteins were purified from the supernatant fraction by nickel-affinity chromatography. A second purification step was carried out by ion exchange chromatography on a HiTrap SP Sepharose XL column (Amersham GE Healthcare, UK). Protein purity was assessed by SDS-PAGE and the identity of each protein was confirmed by mass spectrometry.

Synthetic peptide V39E derived from the S2 HR2 sequence was acquired from Genecust (Luxembourg), with a purity >95%. The peptide (residues 1164-1202) was C-terminally tagged with a SGGY sequence to confer UV absorption at 280 nm. Protein and peptide concentrations were measured by UV absorption measurements at 280 nm with extinction coefficients calculated according to their respective amino acid sequences with the ExPasy ProtParam server (https://web.expasy.org/protparam/) (5).

Circular dichroism (CD)

CD measurements were performed with a Jasco J-715 spectropolarimeter (Tokyo, Japan) equipped with a temperature-controlled cell holder. Measurements of the far-UV CD spectra (260-200 nm) were made with a 1 mm path length quartz cuvette. Spectra were recorded at a scan rate of 100 nm/min, 1 nm step resolution, 1 s response and 1 nm bandwidth. The resulting spectrum was usually the average of 5 scans. Each spectrum was corrected by baseline subtraction using the blank spectrum obtained with the buffer and finally the CD signal was normalized to molar ellipticity ([θ], in deg·dmol^{-1}·cm^{2}). Thermal unfolding was monitored by measuring the CD signal at 222 nm as a function of temperature using a scan rate of 1 °C min^{-1}.

Molecular size characterization

The apparent hydrodynamic radii of the proteins were measured using dynamic light scattering (DLS) in a DynaPro MS-X DLS instrument (Wyatt, Santa Barbara, CA). Dynamics v6 software (Wyatt Technology Corporation, Santa Barbara, CA) was used in data collection and processing. Sets of DLS data were measured at 25 °C with an average number of 50 acquisitions and an acquisition time of 10 s.

Static scattering intensities were measured in a DynaPro MS-X DLS instrument (Wyatt, Santa Barbara, CA) or a Malvern μV instrument (Malvern Panalytical, Malvern, UK) at 25 °C, in 50 mM sodium phosphate buffer pH 7.4, at different concentrations of protein in a range of 0.2-4.5 mg mL^{-1}. The intensities were analysed using the Debye plot as represented by equation 1,

$$K \cdot c / R_{90} = 1 / M_w + 2A_2 c \qquad \textbf{(1)}$$

valid for particles significantly smaller than the wavelength of the incident radiation, where the K is an optical constant of the instrument, c is the particle mass concentration, R_{90} is the Rayleigh ratio of scattered to incident light intensity, Mw is the weight-averaged molar mass, A_2 is the 2^{nd} virial coefficient that is representative of inter-particle interaction strength. M_w can be determined from the intercept of the plot.

Differential scanning calorimetry

DSC experiments were carried out in a MicroCal PEAQ-DSC microcalorimeter equipped with autosampler (Malvern Panalytical, Malvern, UK). Scans were run from 5 to 130°C at a scan rate of 90 °C·h^{-1}. The experiments were carried out in 50 mM sodium phosphate buffer pH 7.4 and 50 mM glycine/HCl buffer pH 2.5. Protein concentration was typically 30 µM. Instrumental baselines were recorded before each experiment with both cells filled with buffer and subtracted from the experimental thermograms of the protein samples. Consecutive reheating runs were carried out to determine the reversibility of the thermal denaturation. The excess heat capacity (ΔC_p) relative to the buffer was calculated from the experimental DSC thermograms using Origin software (OriginLab, Northampton, MA).

Isothermal titration calorimetry

ITC measurements were carried out in a Microcal VP-ITC calorimeter (Malvern Panalytical, Malvern, UK). The proteins were titrated with 25 injections of 5 µL peptide solution at 480 s intervals. Protein concentration in the cell was around 10 µM, while the peptide concentration in the syringe was typically 200–300 µM. The experiments were carried out in 50 mM phosphate buffer (pH 7.4) at 25 °C. The experimental thermograms were baseline corrected and the peaks were integrated to determine the heats produced by each ligand injection. Residual heats due to unspecific binding or ligand dilution were estimated from the final peaks of the titrations. Each heat was normalized per mole of injected ligand. The resulting binding isotherms were fitted using a binding model of independent and equivalent sites, allowing the determination of the binding constant, K_b, the binding enthalpy, ΔH_b, and the binding stoichiometry, n.

Protein crystallization

The freshly purified covS-HR1 protein was dialyzed against 10 mM Tris pH 7.5 buffer and concentrated to ≈9 mg mL^{-1}. The V39E peptide was weighted and dissolved in deionized water. The pH was adjusted to pH 7.5, the solution was centrifuged at 14000 rpm in a microfuge for 30 min at 4°C and its concentration was measured by UV absorption at 280 nm. Then, the necessary volume to achieve the desired concentration was transferred to a clean microtube and lyophilised. Complex solutions (1:2 protein:peptide ratio) were prepared by dissolving the lyophilised peptide prepared previously with the appropriate volume of protein solution. After 15-20 min equilibration, the solution was again centrifuged at 14000 rpm for 30 min.

Screening for initial crystallization conditions was performed by the sitting-drop vapour-diffusion method using commercially available crystal screening kits Structure Screen 1 and 2 Eco Screen from Molecular Dimensions (Suffolk, UK). Droplets consisting of 2 µl complex solution and 2 µl reservoir solution were equilibrated at 298 K against 200 µl reservoir solution in 48-well MRC Maxi Optimization plates (Cambridge, UK). Several favourable conditions were initially identified and were optimized to obtain crystals. The best diffracting crystals were obtained in 0.1M sodium HEPES pH 7.5, 20% (w/v) PEG4000, 10% (v/v) isopropanol.

Diffraction data collection, structure solution and refinement

For data collection, crystals were flash-cooled in liquid nitrogen. Data sets were collected at 100 K at the beamline XALOC at the ALBA synchrotron (Barcelona, Spain), using a wavelength of 0.97926 Å (6). Diffraction data were indexed and integrated with the AutoPROC toolbox (7). Data

scaling was performed using the program Aimless (8) from the CCP4 suite (9). Data collection statistics are collected in Table S3. Solution and refinement of the structures were performed using the PHENIX suite (10). Molecular-replacement phasing using PHASER (11) was performed with the coordinates of the crystallographic structure of post fusion core of 2019-nCoV S2 subunit (PDB entry 6LXT (3)). Manual model building was performed using COOT (12, 13). Refinement was performed using phenix.refine in PHENIX (14). Quality of the structures were checked using MOLPROBITY (10) and PDB_REDO (15). Structural refinement statistics are collected in Table 1. All residues are in the allowed region of the Ramachandran plot. The complex coordinates were deposited at the Protein Data Bank under the accession code 7ZR2.

S protein binding assays

CovS-HR1 proteins ability to bind soluble trimeric Spike (S) proteins was determined by ELISA. Briefly, 96-well ELISA plates (Maxisorp, Nunc) were coated at 4 °C overnight with recombinant trimeric SARS-CoV-2 Spike protein (ProSci Inc., Poway, CA) in 0.1 M bicarbonate buffer (pH 9.6). After saturation with 2% BSA, 0.05% Tween in PBS for 1 h at 25 °C, 0.3-8.0 nM of covS-HR1 molecules (100 µL diluted in 1% BSA 0.05% Tween solution) were added and incubated for 2 h at room temperature. The plate was then washed five times and covS-HR1 binding was detected with 100 µL anti-6X Histag antibody conjugated to horseradish peroxidase (HRP) (Abcam) at 1/10000 dilution incubated for 1 hour at room temperature. Antibody binding was then revealed with tetramethylbenzidine (TMB) substrate buffer, the reaction was stopped with 1 M H_2SO_4 and optical density was read at 450 nm with a Molecular Device Plate Reader equipped with SoftMax Pro 6 program. Background binding was measured in plates without Spike protein and subtracted from the data. The percentage of binding was calculated using the readings with wells coated with His-tagged Spike incubated with PBS buffer instead of covS-HR1 molecules as control for 100% binding.

Virus inhibition assays

One day prior infection, Vero 76 cells were plated on a 96 well plate at 12500 cells/well. 50 µL of serial 4-fold dilutions of covS-HR1 proteins (2-fold concentrated) were incubated with the cells for 30 min. Cells were then infected by adding 50 µL WT SARS-CoV-2 viruses (B1 UK D614G or BA.1 Omicron genotypes) at Multiplicities Of Infection (MOI) of 80. After 2 days, cells are fixed with methanol for 20 min, washed with PBS and stained with anti-Nucleocapside Antibody (Genetex GTX135357) at 1/200 dilution in permwash (B&D) for 45 min at room temperature. Ab is revealed by incubation with a donkey anti-Rabbit monoclonal Ab (Alexa 647; A31573, Invitrogen) diluted at 1/200 in PBS 5% FCS for 45 min at room temperature. In parallel, total cells were detected by Sytox green (S7020, Invitrogen) staining. Total cells (Sytox green positive) and infected cells (nucleocapside positive) were counted using SpectraMax MiniMax Imaging Cytometer (Molecular Devices LLC). The percentage of infected cells was recorded. The 50% inhibitory concentration (IC50) was defined as the protein concentration leading to a 50 % reduction in the percentage of infected cells.

Detection of antibody responses in patients' sera

CovS-HR1 antigenicity against sera of infected patients was determined by ELISA. Sera samples were collected 3 months after recovery from COVID-19 infection. All patients and healthy donors gave their written informed consent (COVID-HUS ethics committee approved, reference CE: 2020-34). 96-well ELISA plates (Maxisorp, Nunc) were coated at 4 °C overnight with covS-HR1 molecules in 0.1 M bicarbonate buffer (pH 9.6). After blocking with 5% non-fat powdered milk in PBS for 1 h at 25 °C, 1/1000 diluted sera of patients after 3 of infection (100 µL diluted in 1% BSA 0.05% Tween solution) were added and incubated for 30 min at room temperature. The plate was then washed five times and covS-HR1 binding was detected with 100 µL goat anti-human IgG antibody conjugated to horseradish peroxidase (HRP) (Abcam Cambridge, UK) at 1/5000 dilution incubated for 1 hour at room temperature. Antibody binding was then revealed with tetramethylbenzidine (TMB) substrate buffer, the reaction was stopped with 1 M H_2SO_4 and

optical density was read at 450 nm with a Molecular Device Plate Reader equipped with SoftMax Pro 6 program. Background binding was measured in plates without covS-HR1 proteins and subtracted from the data. As positive control His-tagged RBD was used to detect the anti-RBD antibodies present in the patients sera. Control sera from healthy patients were also assayed. Data was analysed using Origin software (OriginLab, Northampton, MA).

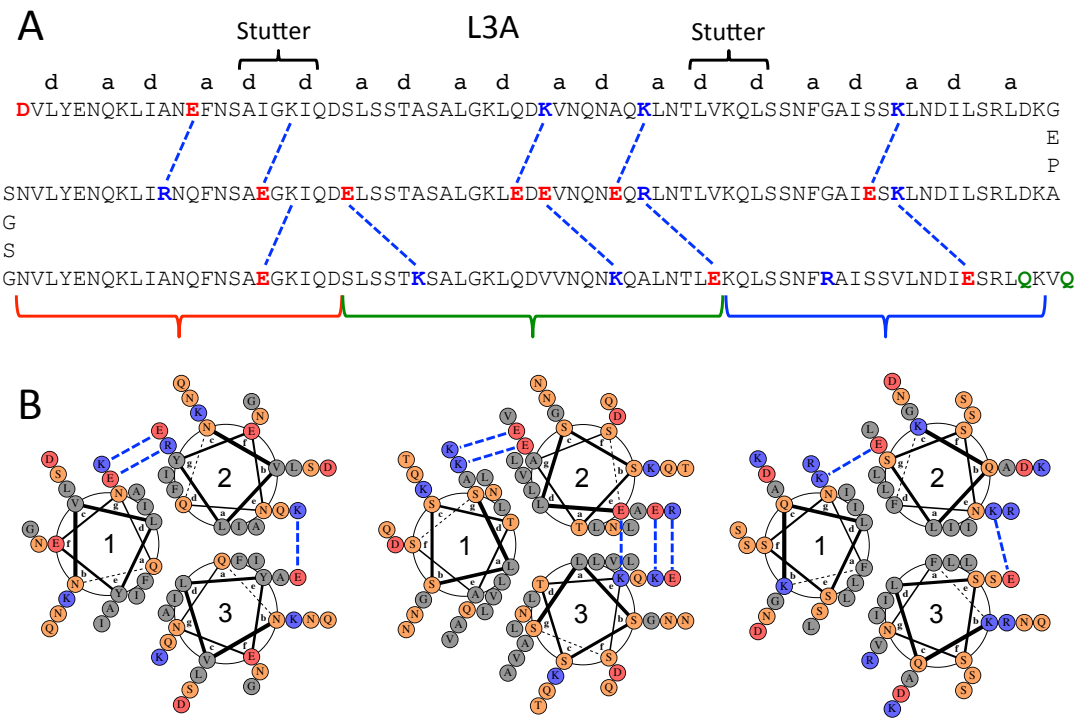


Fig. S1. Sequence and topology of CoVS-HR1 L3A protein. A) Sequence and topology of the L3A variant indicating the engineered amino acid substitutions (highlighted with colours) and the stabilizing interactions (with blue dashed lines). The ‘a’ and ‘d’ characters on top of the first sequence stretch indicate the corresponding positions in a heptad repeat. Two stutters in the canonical repeats are indicated. B) Helix-wheel diagrams corresponding to different sections of the coiled-coils (indicated with brackets under A). The diagrams have been drawn with Drawcoil (https://grigoryanlab.org/drawcoil/). The blue dashed lines indicate the engineered stabilizing interactions in each molecule.

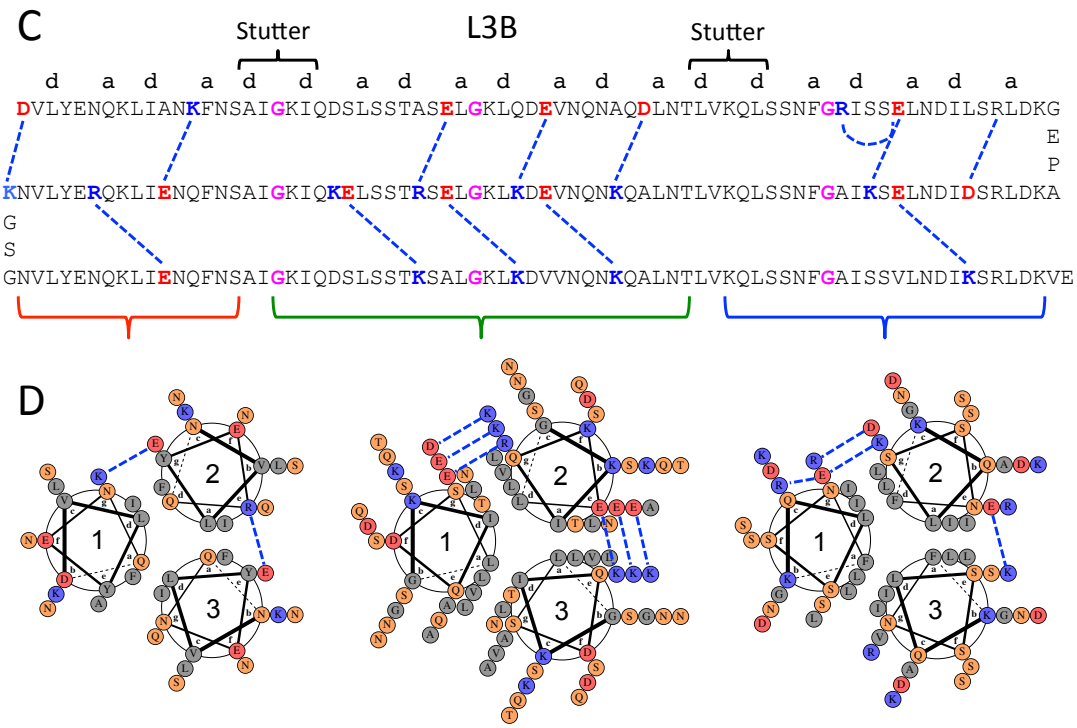


Fig. S2. Sequence and topology of CoVS-HR1 L3B protein. A) Sequence and topology of the L3B variant indicating the engineered amino acid substitutions (highlighted with colours) and the stabilizing interactions (with dashed lines). The ‘a’ and ‘d’ characters on top of the first sequence stretch indicate the corresponding positions in a heptad repeat. Two stutters in the canonical repeats are indicated. The glycine amino acids changed to polar amino acids in the L3C variant are highlighted in magenta. B) Helix- wheel diagrams corresponding to different sections of the coiled-coils (indicated with brackets under A). The diagrams have been drawn with Drawcoil (https://grigoryanlab.org/drawcoil/). The blue dashed lines indicate the engineered stabilizing interactions in each molecule.

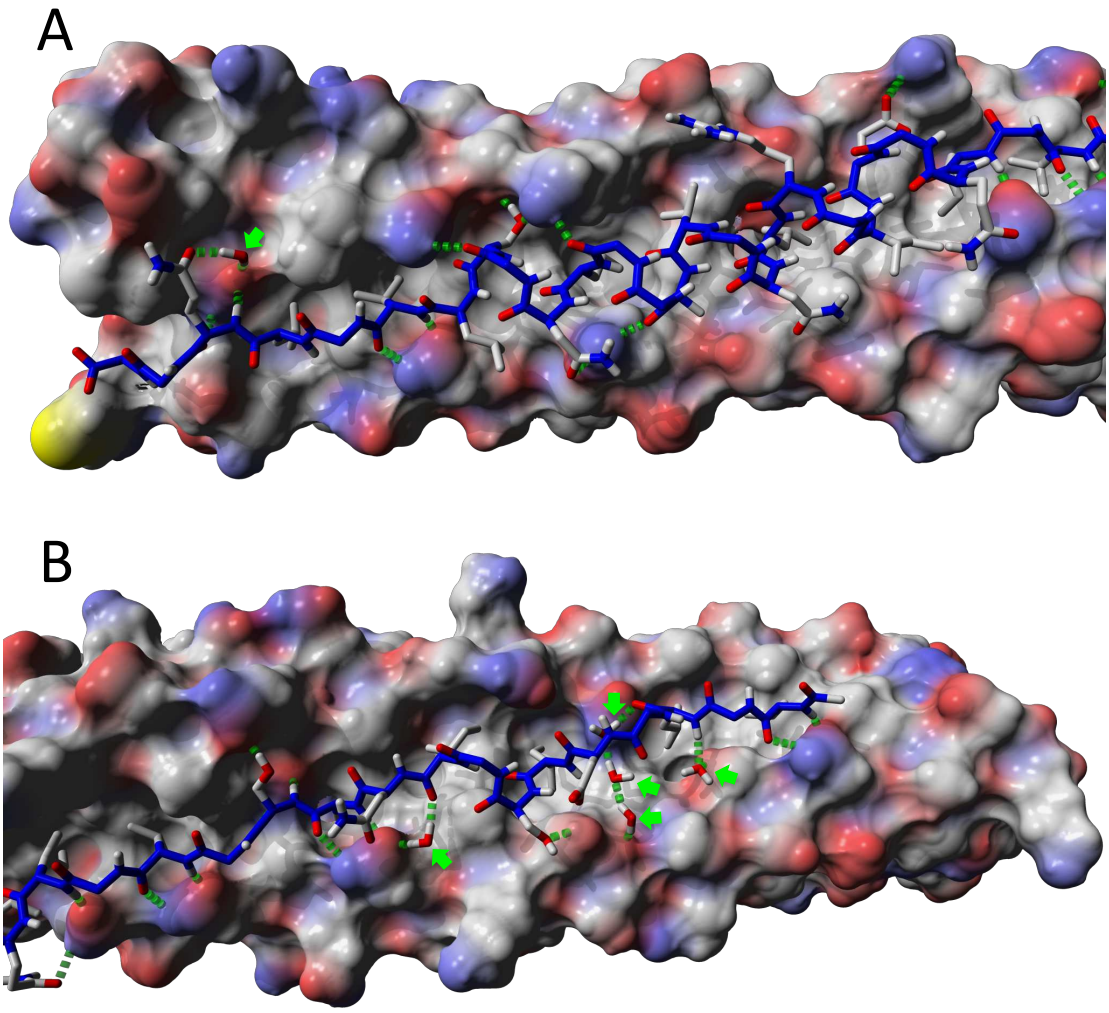


Fig. S3: Interfacial interactions between the CoVS-HR1 L3B protein and the V39E peptide in their crystallographic complex structure. Panel A shows the N-terminal half of the HR1 binding crevice. Panel B shows the C-terminal HR1 region. The L3B protein is represented as molecular surface and coloured according to element type (CPK scheme). The peptide is represented using sticks with backbone atoms coloured in blue and side chains involved in interactions with the protein are coloured in CPK scheme. Several buried water molecules at the protein peptide interface are highlighted with green arrows. Interfacial hydrogen bonds are represented in green dashed lines.

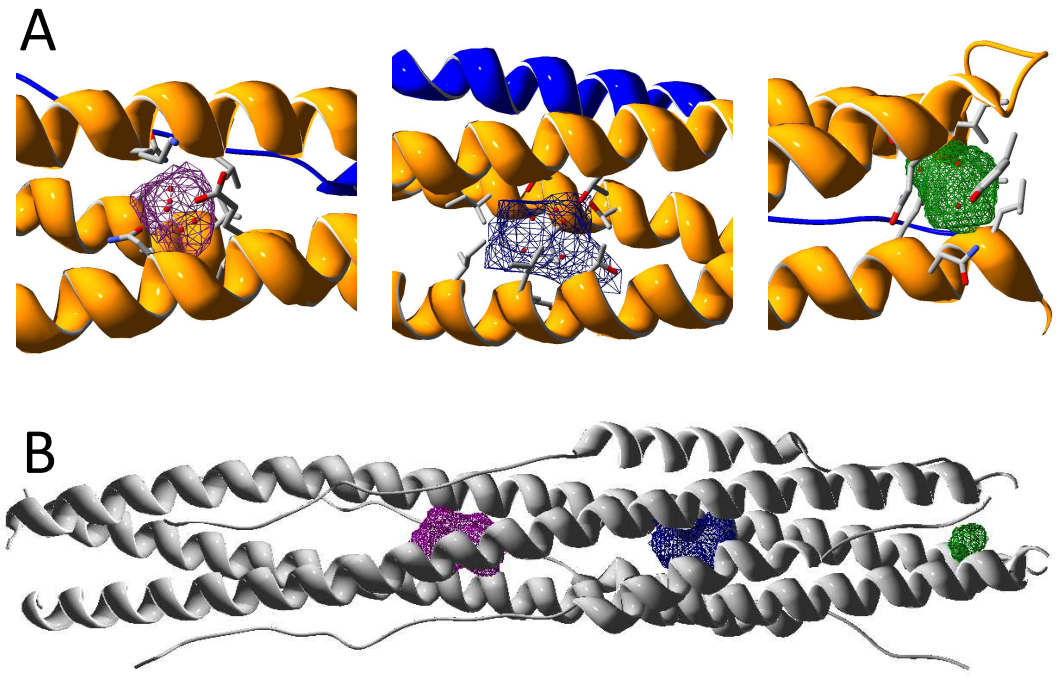


Fig. S4: Internal water-filled cavities inside the HR1 coiled-coil. A) Three sections of the L3B-V39E complex structure showing internal cavities in the HR1 trimeric coiled-coil filled with several buried water molecules well resolved in the crystallographic structure. The cavities are depicted with grids and the water molecules are represented in red spheres. Side chains of HR1 residues lining the cavities are highlighted in stick. B) Crystal structure of the 6-helix bundle post-fusion state of HR1 and HR2 in S2 (PDB id. 6LXT) showing similar cavities to those detected in the L3B-V39E complex structure. No water molecules were resolved in this structure because of a lower resolution.

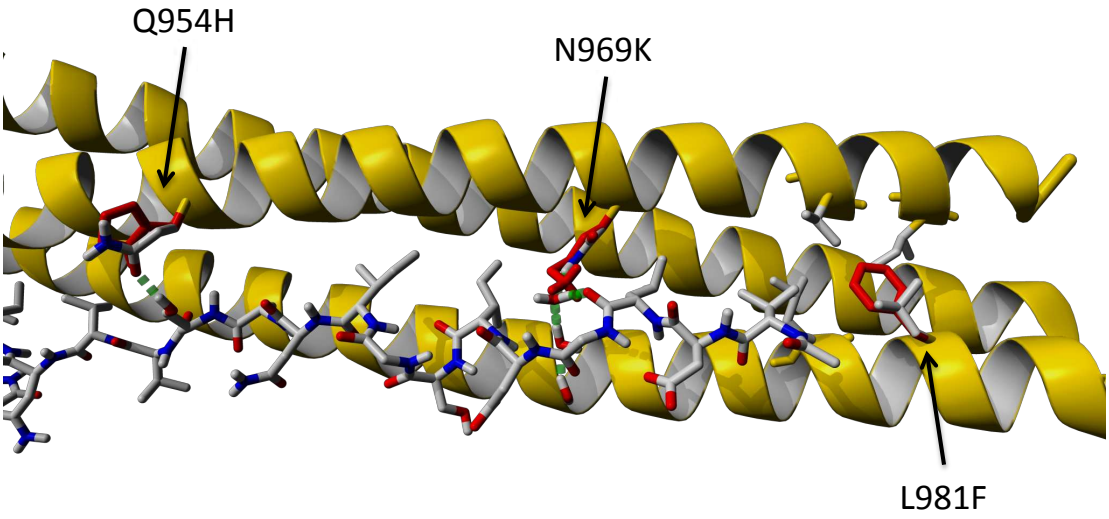


Fig. S5: HR1 mutations in Omicron BA.1 variant. The side chains of amino acids in the L3B protein structurally equivalent to those mutated in Omicron HR1 are highlighted with sticks and indicated with arrows. Original side chains are coloured in CPK scheme and mutant side chains are coloured in red. Similar rotamers have been chosen for the mutant side chains compared to the original ones. Hydrogen bonds between the side chains and the peptide are represented with green dashed lines. Due to its larger size, the K969 side chain should exclude one of interfacial the water molecules bridging the protein-peptide interaction.

Table S1. Amino acid sequences of the reference HR1, HR2 (Uniprot P0DTC2) and the CoVS-HR1 variants.

	Amino acid sequences [b]
HR1	NVLYENQKLI ANQFNSAIGK IQDSLSSTAS ALGKLQDVVN QNAQALNTLV KQLSSNFGAI SSVLNDILSR LDKVE
HR2 (V39E) [a]	VDLGDISGIN ASVVNIQKEI DRLNEVAKNL NESLIDLQE
L3A [c]	**D**VLYENQKLI AN**E**FNSAIGK IQDSLSSTAS ALGKLQD**K**VN QNAQ**K**LNTLV KQLSSNFGAI SS**K**LNDILSR LDKGEPAKDL RSLIDNL**K**S**E** IAGFNSSLQK VLTNL**R**Q**E**NQ NV**E**D**E**LKGLA SATSSL**E**DQI KG**E**ASNFQN**R** ILKQNEYLVN SGSGNVLYEN QKLIANQFNS A**E**GKIQDSLS ST**K**SALGKLQ DVVNQN**K**QAL NTL**E**KQLSSN F**R**AISSVLND I**E**SRL**Q**KV**Q**
L3B	**D**VLYENQKLI AN**K**FNSAIGK IQDSLSSTAS **E**LGKLQD**E**VN QNAQ**D**LNTLV KQLSSNFG**R**I SS**E**LNDILSR LDKGEPAKDL RS**D**IDNL**E**S**K** IAGFNSSLQK VLTNLAQ**K**NQ NV**E**D**K**LKGL**E** S**R**TSSL**EK**QI KGIASNFQN**E** ILKQ**R**EYLVN KGSGNVLYEN QKLI**E**NQFNS AIGKIQDSLS ST**K**SALGKL**K** DVVNQN**K**QAL NTLVKQLSSN FGAISSVLND I**K**SRLDKVE
L3C	**D**VLYENQKLI AN**K**FNSAI**K**K IQDSLSSTAS **E**L**K**KLQD**E**VN QNAQ**D**LNTLV KQLSSNF**KR**I SS**E**LNDILSR LDKGEPAKDL RS**D**IDNL**E**S**K** IA**K**FNSSLQK VLTNLAQ**K**NQ NV**E**D**K**LK**T**L**E** S**R**TSSL**EK**QI K**K**IASNFQN**E** ILKQ**R**EYLVN KGSGNVLYEN QKLI**E**NQFNS AI**K**KIQDSLS ST**K**SAL**K**KL**K** DVVNQN**K**QAL NTLVKQLSSN F**S**AISSVLND I**K**SRLDKVE

[a] The V39E peptide contained a C-terminal SGGY tag and was N-acetylated and C-amydated.
[b] The substituted amino acids from the reference sequence are highlighted in bold.
[c] The three protein variants contained a N-terminal Methionine and a C-terminal Histidine-tag of sequence GGGGSHHHHHH.

Table S2. Thermodynamic parameters of binding of CoVS-HR1 proteins and HR2 V39E peptide measured by ITC.

Protein variant	n	K_b ($\cdot 10^7$ M^{-1})	K_d (nM)	ΔH_b (kJ mol^{-1})	$T \cdot \Delta S_b$ (kJ mol^{-1})	ΔG_b (kJ mol^{-1})
L3A	1.09	3.0 ± 0.5[a]	34 ± 5	−52.0 ± 0.7	−9.4 ± 1.1	−42.3 ± 0.4
L3B	1.00	3.2 ± 0.5	32 ± 5	−30.0 ± 0.3	+12.8 ± 0.7	−42.8 ± 0.4
L3C	1.18	4.7 ± 1.4	21 ± 6	−14.0 ± 0.3	+29.8 ± 1.0	−43.8 ± 0.7

[a]Errors in the parameters are estimated as 95% confidence intervals in the parameters obtained from the fittings using a binding model of n independent and equivalent sites.

Table S3. Data collection and refinement statistics

	PDB entry 7ZR2
Data collection	
Space group	C 1 2 1
Cell dimensions	
a, b, c (Å)	166.78 37.37 43.91
a, b, g (°)	90 95.67 90
Resolution (Å)	18.92 - 1.45 (1.48 - 1.45)
Rmerge	0.059 (0.720)
I / sI	9.4 (1.2)
Completeness (%)	96.3 (75.4)
Redundancy	3.0 (2.4)
Refinement	
Resolution (Å)	18.92- 1.45
No. reflections	46198 (3824)
Rwork / Rfree	0.178 (0.281)/ 0.213 (0.315)
No. atoms	
Protein	2023
Ligand/ion	3
Water	304
B-factors	
Protein	22.45
Ligand/ion	26.48
Water	31.40
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	0.70

***Values in parentheses are for highest-resolution shell.**

References

1. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723 (1997).
2. E. Krieger, G. Vriend, YASARA View - molecular graphics for all devices - from smartphones to workstations. *Bioinformatics* **30**, 2981-2982 (2014).
3. S. Xia *et al.*, Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res.* **30**, 343-355 (2020).
4. Y. Liu, B. Kuhlman, RosettaDesign server for protein design. *Nucleic Acids Res.* **34**, W235-W238 (2006).
5. E. Gasteiger *et al.*, ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**, 3784-3788 (2003).
6. J. Juanhuix *et al.*, Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrot. Radiat.* **21**, 679-689 (2014).
7. C. Vonrhein *et al.*, Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **67**, 293-302 (2011).
8. P. R. Evans, An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **67**, 282-292 (2011).
9. N. Collaborative Computational Project, The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **50**, 760-763 (1994).
10. P. D. Adams *et al.*, PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **66**, 213-221 (2010).
11. G. Bunkoczi *et al.*, Phaser.MRage: automated molecular replacement. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **69**, 2276-2286 (2013).
12. P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **60**, 2126-2132 (2004).
13. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **66**, 486-501 (2010).
14. P. V. Afonine *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **68**, 352-367 (2012).
15. I. de Vries *et al.*, New restraints and validation approaches for nucleic acid structures in PDB-REDO. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **77**, 1127-1141 (2021).

4. Discussion and Future Work.

A chip off the old block; HR1 mimetic proteins as molecular templates.

Gp41 and S2 have remained elusive to in-depth studies due to their intrinsic properties that make them difficult to work with because of their low stability and tendency to aggregate. On the other hand, the proteins described in this thesis constitute faithful representations of a fully exposed HR1 trimer of HIV gp41 and SARS-CoV-2 S2. This is supported by the fact that they exhibit the same biophysical properties, secondary structure and oligomerization state, they can thightly bind to their peptide counterparts replacing the actual HR1 region inhibiting the fusion process and are able to neutralize viral fusion. The crystallographic structures of the complexes with the respective HR2 peptides (PDB IDs: 6R2G and 7ZR2) also show astonishing similarity with the post-fusion structures. In depth analysis of both structures reveals common characteristics between them: the groove is mainly hydrophobic and HR2 peptides interact with it through apolar amino acid side chains. The ridges flanking the groove present polar or charged amino acids alternating between positive- and negatively charge amino acids, with which the HR2 peptides establish hydrogen bonds or electrostatic interactions. Some interactions between HR1 and HR2 regions are mediated by interfacial water molecules and may account for specificity in the HR1-HR2 interaction and could also play a role as an escape mechanism to hide relevant epitopes to the host's immune system or as a mechanism to acquire resistance against inhibitor treatments.

The highly structural similarities of covNHR and covS-HR1 with the respective trimeric HR1 regions have several important implications, in fact, in this thesis we have employed them in a threefold way. First of all, covNHR and covS-HR1 proteins can bind HR2 peptides preventing the formation of a productive 6-HB. This aspect of the molecules has allowed us to exploit them as fusion inhibitors. They display a potent and broad inhibitory activity against both viruses and different strains. A second chief important aspect is that they constitute suitable models for the study the binding determinants of the HR1-HR2 interactions. In this regard, they could also be used as molecular templates in the search of small-molecule binders. HR1 region in SARS-CoV-2 does not present a prominent deep cavity like the HP in HIV but it does display potential hot-spots throughout the HR1 groove such as the fusion-core. Third, just as important, their structural similitudes with epitopes of immunogenic importance allowed us for the possibility of detecting antibodies directed against conformational epitopes. Therefore, the proteins may be used for the generation of polyclonal antibodies, nanobodies or human monoclonal antibodies with potentially broad neutralizing activity. These approaches constitute the major goals of effective vaccines or therapeutic to control the pandemics.

Conformational stability and activity: A structural tit-for-tat relationship.

In order to develop an effective protein-based therapeutic, several requirements have to be met. One of them is conformational stability because it not only affects biological activity but also many other variables important in pharmacological production. A high intrinsic conformational stability can increment production yield, facilitate storage conditions, extend medicine's shelf lives, maintain structural uniqueness, improve *in vivo* half-life, alleviate pharmaceutical formulation and enhance pharmacokinetic and pharmacodynamic properties, inter alia.

Over the course of previous work and this thesis we have carried out several approximations in an effort to stabilize the covNHR and covSHR1 proteins. The first approach implied the mutation of residues at the edge of the anti-parallel helices in positions e and g of the coiled-coil's heptad-repeat to introduce polar and charged amino acids, mainly glutamic acid, lysine and arginine, so that they could form salt-bridges between them inter-connecting the helices, while not interfering with the hydrophobic groove displayed by the parallel helices. This conferred covSHR1 molecules enough conformational stability and solubility to be expressed recombinantly in *E. coli* and stabilize their antiparallel three-helix bundle fold, while reducing exposed hydrophobicity and self-association. Differences in the sites mutated and the mutations introduced can led to differences in T_m values of ~10ºC (as seen between L3A and L3B with T_m values of ~83ºC and ~92ºC respectively). Another approach to stabilize these molecules was to introduce Gly-X substitutions (being X mainly Lysine) in the middle of the α-helices to increase α-helix propensity. This led to a further increase of ~10ºC (if we compare L3B and L3C proteins with T_m values of ~92ºC and ~103ºC respectively).

Polar-to-hydrophobic mutations of core-buried amino acids of covNHR lead to increases in T_m of over 20ºC in the most extreme cases, where T_m raised over 130ºC for Q13I and Q23I mutants, compared with a T_m value of 107.9ºC of covNHR. This approach, however, had the drawback of increasing self-association of the proteins, although this did not compromise their inhibitory activity. The other simple but still effective approximation used to increase conformational stability was the addition of intra-chain disulfide bonds. This approach improved almost systematically the T_m values of the constructs by approximately 20ºC per disulfide bond.

All these strategies successfully achieved their objective of stabilizing the proteins and further stabilization strategies may lay in a combination of them. New constructs combining two or more of these approaches may yield hyper-thermostable proteins and may be subject matter in future investigations.

With regard to the influence of conformational stability upon inhibitory activity two different observations arose from publications 2, 3 and 4. In publication 2, the effects in stability did not translate into significant increases in inhibitory activity. There may be two main reasons behind this: on one side, full-length mimetics of the NHR region may be sterically hindered due to their relatively large molecular size. On the other hand, the lack of accessibility of the HP-binding domain within the structure of the native envelope may difficult the accessibility of fusion inhibitors that target this region. Thereby, a

kinetic limit might be limiting the final inhibitory activity of these molecules. This has also been described for other fusion inhibitors that are kinetically restricted because their target is only transiently exposed during fusion. They are limited by the rate of association of the inhibitor to its target. The most probable scenario must be a combination of those two reasons.

In publication 3 and 4, however, we did identify a clear correlation between conformational stability and inhibitory potency. This can be justified by several reasons, possibly complementary between each other. Due to its smaller size, covNHR-N-dSS may be less sterically hindered than full-length covNHR constructs to access its target in gp41 in the context of the virus-cell assembly. It is also possible that the respective targets of covNHR-N-dSS and L3C in HR2 are natively more exposed or become accessible during a longer time interval during fusion. Another possibility is that, due to their lower binding affinity to their targets and their less potent activity they have not yet reached the above described kinetic limit for inhibition potency.

In the investigations concerning this thesis we have also noticed that the stabilization of the proteins does not modify, or only slightly modifies, the affinity to their respective peptide targets. They exhibit similar binding Gibbs energies to the less-stabilized parent molecules; therefore, there must be a kinetic reason that explains the increase in activity. If the activity is kinetically restricted because the target of the molecules is only transiently exposed, we should try to design inhibitor that binds as fast as possible to their target. This means that the activation energy barrier of the binding between HR1 inhibitors to the transiently-exposed HR2 region should be reduced as much as possible. CovNHR-N-SS, covNHR-N-dSS and L3C proteins may have achieved that by reducing their conformational flexibility while simultaneously stabilizing a predisposed conformation prone to HR2 binding. This predisposed conformation must have reduced the binding entropic barrier associated to the entropic cost acquiring the binding-competent conformation.

Long-range allosteric communication between sub-domains.

An interesting finding in publication 2 is that mutations at core positions of the coiled-coil not only affect the overall stability of the NHR trimer but they also influence the interactions established with the counterpart CHR region. This can be deduced from the effect of the mutations Q13I introduced at the core lining with NTP region. They destabilize the interaction between Y24L peptide and the NTP because they remove an interaction mediated by an interfacial water molecule. These mutations also alter the NHR core stability and those alterations are transmitted to other pockets, the MP and HP, increasing the binding affinity to CHR peptides. This scenario is also evidenced in a previous publication from our group[144], where two mutations introduced at the NTP that replaced the cavity left by this interfacial water molecule led to a decrease in binding affinity at the NTP but a compensation improvement of binding affinity at the MP and HP.

This long-range pocket communication is also supported by the findings in publication 1 where we found that ANS, an amphipathic probe specifically targeting the HP, gets its

binding mode influenced by binding of the Y24L peptide, which interacts with the NTP and MP. The results from this publication also support that the binding to one pocket affect the structural tightening contribution of binding to other pockets. Correlation between structural tightening and binding to different pockets was also found in publication 2.

The combined results presented in this thesis allow us to understand better the connections established between sub-domains in the NHR trimer. The alteration of one site, whether it is caused by a mutation or by the interaction with a binder molecule (a peptide or a small-molecule binder), may be transmitted to other distant sites throughout the structure of the coiled-coil. This could mean that a bi-directional communication may govern the overall arrangement of the NHR-CHR interaction. More specifically, a perturbation in the core structure of the NHR can be transmitted to the CHR binding interface. It may also happen the other way: the adaptation made upon a CHR peptide binding may be transmitted to the NHR core and other pockets and/or sub-domains. This hypothesis may explain a resistance mechanism used by viruses to escape from inhibitors or other host's mediated responses. If an inhibitor targets a relevant region in gp41 the virus may escape from it by mutating it, reducing the affinity at this site but improving the affinity at other sites still maintaining significant infectivity.

5. Future Perspectives:

Small-molecule inhibitors directed against gp41. Does size matter?

Protein-based therapeutic strategies are currently highly successful in terms of their clinical utility [149150151] but we cannot ignore the potential that small-molecules still can offer in the treatment of certain diseases. Small-molecule therapeutics typically exhibit simpler structures that normally lead to more predictable pharmacokinetics and pharmacodynamics; their development often involves simpler manufacturing; many of them are highly stable and orally bioavailable and relatively more affordable[152]. Those benefits, despite the drawbacks described earlier in the Introduction, still make worth the investment in finding small-molecules capable of inhibition HIV-1 fusion. In order to fully exploit the potential of our covNHR molecules as viral target templates we propose to perform structure similarity virtual screening of compound libraries to identify drug-like compounds with analogous structures to that studied in publication 1. A plausible approximation may be search for molecules mimicking the structure of the ANS dimer or even larger in size to fully occupy the hydrophobic pocket. Even larger molecules may also be interesting as interactions with other pockets might increase the binding affinity and their capacity to disturb the formation of the 6-HB. Another interesting approach can be the use of fragment-based drug discovery techniques[153] that may allow us to find structural minimal units that bind to the same or different pockets and then tethering them together to obtain drug-like lead compounds. The knowledge drawn from publication 1 concerning important interactions involved in ANS binding may guide us to try to co-crystalize this or other structurally similar molecule with covNHR or other protein variants. This result would set the accomplishment of a milestone, since no other small-molecule has been successfully co-crystallized with any gp41 mimetic to date.

Exploring new regions to target: Not all those who wander are lost.

We have discussed the important direct correlation existent between conformational stability and activity for non-kinetically restricted inhibitors. Those inhibitors must gather at least two principles: their target region must be accessible, both spatially and temporally; and they should be small enough so that their molecular size doesn't hinder the accessibility to the otherwise exposed target. We have achieved those requirements with the results showed in publication 3. Due to the accessibility limitations of the upstream CHR region, involved in the tryptophan collar clasp, it seems more reasonable to target, the so-far unexplored downstream MPER region. This region has already been targeted by antibodies (such as 2F5, 4E10 or 10E8) meaning that it fulfills requisite of being accessible. So, the immediate next step would be designing miniproteins covering the N-terminal sub-domain of NHR that also incorporate FPPR and FP mimicking regions (complementary to the MPER region). This strategy could yield a protein inhibitor with an even-higher binding affinity and whose inhibitory activity may overcome the currently existing inhibition limit in the low nM range with the objective of achieving sub-nM activities.

HR1-based COVID-19 fusion inhibitors: Taking the plunge.

CovSHR1 proteins may be subjected to the same strategies described earlier for covNHR molecules to improve their stability. As a result, a combination of polar-to-hydrophobic core mutations with the Introduction of cysteine bridges seems like a feasible strategy to continue stabilizing these molecules, especially L3C, the most promising one.

To study in more detail the binding determinants of the hydrophobic groove in HR1 a future direction may involve the design of smaller molecules encompassing sub-domains of the HR1 region.

Due to the conservation level of the HR1 region and that the molecules developed in this thesis constitute accurate representations of it we are planning to use them to search different molecular binders that could interfere with the fusion process. The search of small-molecule binders seems more difficult than in HIV as there is no prominent cavity in the HR1 surface and there is currently not a detailed characterization of the interactions established throughout the groove. However, a more interesting approach arises by using these molecules to elicit immune responses, mainly neutralizing responses, where nanobodies or antibodies could be generated for therapeutic purposes.

6. Conclusions.

The most relevant conclusions drawn from the studies presented as part of this thesis can be summarized as follows:

1. ANS binds specifically to the hydrophobic pocket in covNHR in two different binding modes.
2. The ANS binding mode is allosterically modified by the occupancy of other distant NHR pockets by a CHR peptide, demonstrating that the HP shows a marked conformational flexibility and that communication exists between different regions within the NHR crevice.
3. In its stabilized form, the HP can hold two ANS molecules that bind cooperatively, mimicking functionally relevant interactions at the HP.
4. ANS-based derivative molecules may constitute potential drug candidates, providing a valuable class of new lead structures for the design of potent HIV fusion inhibitors.
5. CovNHR constructs can serve as accurate molecular templates in the search of small-molecule ligands mimicking the gp41 NHR pre-hairpin intermediate structure.
6. Replacing buried polar Thr or Gln side chains for Ile in three core positions of the covNHR protein leads to very strong increases in thermal stability but favors self-association of the proteins into dimeric species.
7. The observed effects of mutations at the core of covNHR support a role of buried polar interactions inside the NHR coiled-coil to contribute to structural specificity and uniqueness at a cost of stability in the HIV-1 gp41 NHR region.
8. The effects of mutations at the core of covNHR are not restricted locally but extend to other regions influencing their conformational properties and thus affecting the binding interactions with CHR peptides. This also corroborates a coupling that exists between CHR binding and structural tightening throughout the covNHR's structure.
9. Disulfide bonds introduced in covNHR molecule's structure produce strong stabilization effects that result in tighter interactions with CHR due to reduced conformational dynamics at the peptide-protein complex interface.
10. Potent and broad HIV inhibitory activity in the tens of nanomolar range can be achieved with a small but stable covNHR construct targeting the C-terminal half of CHR but not targeting the preserved HP-binding motif.
11. There is significant correlation between conformational stability and HIV inhibitory potency.

12. A design approach similar to that used in covNHR proteins has been successfully applied to build the CoVS-HR1 constructs, which constitute faithful representations of the HR1 trimer in pre-hairpin intermediate conformation of SARS-CoV-2 S2.

13. The crystallographic structure of the complex between a CoV-S-HR1 protein and a HR2 peptide shows interactions virtually identical to those of the S2 post-fusion structure.

14. The CoVS-HR1 proteins can be used to study in detail the structural and thermodynamic determinants of binding to their HR2 peptide counterparts, helping the overall understanding of the HR1-HR2-mediated fusion process.

15. CovS-HR1 proteins are broad and effective inhibitors of virus-fusion infection *in vitro*.

16. Sera from COVID-19 convalescent patients have antibodies that bind to the CoVS-HR1 proteins emphasizing their capacity to mimic the SARS-CoV-2 HR1 region and implying that the HR1 region is immunogenic at some point during the course of the infection.

17. The antibody response against the CoVS-HR1 proteins correlates positively with their stability but not with their HR1 sequence preservation, suggesting that the proteins imitate conformational HR1 epitopes of potential immunological relevance.

Las conclusiones más relevantes extraídas de los estudios presentados como parte de esta tesis se pueden resumir de la siguiente manera:

1. El ANS se une específicamente al bolsillo hidrofóbico en covNHR con dos modos de unión diferentes.
2. El modo de unión del ANS se modifica alostéricamente mediante la ocupación de otros bolsillos distantes de NHR por un péptido CHR, lo que demuestra que el HP muestra una marcada flexibilidad conformacional y que existe comunicación entre diferentes regiones dentro del surco NHR.
3. En su forma estabilizada, el HP puede contener dos moléculas de ANS que se unen cooperativamente, imitando interacciones funcionalmente relevantes en el HP.
4. Las moléculas derivadas del ANS y basadas en él pueden constituir posibles candidatos a fármacos, proporcionando una valiosa clase de nuevas estructuras líder para el diseño de potentes inhibidores de la fusión del VIH.
5. Los constructos covNHR pueden servir como plantillas moleculares precisas en la búsqueda de ligandos de molécula pequeña que imiten la estructura intermedia pre-horquilla del NHR de gp41.
6. Reemplazar cadenas laterales polares de Thr o Gln enterradas por Ile en tres posiciones en el núcleo de la proteína covNHR conduce a aumentos muy fuertes en la estabilidad térmica pero favorece la autoasociación de las proteínas en especies diméricas.
7. Los efectos observados de las mutaciones en el núcleo de covNHR respaldan el papel de las interacciones polares enterradas dentro de la hélice superenrollada de NHR para contribuir a la especificidad estructural y a la singularidad a costa de la estabilidad en la región NHR de HIV-1 gp41.
8. Los efectos de las mutaciones en el núcleo de covNHR no están restringidos localmente, sino que se extienden a otras regiones influyendo en sus propiedades conformacionales y, por lo tanto, afectando las interacciones de unión con los péptidos CHR. Esto también corrobora un acoplamiento existente entre la unión de CHR y el estrechamiento estructural en toda la estructura de covNHR.
9. Los enlaces disulfuro introducidos en la estructura de la molécula covNHR producen fuertes efectos de estabilización que resultan en interacciones más estrechas con CHR debido a una dinámica conformacional reducida en la interfaz del complejo péptido-proteína.
10. Se puede lograr una potente y amplia actividad inhibidora del VIH en el rango de decenas de nanomolar con un constructo covNHR pequeño pero estable dirigido a la mitad C-terminal de CHR y sin dirigirse frente al conservado motivo de unión a HP.
11. Existe una correlación significativa entre la estabilidad conformacional y la potencia inhibidora del VIH.

12. Se ha aplicado con éxito un enfoque de diseño similar al utilizado en las proteínas covNHR para construir los constructos CoVS-HR1, que constituyen representaciones fieles del trímero HR1 en la conformación intermedia previa a la horquilla de la proteína S2 del SARS-CoV-2.

13. La estructura cristalográfica del complejo entre una proteína CoV-S-HR1 y un péptido HR2 muestra interacciones prácticamente idénticas a las de la estructura post-fusión de S2.

14. Las proteínas CoVS-HR1 se pueden usar para estudiar en detalle los determinantes estructurales y termodinámicos de la unión a sus contrapartes peptídicas HR2, lo que ayuda a la comprensión general del proceso de fusión mediado por HR1-HR2.

15. Las proteínas CovS-HR1 son inhibidores amplios espectro y eficaces de la infección mediada por la fusión viral *in vitro*.

16. Los sueros de pacientes convalecientes de COVID-19 tienen anticuerpos que se unen a las proteínas CoVS-HR1, lo que enfatiza su capacidad para imitar la región HR1 del SARS-CoV-2 e implica que la región HR1 es inmunogénica en algún momento durante el curso de la infección.

17. La respuesta de anticuerpos contra las proteínas CoVS-HR1 se correlaciona positivamente con su estabilidad, pero no con la preservación de su secuencia HR1, lo que sugiere que las proteínas imitan epítopos conformacionales de HR1 de potencial relevancia inmunológica.

Personal Contribution.

In this section I will highlight the individual contributions I carried out within the publications used in this thesis:

Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors[154]**.**

- Protein expression and purification.
- Fluorescence assays.
- Circular dichroism experiments.
- Isothermal Titration Calorimetry studies.
- Dynamic Light Scattering analyses.
- Docking calculations.
- Molecular Dynamics studies.
- Writing - original draft.

Extremely Thermostabilizing Core Mutations in Coiled-Coil Mimetic Proteins of HIV-1 gp41 Produce Diverse Effects on Target Binding but Do Not Affect Their Inhibitory Activity[155]**.**

- Protein expression and purification.
- Thermal Stability Studies by Differential Scanning Calorimetry (but not their analyses).
- Isothermal Titration Calorimetry studies.
- Dynamic Light Scattering analyses

Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion[156]**.**

- Protein expression and purification.
- Molecular Dynamics studies.
- Thermal Stability Studies by Circular Dichroism.
- Dynamic Light Scattering analyses.
- Isothermal Titration Calorimetry studies.
- *In vitro* binding assays to the trimeric Envelopes.
- Writing - original draft.

Novel proteins mimicking SARS-CoV-2 epitopes with broad inhibitory activity.

- Protein expression and purification. Only L3B and L3C proteins.
- Circular Dichroism and Dynamic Light Scattering characterization of L3B protein.
- Isothermal Titration Calorimetry studies with V39E peptide.
- Dynamic Light Scattering experiments with L3B protein.
- In vitro binding assays to the trimeric Spike.
- Reactivity assays of CoVS-HR1 proteins against COVID-19 convalescent patient's sera.

References

1. Hippocrates & Jones, W. H. S. *Hippocrates: With an English Translation by WHS Jones*. (Harvard University Press, 1923).
2. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science (80-.).* **366**, (2019). DOI: 10.1126/science.aax3100.
3. Gibb, R., Franklinos, L. H. V., Redding, D. W. & Jones, K. E. Ecosystem perspectives are needed to manage zoonotic risks in a changing climate. *BMJ* **371**, 1–7 (2020). DOI: 10.1136/bmj.m3389.
4. Rupasinghe, R., Chomel, B. B. & Martínez-López, B. Climate change and zoonoses: A review of the current status, knowledge gaps, and future trends. *Acta Trop.* **226**, 106225 (2022). DOI: 10.1016/j.actatropica.2021.106225.
5. Romanello, M. *et al.* The 2021 report of the Lancet Countdown on health and climate change: code red for a healthy future. *Lancet* **398**, 1619–1662 (2021). DOI: 10.1016/S0140-6736(21)01787-6.
6. Mills, J. N., Gage, K. L. & Khan, A. S. Potential influence of climate change on vector-borne and zoonotic diseases: A review and proposed research plan. *Environ. Health Perspect.* **118**, 1507–1514 (2010). DOI: 10.1289/ehp.0901389.
7. Naicker, P. R. The impact of climate change and other factors on zoonotic diseases. *Arch. Clin. Microbiol.* **2**, 2–7 (2011). DOI: 10.3823/226.
8. Barré-Sinoussi, F. *et al.* Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science (80-.).* **220**, 868–871 (1983). DOI: 10.1126/science.6189183.
9. Gallo, R. C. *et al.* Isolation of Human T-Cell Leukemia Virus in Acquired Immune Deficiency Syndrome (AIDS). *Science (80-.).* **220**, 865–867 (1983). DOI: 10.1126/science.6601823.
10. Ksiazek, T. G. *et al.* A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **348**, 1953–1966 (2003). DOI: 10.1056/NEJMoa030781.
11. Drosten, C. *et al.* Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. *N. Engl. J. Med.* **348**, 1967–1976 (2003). DOI: 10.1056/NEJMoa030747.
12. Subbarao, K. *et al.* Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science (80-.).* **279**, 393–396 (1998). DOI: 10.1126/science.279.5349.393.
13. Lai, S. *et al.* Global epidemiology of avian influenza A H5N1 virus infection in humans, 1997-2015: A systematic review of individual case data. *Lancet Infect. Dis.* **16**, e108–e118 (2016). DOI: 10.1016/S1473-3099(16)00153-5.
14. Emond, R. T., Evans, B., Bowen, E. T. & Lloyd, G. A case of Ebola virus infection. *BMJ* **2**, 541–544 (1977). DOI: 10.1136/bmj.2.6086.541.
15. Baize, S. *et al.* Emergence of Zaire Ebola Virus Disease in Guinea. *N. Engl. J. Med.* **371**, 1418–1425 (2014). DOI: 10.1056/nejmoa1404505.

16. Duffy, M. R. *et al.* Zika Virus Outbreak on Yap Island, Federated States of Micronesia. *N. Engl. J. Med.* **360**, 2536–2543 (2009). DOI: 10.1056/nejmoa0805715.

17. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020). DOI: 10.1016/S0140-6736(20)30183-5.

18. Roberts, D. L., Rossman, J. S. & Jarić, I. Dating first cases of COVID-19. *PLoS Pathog.* **17**, 1–10 (2021). DOI: 10.1371/journal.ppat.1009620.

19. Centers for Disease Control and Prevention. Pneumocystis Pneumonia - Los Angeles. *Morbidity and Mortality Weekly Report.* 30(21);1-3 https://www.cdc.gov/mmwr/preview/mmwrhtml/june_5.htm (1981).

20. Broder, S. & Gallo, R. C. A Pathogenic Retrovirus (HTLV-III) Linked to AIDS. *N. Engl. J. Med.* **311**, 1292–1297 (1984). DOI: 10.1056/NEJM198411153112006.

21. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**, 9–17 (1985). DOI: 10.1016/0092-8674(85)90303-4.

22. Allan, J. S. *et al.* Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science (80-.).* **228**, 1091–1094 (1985). DOI: 10.1126/science.2986290.

23. Clavel, F. *et al.* Isolation of a New Human Retrovirus African Patients with AIDS. *Science (80-.).* **233**, 343–346 (1986).

24. Montagnier, L. Historical essay: A history of HIV discovery. *Science (80-.).* **298**, 1727–1728 (2002). DOI: 10.1126/science.1079027.

25. Simpson, S., Kaufmann, M. C., Glozman, V. & Chakrabarti, A. Disease X: accelerating the development of medical countermeasures for the next pandemic. *Lancet Infect. Dis.* **20**, e108–e115 (2020). DOI: 10.1016/S1473-3099(20)30123-7.

26. Kuthya, S., Anthony, C. L., Fashina, T., Yeh, S. & Shantha, J. G. World health organization high priority pathogens: Ophthalmic disease findings and vision health perspectives. *Pathogens* **10**, 1–11 (2021). DOI: 10.3390/pathogens10040442.

27. Jiang, S. & Shi, Z. L. The First Disease X is Caused by a Highly Transmissible Acute Respiratory Syndrome Coronavirus. *Virol. Sin.* **35**, 263–265 (2020). DOI: 10.1007/s12250-020-00206-5.

28. Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, D. B. and M. R. Coronavirus Pandemic (COVID-19). *Published online at OurWorldInData.org* https://ourworldindata.org/coronavirus (2020).

29. Korber, B. *et al.* Timing the Ancestor of the HIV-1 Pandemic Strains. *Science (80-.).* **288**, 1789–1796 (2000). DOI: 10.1126/science.288.5472.1789.

30. UNAIDS. Global AIDS Update 2021. *World Heal. Organ.* (2021).

31. Deeks, S. G., Overbaugh, J., Phillips, A. & Buchbinder, S. HIV infection. *Nat. Rev. Dis. Prim.* **1**, 15035 (2015). DOI: 10.1038/nrdp.2015.35.

32. McLaren, P. J. & Fellay, J. HIV-1 and human genetic variation. *Nat. Rev. Genet.* **22**, 645–657 (2021). DOI: 10.1038/s41576-021-00378-0.

33. Li, G., Jiang, Y. & Zhang, L. HIV upsurge in China's students. *Science* **364**, 711 (2019). DOI: 10.1126/science.aay0799.

34. Steichen, J. M. *et al.* A generalized HIV vaccine design strategy for priming of broadly neutralizing antibody responses. *Science (80-.).* **366**, 1–13 (2019). DOI: 10.1126/science.aax4380.

35. Ng'uni, T., Chasara, C. & Ndhlovu, Z. M. Major Scientific Hurdles in HIV Vaccine Development: Historical Perspective and Future Directions. *Front. Immunol.* **11**, 1–17 (2020). DOI: 10.3389/fimmu.2020.590780.

36. Esté, J. A. & Telenti, A. HIV entry inhibitors. *Lancet* **370**, 81–88 (2007). DOI: 10.1016/S0140-6736(07)61052-6.

37. Gochin, M. & Zhou, G. Amphipathic Properties of HIV-1 gp41 Fusion Inhibitors. *Curr. Top. Med. Chem.* **11**, 3022–3032 (2011). DOI: 10.2174/156802611798808488.

38. Levy, J. A. Pathogenesis of human immunodeficiency virus infection. *Microbiol. Rev.* **57**, 183–289 (1993). DOI: 10.1128/mr.57.1.183-289.1993.

39. Shang, J. *et al.* Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci.* **117**, 11727–11734 (2020). DOI: 10.1073/pnas.2003138117.

40. Nicola, M. *et al.* The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *Int. J. Surg.* **78**, 185–193 (2020). DOI: 10.1016/j.ijsu.2020.04.018.

41. Max Roser and Hannah Ritchie. HIV / AIDS. *Published online at OurWorldInData.org* https://ourworldindata.org/hiv-aids (2018).

42. Illanes-álvarez, F., Márquez-Ruiz, D., Márquez-Coello, M., Cuesta-Sancho, S. & Girón-González, J. A. Similarities and differences between hiv and sars-cov-2. *Int. J. Med. Sci.* **18**, 846–851 (2021). DOI: 10.7150/ijms.50133.

43. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020). DOI: 10.1016/ S1473-3099(20)30120-1.

44. Teixeira, C., Gomes, J. R. B., Gomes, P. & Maurel, F. Viral surface glycoproteins, gp120 and gp41, as potential drug targets against HIV-1: Brief overview one quarter of a century past the approval of zidovudine, the first anti-retroviral drug. *Eur. J. Med. Chem.* **46**, 979–992 (2011). DOI: 10.1016/j.ejmech.2011.01.046.

45. Blumenthal, R., Durell, S. & Viard, M. HIV entry and envelope glycoprotein-mediated fusion. *J. Biol. Chem.* **287**, 40841–40849 (2012). DOI: 10.1074/jbc.R112.406272.

46. Chen, J. *et al.* Visualizing the translation and packaging of HIV-1 full-length RNA. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 6145–6155 (2020). DOI: 10.1073/pnas.1917590117.

47. Krebs FC, Hogan TH, Quiterio S, Gartner S, W. B. *Lentiviral LTR-directed Expression, Sequence Variation, and Disease Pathogenesis*. (HIV Sequence Compendium 2001. Los Alamos, NM, LA-UR 02–2877.: Theoretical Biology

and Biophysics Group, Los Alamos National Laboratory, 2001).

48. Sauter, D. *et al.* Human Tetherin Exerts Strong Selection Pressure on the HIV-1 Group N Vpu Protein. *PLoS Pathog.* **8**, e1003093 (2012). DOI: 10.1371/journal.ppat.1003093.

49. Seitz, R. Human Immunodeficiency Virus (HIV). *Transfus. Med. Hemotherapy* **43**, 203–222 (2016). DOI: 10.1159/000445852.

50. Vicenzi, E. & Poli, G. Novel factors interfering with human immunodeficiency virus-type 1 replication in vivo and in vitro. *Tissue Antigens* **81**, 61–71 (2013). DOI: 10.1111/tan.12047.

51. Yager & Konan. Sphingolipids as Potential Therapeutic Targets against Enveloped Human RNA Viruses. *Viruses* **11**, 912 (2019). DOI: 10.3390/v11100912.

52. Khailany, R. A., Safdar, M. & Ozaslan, M. Genomic characterization of a novel SARS-CoV-2. *Gene Reports* **19**, 100682 (2020). DOI: 10.1016/j.genrep.2020.100682.

53. Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M. & Bergeron, J. R. C. Structural Characterization of SARS-CoV-2: Where We Are, and Where We Need to Be. *Front. Mol. Biosci.* **7**, (2020). DOI: 10.3389/fmolb.2020.605236.

54. Mousavizadeh, L. & Ghasemi, S. Genotype and phenotype of COVID-19 : Their roles in pathogenesis. *J. Microbiol. Immunol. Infect.* **54**, 159–163 (2021). DOI: 10.1016/j.jmii.2020.03.022.

55. Huthoff, H. & Towers, G. J. Restriction of retroviral replication by APOBEC3G/F and TRIM5α. *Trends Microbiol.* **16**, 612–619 (2008). DOI: 10.1016/j.tim.2008.08.013.

56. Evans, D. T., Serra-Moreno, R., Singh, R. K. & Guatelli, J. C. BST-2/tetherin: a new component of the innate immune response to enveloped viruses. *Trends Microbiol.* **18**, 388–396 (2010). DOI: 10.1016/j.tim.2010.06.010.

57. Engelman, A. & Cherepanov, P. The structural biology of HIV - 1 : mechanistic and therapeutic insights. *Nat. Publ. Gr.* **10**, 279–290 (2012). DOI: 10.1038/nrmicro2747.

58. Gheysen, D. *et al.* Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. *Cell* **59**, 103–112 (1989). DOI: 10.1016/0092-8674(89)90873-8.

59. Sundquist, W. I. & Krausslich, H.-G. HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* **2**, a006924–a006924 (2012). DOI: 10.1101/cshperspect.a006924.

60. Pettit, S. C. *et al.* The p2 domain of human immunodeficiency virus type 1 Gag regulates sequential proteolytic processing and is required to produce fully infectious virions. *J. Virol.* **68**, 8017–8027 (1994). DOI: 10.1128/jvi.68.12.8017-8027.1994.

61. Freed, E. O. HIV-1 assembly, release and maturation. *Nat. Rev. Microbiol.* **13**, 484–496 (2015). DOI: 10.1038/nrmicro3490.

62. Parry, R., Gifford, R. J., Lytras, S., Ray, S. C. & Coin, L. J. M. No evidence of SARS-CoV-2 reverse transcription and integration as the origin of chimeric

transcripts in patient tissues. *Proc. Natl. Acad. Sci.* **118**, 1–2 (2021). DOI: 10.1073/pnas.2109066118.

63. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271–280.e8 (2020). DOI: 10.1016/j.cell.2020.02.052.

64. V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021). DOI: 10.1038/s41579-020-00468-6.

65. Pizzato, M. *et al.* SARS-CoV-2 and the Host Cell: A Tale of Interactions. *Front. Virol.* **1**, 1–29 (2022). DOI: 10.3389/fviro.2021.815388.

66. Trougakos, I. P. *et al.* Insights to SARS-CoV-2 life cycle, pathophysiology, and rationalized treatments that target COVID-19 clinical complications. *J. Biomed. Sci.* **28**, 1–18 (2021). DOI: 10.1186/s12929-020-00703-5.

67. Medina-O'donnell, M. *et al.* Oleanolic Acid Derivatives as Potential Inhibitors of HIV-1 Protease. *J. Nat. Prod.* **82**, (2019). DOI: 10.1021/acs.jnatprod.9b00649.

68. Wei, X. *et al.* Emergence of Resistant Human Immunodeficiency Virus Type 1 in Patients Receiving Fusion Inhibitor (T-20) Monotherapy. *Antimicrob. Agents Chemother.* **46**, 1896–1905 (2002). DOI: 10.1128/AAC.46.6.1896-1905.2002.

69. Nováková, L., Pavlík, J., Chrenková, L., Martinec, O. & Červený, L. Current antiviral drugs and their analysis in biological materials – Part II: Antivirals against hepatitis and HIV viruses. *J. Pharm. Biomed. Anal.* **147**, 378–399 (2018). DOI: 10.1016/j.jpba.2017.07.003.

70. Saag, M. S. *et al.* Antiretroviral Drugs for Treatment and Prevention of HIV Infection in Adults: 2020 Recommendations of the International Antiviral Society-USA Panel. *JAMA - J. Am. Med. Assoc.* **324**, 1651–1669 (2020). DOI: 10.1001/jama.2020.17025.

71. Menéndez-Arias, L. Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Res.* **98**, 93–120 (2013). DOI: 10.1016/j.antiviral.2013.01.007.

72. Yang, H. & Rao, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nat. Rev. Microbiol.* **19**, 685–700 (2021). DOI: 10.1038/s41579-021-00630-8.

73. Cully, M. A tale of two antiviral targets — and the COVID-19 drugs that bind them. *Nat. Rev. Drug Discov.* **21**, 3–5 (2022). DOI: 10.1038/d41573-021-00202-8.

74. FDA Approves First Treatment for COVID-19. *October 22* https://www.fda.gov/news-events/press-announcements/fda-approves-first-treatment-covid-19 (2020).

75. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Final Report. *N. Engl. J. Med.* **383**, 1813–1826 (2020). DOI: 10.1056/nejmoa2007764.

76. FDA Authorizes Pfizer-BioNTech COVID-19 Vaccine. *Med. Lett.* **63**, 1–2 (2021).

77. Oliver, S. E. The advisory committee on immunization practices' interim recommendation for use of moderna COVID-19 vaccine—United States,

December 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **69**, (2021).

78. Food and Drug Administration. Janssen COVID-19 vaccine emergency use authorization. *Silver Spring, MD: US Department of Health and Human Services, Food and Drug Administration;* https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/Janssen-covid-19-vaccine (2021).

79. Mouffak, S., Shubbar, Q., Saleh, E. & El-Awady, R. Recent advances in management of COVID-19: A review. *Biomed. Pharmacother.* **143**, 112107 (2021). DOI: 10.1016/j.biopha.2021.112107.

80. Food and Drug Administration. Emergency Use Authorization for Vaccines Explained. *11/20/2020* https://www.fda.gov/vaccines-blood-biologics/vaccines/emergency-use-authorization-vaccines-explained (2020).

81. Barouch, D. H. Challenges in the development of an HIV-1 vaccine. *Nature* **455**, 613–619 (2008). DOI: 10.1038/nature07352.

82. Yi, H. A., Fochtman, B. C., Rizzo, R. C. & Jacobs, A. Inhibition of HIV Entry by Targeting the Envelope Transmembrane Subunit gp41. *Curr. HIV Res.* **14**, 283–294 (2016). DOI: 10.1001/jamasurg.2014.1086.Feasibility.

83. Klasse, P. J. The molecular basis of HIV entry. *Cell. Microbiol.* **14**, 1183–1192 (2012). DOI: 10.1111/j.1462-5822.2012.01812.x.

84. Haqqani, A. A. & Tilton, J. C. Entry inhibitors and their use in the treatment of HIV-1 infection. *Antiviral Res.* **98**, 158–170 (2013). DOI: 10.1016/j.antiviral.2013.03.017.

85. Pancera, M. *et al.* Structure and immune recognition of trimeric pre-fusion HIV-1 Env. *Nature* **514**, 455–461 (2014). DOI: 10.1038/nature13808.

86. Wyatt, R. The HIV-1 Envelope Glycoproteins: Fusogens, Antigens, and Immunogens. *Science (80-.).* **280**, 1884–1888 (1998). DOI: 10.1126/science.280.5371.1884.

87. Tilton, J. C. & Doms, R. W. Entry inhibitors in the treatment of HIV-1 infection. *Antiviral Res.* **85**, 91–100 (2010). DOI: 10.1016/j.antiviral.2009.07.022.

88. Lu, K., Asyifah, M. R., Shao, F. & Zhang, D. Development of HIV-1 fusion inhibitors targeting gp41. *Curr. Med. Chem.* **21**, 1976–96 (2014). DOI: 10.2174/0929867321666131218094559.

89. Allen, W. J. & Rizzo, R. C. Computer-Aided Approaches for Targeting HIVgp41. *Biology (Basel).* **1**, 311–338 (2012). DOI: 10.3390/biology1020311.

90. Abdelrahman, Z., Li, M. & Wang, X. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Front. Immunol.* **11**, (2020). DOI: 10.3389/fimmu.2020.552909.

91. Du, L. *et al.* The spike protein of SARS-CoV - A target for vaccine and therapeutic development. *Nat. Rev. Microbiol.* **7**, 226–236 (2009). DOI: 10.1038/nrmicro2090.

92. Kirtipal, N., Bharadwaj, S. & Kang, S. G. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect. Genet. Evol.* **85**, 104502 (2020). DOI: 10.1016/j.meegid.2020.104502.

93. Huang, Y., Yang, C., Xu, X. feng, Xu, W. & Liu, S. wen. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020). DOI: 10.1038/s41401-020-0485-4.

94. Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science (80-.).* **369**, 1586–1592 (2020). DOI: 10.1126/science.abd4251.

95. Duan, L. *et al.* The SARS-CoV-2 Spike Glycoprotein Biosynthesis, Structure, Function, and Antigenicity: Implications for the Design of Spike-Based Vaccine Immunogens. *Front. Immunol.* **11**, 1–12 (2020). DOI: 10.3389/fimmu.2020.576622.

96. Woo, H. *et al.* Developing a fully glycosylated full-length SARS-COV-2 spike protein model in a viral membrane. *J. Phys. Chem. B* **124**, 7128–7137 (2020). DOI: 10.1021/acs.jpcb.0c04553.

97. Pan, J., Peng, H., Chen, B. & Harrison, S. C. Cryo-EM Structure of Full-length HIV-1 Env Bound With the Fab of Antibody PG16. *J. Mol. Biol.* **432**, 1158–1168 (2020). DOI: 10.1016/j.jmb.2019.11.028.

98. Cai, L. & Jiang, S. Development of peptide and small-molecule HIV-1 fusion inhibitors that target gp41. *ChemMedChem* **5**, 1813–1824 (2010). DOI: 10.1002/cmdc.201000289.

99. Caillat, C. *et al.* Structure of hiv-1 gp41 with its membrane anchors targeted by neutralizing antibodies. *Elife* **10**, 1–26 (2021). DOI: 10.7554/ELIFE.65005.

100. Wells, A. & Coyne, C. B. Inhibiting Ebola virus and SARS-CoV-2 entry. *Science (80-.).* **370**, 167–169 (2020).

101. Pollard, C. A., Morran, M. P. & Nestor-Kalinoski, A. L. The covid-19 pandemic: A global health crisis. *Physiol. Genomics* **52**, 549–557 (2020). DOI: 10.1152/physiolgenomics.00089.2020.

102. Scialo, F. *et al.* ACE2: The Major Cell Entry Receptor for SARS-CoV-2. *Lung* **198**, 867–877 (2020). DOI: 10.1007/s00408-020-00408-4.

103. Jackson, C. B., Farzan, M., Chen, B. & Choe, H. Mechanisms of SARS-CoV-2 entry into cells. *Nat. Rev. Mol. Cell Biol.* **23**, 3–20 (2022). DOI: 10.1038/s41580-021-00418-x.

104. Liu, S., Wu, S. & Jiang, S. HIV Entry Inhibitors Targeting gp41: From Polypeptides to Small-Molecule Compounds. *Curr. Pharm. Des.* **13**, 143–162 (2007). DOI: 10.2174/138161207779313722.

105. Sodroski, J. G. HIV-1 Entry Inhibitors in the Side Pocket. *Cell* **99**, 243–246 (1999). DOI: 10.1016/S0092-8674(00)81655-4.

106. Melikyan, G. B. Membrane Fusion Mediated by Human Immunodeficiency Virus Envelope Glycoprotein. in *Current Topics in Membranes* vol. 68 81–106 (Elsevier Inc., 2011). DOI: 10.1016/B978-0-12-385891-7.00004-0.

107. Bewley, C. A., Louis, J. M., Ghirlando, R. & Marius Clore, G. Design of a novel peptide inhibitor of HIV fusion that disrupts the internal trimeric coiled-coil of gp41. *J. Biol. Chem.* **277**, 14238–14245 (2002). DOI: 10.1074/jbc.M201453200.

108. Jiang, S., Lin, K., Strick, N. & Neurath, A. R. HIV-1 inhibition by a peptide. *Nature* **365**, 113–113 (1993). DOI: 10.1038/365113a0.

109. Jiang, S. B., Lin, K., Strick, N. & Neurath, A. R. Inhibition of HIV-1 Infection by a Fusion Domain Binding Peptide from the HIV-1 Envelope Glycoprotein GP41. *Biochem. Biophys. Res. Commun.* **195**, 533–538 (1993). DOI: 10.1006/bbrc.1993.2078.

110. Debnath, A. K. Progress in identifying peptides and small-molecule inhibitors targeted to gp41 of HIV-1. *Expert Opin. Investig. Drugs* **15**, 465–478 (2006). DOI: 10.1517/13543784.15.5.465.

111. Kilby, J. M. *et al.* Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **4**, 1302–1307 (1998). DOI: 10.1038/3293.

112. Yu, F. *et al.* Approaches for identification of HIV-1 entry inhibitors targeting gp41 pocket. *Viruses* **5**, 127–149 (2013). DOI: 10.3390/v5010127.

113. Jiang, S. *et al.* N-Substituted Pyrrole Derivatives as Novel Human Immunodeficiency Virus Type 1 Entry Inhibitors That Interfere with the gp41 Six-Helix Bundle Formation and Block Virus Fusion N-Substituted Pyrrole Derivatives as Novel Human Immunodeficiency Virus Type 1 E. *Antimicrob. Agents. Ch.* **48**, 4349–4359 (2004). DOI: 10.1128/AAC.48.11.4349.

114. Crespillo, S. *et al.* Single-chain protein mimetics of the N-terminal heptad-repeat region of gp41 with potential as anti–HIV-1 drugs. *Proc. Natl. Acad. Sci.* **111**, 18207–18212 (2014). DOI: 10.1073/pnas.1413592112.

115. Li, W., Lu, L., Li, W. & Jiang, S. Small-molecule HIV-1 entry inhibitors targeting gp120 and gp41: a patent review (2010-2015). *Expert Opin. Ther. Pat.* **3776**, 1–13 (2017). DOI: 10.1080/13543776.2017.1281249.

116. He, Y. *et al.* Design and evaluation of sifuvirtide, a novel HIV-1 fusion inhibitor. *J. Biol. Chem.* **283**, 11126–11134 (2008). DOI: 10.1074/jbc.M800200200.

117. Root, M. J., Kay, M. S. & Kim, P. S. Protein design of an HIV-1 entry inhibitor. *Science* **291**, 884–888 (2001). DOI: 10.1126/science.1057453.

118. Welch, B. D., VanDemark, A. P., Heroux, A., Hill, C. P. & Kay, M. S. Potent D-peptide inhibitors of HIV-1 entry. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16828–33 (2007). DOI: 10.1073/pnas.0708109104.

119. Welch, B. D. *et al.* Design of a Potent d -Peptide HIV-1 Entry Inhibitor with a Strong Barrier to Resistance . *J. Virol.* **84**, 11235–11244 (2010). DOI: 10.1128/jvi.01339-10.

120. Eckert, D. M., Malashkevich, V. N., Hong, L. H., Carr, P. A. & Kim, P. S. Inhibiting HIV-1 entry: Discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **99**, 103–115 (1999). DOI: 10.1016/S0092-8674(00)80066-5.

121. Miller, S. M., Simon, R. J., Ng, S., Zuckermann, R. N. & Kerr, J. M. Comparison of the proteolytic susceptibilities of Homologous L-amino Acid, D-Amino Acid and N-substituted Glycine Peptide and Peptoid Oligomers, Miller et al 1995.pdf. *Drug Dev. Res.* **32**, 20–32 (1995).

122. Zhao, L. & Lu, W. Mirror image proteins. *Curr. Opin. Chem. Biol.* **22**, 56–61 (2014). DOI: 10.1016/j.cbpa.2014.09.019.

123. Pang, W., Tom, S. C. & Zheng, Y. T. Current peptide HIV type-1 fusion

inhibitors. *Antivir. Chem. Chemother.* **20**, 1–18 (2009). DOI: 10.3851/IMP1369.

124. Orkin, C. *et al.* Opening the door on entry inhibitors in HIV: Redefining the use of entry inhibitors in heavily treatment experienced and treatment-limited individuals living with HIV. *HIV Med.* 1–11 (2022) doi:10.1111/hiv.13288. DOI: 10.1111/hiv.13288.

125. Chong, H. *et al.* Biophysical Property and Broad Anti-HIV Activity of Albuvirtide, a 3-Maleimimidopropionic Acid-Modified Peptide Fusion Inhibitor. *PLoS One* **7**, e32599 (2012). DOI: 10.1371/journal.pone.0032599.

126. Zhang, H. *et al.* Combination of long-acting HIV fusion inhibitor albuvirtide and LPV/r showed potent efficacy in HIV-1 patients. *AIDS Res. Ther.* **13**, 18–21 (2016). DOI: 10.1186/s12981-016-0091-1.

127. Saag, M. S. New and investigational antiretroviral drugs for HIV infection: Mechanisms of action and early research findings. *Top. Antivir. Med.* **20**, 162–167 (2012).

128. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. Core Structure of gp41 from the HIV Envelope Glycoprotein. *Cell* **89**, 263–273 (1997). DOI: 10.1016/S0092-8674(00)80205-6.

129. Chan, D. C., Chutkowski, C. T. & Kim, P. S. Evidence that a prominent cavity in the coiled coil of HIV type 1 gp41 is an attractive drug target. *Med. Sci.* **95**, 15613–15617 (1998). DOI: 10.1073/pnas.95.26.15613.

130. Di, L., Kerns, E. & Carter, G. Drug-Like Property Concepts in Pharmaceutical Design. *Curr. Pharm. Des.* **15**, 2184–2194 (2009). DOI: 10.2174/138161209788682479.

131. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44**, 235–249 (2000). DOI: 10.1016/S1056-8719(00)00107-6.

132. Lu, K; Asyifah, MR; Shao, F; Zhang, D. Development of HIV-1 Fusion Inhibitors Targeting gp41. *Curr. Med. Chem.* **21**, 1976–1996 (2014).

133. Miller, M. D. *et al.* A human monoclonal antibody neutralizes diverse HIV-1 isolates by binding a critical gp41 epitope. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14759–14764 (2005). DOI: 10.1073/pnas.0506927102.

134. Corti, D. *et al.* Analysis of Memory B Cell Responses and Isolation of Novel Monoclonal Antibodies with Neutralizing Breadth from HIV-1-Infected Individuals. *PLoS One* **5**, e8805 (2010). DOI: 10.1371/journal.pone.0008805.

135. Rubio, A. A. *et al.* A Derivative of the D5 Monoclonal Antibody That Targets the gp41 N-Heptad Repeat of HIV-1 with Broad Tier-2-Neutralizing Activity. *J. Virol.* **95**, e02350-20 (2021). DOI: 10.1128/JVI.02350-20.

136. Eckert, D. M. & Kim, P. S. Mechanisms of Viral Membrane Fusion and Its Inhibition. *Annu. Rev. Biochem.* **70**, 777–810 (2001). DOI: 10.1146/annurev.biochem.70.1.777.

137. Eckert, D. M. & Kim, P. S. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci.* **98**, 11187–11192 (2001). DOI: 10.1073/pnas.201392898.

138. Wild, C., Oas, T., Mcdanal, C., Bolognesi, D. & Matthews, T. A synthetic

peptide inhibitor of human immunodeficiency virus replication: Correlation between solution structure and viral inhibition. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10537–10541 (1992). DOI: 10.1073/pnas.89.21.10537.

139. Eckert, D. M. & Kim, P. S. Design of potent inhibitors of HIV-1 entry from the gp41 N-peptide region. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11187–11192 (2001). DOI: 10.1073/pnas.201392898.

140. Xing, L. *et al.* A Five-Helix-Based SARS-CoV-2 Fusion Inhibitor Targeting Heptad Repeat 2 Domain against SARS-CoV-2 and Its Variants of Concern. *Viruses* **14**, 597 (2022). DOI: 10.3390/v14030597.

141. Louis, J. M., Bewley, C. A. & Clore, G. M. Design and Properties of NCCG-gp41, a Chimeric gp41 Molecule with Nanomolar HIV Fusion Inhibitory Activity. *J. Biol. Chem.* **276**, 29485–29489 (2001). DOI: 10.1074/jbc.C100317200.

142. Ni, L., Gao, G. F. & Tien, P. Rational design of highly potent HIV-1 fusion inhibitory proteins: Implication for developing antiviral therapeutics. *Biochem. Biophys. Res. Commun.* **332**, 831–836 (2005). DOI: 10.1016/j.bbrc.2005.05.037.

143. Pang, W. *et al.* Recombinant protein of heptad-repeat HR212, a stable fusion inhibitor with potent anti-HIV action in vitro. *Virology* **377**, 80–87 (2008). DOI: 10.1016/j.virol.2008.04.013.

144. Jurado, S. *et al.* Structural and Thermodynamic Analysis of HIV-1 Fusion Inhibition Using Small gp41 Mimetic Proteins. *J. Mol. Biol.* **431**, 3091–3106 (2019). DOI: 10.1016/j.jmb.2019.06.022.

145. Jurado, S., Cano-Muñoz, M., Polo-Megías, D., Conejero-Lara, F. & Morel, B. Thermodynamic dissection of the interface between HIV-1 gp41 heptad repeats reveals cooperative interactions and allosteric effects. *Arch. Biochem. Biophys.* **688**, 108401 (2020). DOI: 10.1016/j.abb.2020.108401.

146. Root, M. J., Kay, M. S. & Kim, P. S. Protein Design of an HIV-1 Entry Inhibitor. *Science (80-.).* **291**, 884–888 (2001). DOI: 10.1126/science.1057453.

147. Steger, H. K. & Root, M. J. Kinetic dependence to HIV-1 entry inhibition. *J. Biol. Chem.* **281**, 25813–25821 (2006). DOI: 10.1074/jbc.M601457200.

148. Jurado, S. *et al.* Probing Vulnerability of the gp41 C-Terminal Heptad Repeat as Target for Miniprotein HIV Inhibitors. *J. Mol. Biol.* **432**, 5577–5592 (2020). DOI: 10.1016/j.jmb.2020.08.010.

149. Dimitrov, D. S. *Therapeutic Proteins*. *Methods in Molecular Biology* vol. 899 (Humana Press, 2012). DOI: 10.1007/978-1-61779-921-1.

150. Gurevich, E. V & Gurevich, V. V. Therapeutic Potential of Small Molecules and Engineered Proteins. in *Arrestins - Pharmacology and Therapeutic Potential. Handbook of Experimental Pharmacology*, 1–12 (2014). doi:10.1007/978-3-642-41199-1_1. DOI: 10.1007/978-3-642-41199-1_1.

151. Maurer, T. S., Edwards, M., Hepworth, D., Verhoest, P. & Allerton, C. M. N. Designing small molecules for therapeutic success : A contemporary perspective. *Drug Discov. Today* **27**, 538–546 (2022). DOI: 10.1016/j.drudis.2021.09.017.

152. Ngo, H. X. & Garneau-Tsodikova, S. What are the drugs of the future? *Medchemcomm* **9**, 757–758 (2018). DOI: 10.1039/c8md90019a.

153. Li, Q. Application of Fragment-Based Drug Discovery to Versatile Targets. *Front. Mol. Biosci.* **7**, 1–13 (2020). DOI: 10.3389/fmolb.2020.00180.

154. Cano-Muñoz, M., Jurado, S., Morel, B. & Conejero-Lara, F. Conformational flexibility of the conserved hydrophobic pocket of HIV-1 gp41. Implications for the discovery of small-molecule fusion inhibitors. *Int. J. Biol. Macromol.* **192**, 90–99 (2021). DOI: 10.1016/j.ijbiomac.2021.09.198.

155. Cano-muñoz, M. *et al.* Extremely thermostabilizing core mutations in coiled-coil mimetic proteins of hiv-1 gp41 produce diverse effects on target binding but do not affect their inhibitory activity. *Biomolecules* **11**, 1–16 (2021). DOI: 10.3390/biom11040566.

156. Cano-Muñoz, M. *et al.* Conformational Stabilization of Gp41-Mimetic Miniproteins Opens Up New Ways of Inhibiting HIV-1 Fusion. *Int. J. Mol. Sci.* **23**, 2794 (2022). DOI: 10.3390/ijms23052794.