UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

DEPARTMENT OF INFORMATIONAL ENGINEERING

PHD PROGRAM IN TECHNOLOGY FOR HEALTH

XXXVII CYCLE

Specific molecular modelling and dynamics to study the structure/function relationship in clonogenic HIV-1 p17 protein variants: preliminary set up of a computational drug repositioning pipeline

---

Candidate:
Alessandro Rondina

Supervisor:
Prof. Marco Rusnati

Co-Supervisor:
Francesca Caccuri

SSD MED/04

INDEX

0. RIASSUNTO

Nel mio progetto di ricerca ho studiato le varianti della proteina di matrice p17 di HIV-1 (vp17s), caratterizzate da inserzioni di amminoacidi nella regione C-terminale della proteina virale. A differenza della p17 wild-type (refp17), queste varianti mostrano una forte capacità di promuovere la crescita delle cellule B e sono prevalentemente riscontrate nei pazienti affetti da linfoma rispetto ai pazienti HIV-1 positivi senza linfoma, suggerendo un loro ruolo chiave nella linfomagenesi nei pazienti HIV-1+. Attraverso l'analisi genomica, abbiamo identificato hot spots di mutazione nel gene Gag, concentrandoci in particolare sulla porzione della matrice (MA), che sembrano originarsi da un arresto della trascrittasi inversa (RT) durante la replicazione virale. Questo fenomeno è probabilmente guidato da sequenze palindromiche e ripetitive, oltre a un elevato contenuto di adenina, creando instabilità genomica che favorisce eventi di ricombinazione e la generazione di vp17s.

Il progetto ha incluso anche un approccio sperimentale per validare le analisi bioinformatiche: i miei colleghi hanno misurato la frequenza degli eventi di ricombinazione associati agli hot spots mutazionali identificati. I risultati hanno confermato un ampio spettro di eventi di ricombinazione che si verificano in queste regioni instabili, contribuendo alla generazione di vp17s con attività clonogenica.

Per approfondire i meccanismi molecolari alla base del potenziale clonogenico delle vp17s, ho condotto studi computazionali, come le simulazioni di dinamica molecolare (MDs). Le MDs hanno rivelato che le vp17s subiscono cambiamenti conformazionali significativi, alterando notevolmente la loro struttura secondaria ed il network di legami idrogeno. In particolare, i residui idrofobici Trp16 e Tyr29 si dissociano dal loro nucleo idrofobico, cruciale per la stabilità di refp17. Questo distacco espone epitopi funzionali, in grado ipoteticamente di attivare vie di segnalazione proliferativa nelle cellule B. Mutazioni sito-specifiche di Trp16 e Tyr29 in refp17 hanno prodotto una proteina con attività clonogenica, confermando l'importanza di questi residui nella regolazione della crescita cellulare. Inoltre, ho esplorato approcci di riposizionamento farmacologico per identificare farmaci esistenti in grado di contrastare le vp17s e inibire la loro attività clonogenica.

Le MDs e gli studi di docking hanno identificato alcune molecole che potrebbero interagire con specifici residui delle vp17, suggerendo nuovi potenziali trattamenti per prevenire lo sviluppo di linfomi nei pazienti HIV-1+.

Infine, ho trascorso sei mesi nel laboratorio di Pietro Liò presso il Computer Laboratory dell'Università di Cambridge, focalizzandomi sull'integrazione dei metodi di simulazione di dinamica molecolare con metodologie di intelligenza artificiale e sviluppando strategie di validazione robuste per i modelli generativi. Le conoscenze acquisite da questa esperienza saranno applicate per approfondire i miei futuri studi su target terapeutici sia virali che non virali.

In conclusione, il mio lavoro ha contribuito a chiarire i meccanismi molecolari alla base dell'evoluzione e della funzione delle vp17s, identificando vulnerabilità strutturali che potrebbero essere sfruttate per lo sviluppo di nuove strategie terapeutiche. Questo studio non solo avanza la comprensione delle vp17s nella patogenesi dell'HIV, ma apre anche nuove prospettive per interventi farmacologici mirati a prevenire i linfomi associati.

1. ABBREVIATIONS

aa: amino acid
AIDS: Acquired Immunodeficiency Syndrome
αD: alphaD
αP: alphaP
AMBER: Assisted Model Building and Energy Refinement
AP-3: adaptor protein-3
ART: Antiretroviral therapy
BPs: binding pockets
CA: capsid
ECs: endothelial cells
Ed: dihedral energies
ED: EthreshD
EGFR: epidermal growth factor receptor
Ep: potential energy
EP: EthreshP
FPLC: fast protein liquid chromatograph
GROMOS: GROningen MOlecular Simulation
HAART: highly active antiretroviral therapy
H1-H5: alpha-helices 1-5
HC: hydrophobic cluster
HIV: Human Immunodeficiency Virus
HPV: human papillomavirus
HSPGs: heparan sulfate proteoglycans
IFN-γ: cytokine interferon-gamma
ILV: isoleucine, leucine, and valine
Kd: dissociation constant
KLH: Keyhole Limpet Hemocyanin
LINCS: LINear Constraint Solver
LTRs: Long Terminal Repeats
MA: matrix
mAb: monoclonal antibody
MCP-1: monocyte chemoattractant protein-1
MDs: Molecular Dynamics simulations
Myr: myristyl group
NC: nucleocapsid
NHL: non-Hodgkin lymphoma
NK: natural killer
NMR: nuclear magnetic resonance
P17Rs: p17 receptors
PAR1: protease-activated receptor 1
PCA: principal component analysis
PCs: Principal Components
pDCs: plasmacytoid dendritic cells
p17W16A+Y29A: p17 carrying W16A and Y29A mutations
PIP2: phosphatidylinositol-4,5-bisphosphate
PIC: pre-integration complex
PI: propidium iodide
PM: plasma membrane
PS: phosphatidylserine
refp17: wild-type p17
RT: reverse transcriptase
SD1: major splice donor site
SIV: simian immunodeficiency virus
TNF: tumor necrosis factor
vp17s: p17 variants
WT: wild type
Xn: chi-n
Φ: phi
Ψ: psi

2. INTRODUCTION

2.1. Human Immunodeficiency Virus Type 1

HIV, the acronym for Human Immunodeficiency Virus, is the causative agent of AIDS (Acquired Immunodeficiency Syndrome). Since its first identification in the early 1980s, HIV has become one of the most serious global public health issues, affecting millions of people worldwide.

HIV is a single-stranded RNA virus that belongs to the Retroviridae family, which includes other viruses known to cause neoplastic diseases in certain animal species (Gonda M.A. et al., 1985). Among retroviruses, HIV is classified within the Lentivirus genus, distinguished by a particularly long and complex replication cycle, characterized by a latent phase that can last for many years before clinical symptoms appear (Myers G. et al., 1992).

The life cycle of HIV is marked by the integration of its viral genome into the host cell's DNA mediated by reverse transcriptase (RT), an RNA-dependent DNA polymerase (Basu V. P. et al., 2008) that allows the virus to establish a persistent infection in host cells. HIV is primarily transmitted through contact with infected blood, semen, vaginal secretions, and breast milk (La Placa M., Principles of Medical Microbiology, 524-548; 561-569; 665-679, ed. Esculapio). The most common mode of transmission is sexual, but the virus can also be spread through the use of contaminated needles, infected blood transfusions, and from mother to child during pregnancy, childbirth, or breastfeeding (Re M.C., The Virology Laboratory in HIV Diagnosis and Patient Monitoring).

Once inside the body, HIV targets the cell of the immune system, including CD4+ T lymphocytes, macrophages, and dendritic cells (Frankel A.D. et al., 1998) that, over time, are destroyed by the virus, leading to progressive immunodeficiency. The initial stage of infection, often asymptomatic or marked by flu-like symptoms, can last from several weeks to several years. Without treatment, the infection progresses to more advanced stages, eventually resulting in AIDS, a state of severe immunodeficiency where the body becomes vulnerable to opportunistic infections and cancers, ultimately leading to death (Macher A.M. et al., 1988).

Since the discovery of HIV, huge progress has been made in managing the infection. Antiretroviral therapy (ART) has revolutionized the prognosis of HIV, transforming what was once a fatal disease into a manageable chronic condition (Seddon J. et al., 2011). ART works by suppressing viral replication, reducing viral load in the blood to undetectable levels, and allowing partial recovery of the immune system. However, ART is not a cure, as the virus can remain latent in infected cells, and discontinuing therapy leads to a resurgence of viral replication (Finzi D. et al., 1997). Despite the improvements in understanding the virus and developing effective therapies, HIV continues to pose substantial medical and social challenges (Sharp P.M. et al., 2011). Access to treatment is still limited in many parts of the world, particularly in developing countries. Research continues to focus not only on improving existing therapies but also on developing an effective vaccine and curative strategies that could completely eradicate the virus from the body (Cohen J., 2014).

2.1.1. The HIV Virion: Structure, Components, and Maturation

The HIV virion has a diameter ranging from 100 to 120 nanometers and is composed of 3 main parts: the envelope, matrix, and core, each performing specific and critical functions essential for the virus's infectivity and survival (**Fig. 1**).

The envelope, which forms the outermost layer of the HIV virion, is composed of a lipid bilayer that originates from the host cell membrane during the budding process (Freed E.O. et al., 2007). This lipid layer provides protection to the virion and plays a crucial role in the recognition and attachment to target cells. Embedded within the envelope are two essential viral glycoproteins:

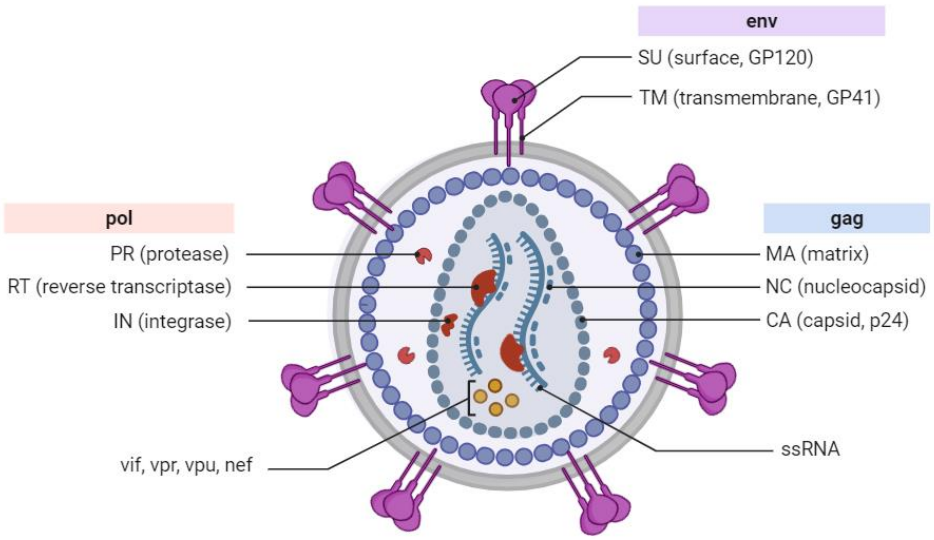

***Fig. 1 - HIV-1 virion**. The virion consists of a lipid bilayer with envelope glycoproteins (GP120, GP41) essential for host cell entry. Inside, the capsid (p24) encloses ssRNA genomes and enzymes (PR, RT, IN) necessary for replication. Structural (MA, NC) and accessory proteins (Vif, Vpr, Vpu, Nef) support infection and immune evasion.*

gp120 and gp41, both derived from the proteolytic cleavage of the gp160 precursor (Zhu P. et al., 2006). gp120 is crucial for virus's entry into host cells, as it binds to CD4 receptors on the surface of T lymphocytes. (Ganser-Pornillos B.K. et al., 2008). Following this binding event, gp41 facilitates the fusion between the viral and host cell membranes. This process is enabled by a fusogenic peptide within gp41 rich in glycine and hydrophobic residues that effectively interact with the cellular membrane promoting the merging of the two lipid bilayers (Freed E.O. et al., 2015). Beneath the envelope lies the virus matrix, composed of the p17 protein that contributes to the structural stability of the virion and plays a role in the proper assembly of the envelope glycoproteins (Muriaux D. et al., 2010).

The innermost part of the virion is the core or nucleocapsid, that houses the viral genome required for replication. The core is surrounded by a protein capsid made of the p24 protein that protects the viral genome and is essential for the virus's maturation and infectivity (**Fig. 2**) (Peng C. et al., 1989). Inside the core are two copies of the positive-sense RNA closely associated with nucleocapsid proteins p7 and p9, which protect them from degradation and assist in their encapsidation. The core also contains the enzymes for viral replication: RT, that converts viral RNA into DNA; integrase, that inserts the viral DNA into the host cell genome; protease, that processes viral polyproteins to produce the mature proteins for the assembly of new virions (Ohishi M. et al., 2011).

HIV virion assembly begins with the production of Gag and Gag-Pol polyproteins, that consists of 4 main domains: MA (matrix), CA (capsid), NC (nucleocapsid), and p6. The viral protease, also encoded within the Gag-Pol polyprotein, cleaves the Gag polyproteins into mature proteins, each serving specific functions in virion assembly. MA anchors the Gag protein to the viral envelope and incorporates the gp120 and gp41 into the assembling virion (Freed E.O. et al., 2015). CA mediates key interactions between Gag proteins during core formation. These interactions are critical for the virion's structural stability (Ganser-Pornillos B.K. et al., 2008). NC binds the viral RNA, facilitating its encapsidation (Muriaux D. et al., 2010). P6 interacts with a group of cellular proteins that facilitate the budding of the virion from the host cell (Ganser-Pornillos B.K. et al., 2008).

A crucial aspect of maturation is also the conversion of the RNA genome from a "kissing-loop" dimeric structure to an "extended-duplex" conformation, stabilized by the NC proteins, that is essential for genomic stability and proper integration into the host DNA (Ohishi M. et al., 2011).

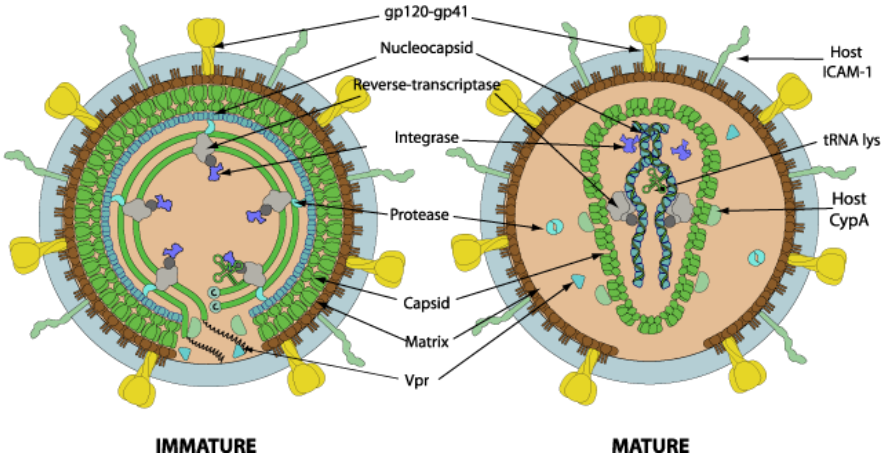

Fig. 2. Immature (left) and mature (right) HIV virion. *The immature virion has an incomplete capsid structure and uncleaved Gag polyproteins. In the mature virion, the capsid is fully formed, enclosing the RNA genome and viral enzymes, following Gag and Gag-Pol cleavage by protease, which is essential for infectivity.*

2.1.2. The HIV-1 Genome: Structure and Functions

As mentioned above, the genome of HIV consists of two strands of single-stranded, positive-sense RNA contains all the information necessary for the virus to invade human cells, integrate into the host genome, and produce and release new viral particles (**Fig. 3**) (Rambaut A. et al., 2004).
It is characterized by the presence of terminal repeat sequences known as Long Terminal Repeats (LTRs), located at both ends of the viral RNA that play a crucial role in regulating viral gene expression, affecting both the encapsidation of viral RNA into new particles and the integration of the viral genome into the host's DNA (Quinn T.C., 1994). Additionally, they serve as binding sites for numerous transcription factors, both viral and cellular, making them central to the modulation of the viral life cycle (Hahn B.H. et al., 2000).

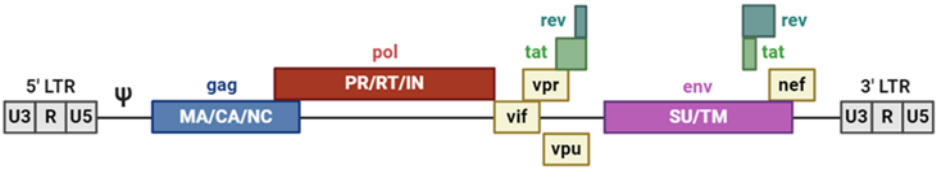

Fig. 3. HIV-1 genome*. The HIV-1 genome consists of structural (gag, pol, env), regulatory (tat, rev), and accessory (vif, vpr, vpu, nef) genes flanked by LTRs (long terminal repeats). These regions encode proteins essential for replication, assembly, and immune evasion.*

The HIV-1 genome contains also 9 major structural, regulatory and accessory genes that encode proteins essential for the assembly and function of viral particles (Sharp P.M. et al., 2011). Structural genes include gag encodes a polyprotein that is subsequently cleaved into several

structural proteins, including the matrix protein (MA), capsid protein (CA), nucleocapsid protein (NC). *Pol* encodes RT, protease and integrase. *Env* encodes gp120 and gp41.

Regulatory and accessory genes include: *tat* that encodes a potent activator of viral transcription which binds to the Trans-Activation Response element on the viral RNA, significantly enhancing the efficiency of viral transcription, thus enabling rapid production of viral proteins necessary for replication (Plantier J.C. et al., 2009). *Rev*, that encodes a protein that regulates the export of unspliced viral mRNA from the nucleus to the cytoplasm, which is essential for the synthesis of the structural and enzymatic proteins of the virus (Frankel A.D. et al., 1998). *Nef*, that downregulates CD4 receptors and class I MHC molecules on the surface of infected cells, helping the virus to evade immune surveillance and increasing viral infectivity (Frankel A.D. et al., 1998). Vif, that encodes a protein that counteracts the cellular protein APOBEC3G, which would otherwise induce lethal mutations in the viral genome, thus compromising viral replication (Basu V.P. et al., 2008). Vpu, that facilitates the release of mature virions from infected cell and degrades the CD4 receptor in the endoplasmic reticulum, preventing viral glycoproteins from becoming trapped and ensuring their expression on the surface of new viral particles (Wilusz j., 2013). Vpr, that plays a role in importing the viral DNA pre-integration complex into the nucleus and arrests the host cell cycle in the G2 phase, creating a more favourable environment for viral replication (Ocwieja K.E. et al., 2012).

HIV genome also contains the Primer Binding Site, which binds the tRNA required to initiate reverse transcription of viral RNA into DNA, the encapsidation signal (Ψ), necessary for the proper packaging of the viral genome into new virions, the Dimerization Initiation Site, which facilitates the incorporation of two RNA strands into the virion (Plantier J.C. et al., 2009), the major splice donor site (SD1), crucial for the production of subgenomic mRNA encoding viral accessory and regulatory proteins and the polypurine tract, critical for viral mRNA maturation by enabling polyadenylation, a process that stabilizes the mRNA and facilitates translation (Van Heuverswyn F. et al., 2006).

This highly remarkable variability of the HIV-1 genome, driven by the low fidelity of RT and the potential for recombination between different viral strains during co-infection, contributes significantly to the virus's ability to evolve rapidly. This genetic diversity poses significant challenges for vaccine development and antiviral treatment, as the virus can quickly adapt and develop resistance to therapeutic agents (Korber B. et al., 1998).

2.1.3. The HIV-1 Replication Cycle

The HIV-1 replication cycle is a highly sophisticated process that enables the virus to infect host cells, replicate, and spread throughout the organism. It begins with the recognition of target CD4-expressing cells, including T helper lymphocytes, dendritic cells, macrophages and microglial cells in the nervous system (**Fig. 4**). The high affinity binding of gp120 to the CD4 receptor establishes the initial contact between the virus and the host cell but, alone, it is insufficient for viral entry into the cell, being other specific co-receptors required: R5 viruses, which primarily infect macrophages and CD4+ T cells, utilize CCR5 co-receptor, while X4 viruses, which selectively infect T lymphocytes, utilize CXCR4 co-receptor. Dual-tropic HIV-1 strains, known as R5X4 viruses are capable of infecting both cell types. In certain cases, HIV-1 can infect cells that do not express the CD4 receptor, such as neural cells, although the precise mechanism of entry into these cells remains unclear.

Following the initial binding to CD4, gp120 undergoes a series of conformational changes that expose the binding sites for co-receptors. These changes induce further modifications in gp120, allowing the exposure of the fusion peptide located at the N-terminal end of gp41. At this point, gp41 inserts into the host cell membrane, facilitating the fusion between the viral and cellular membranes. This fusion process mediates the transport of the viral core into the cytoplasm of the host cell, while the viral envelope remains outside the cell.

Once inside the cytoplasm, the virus enters the uncoating phase, where the viral core's protein shell is degraded, releasing the viral genome and the enzymes, including RT. Reverse transcription thus occurs in the cytoplasm of the host cell, beginning with the synthesis of a complementary DNA strand from the viral RNA followed by the degradation of the RNA strand by the RNase H enzyme, that leaves only the DNA strand. RT then synthesizes a second DNA strand complementary to the first, completing the formation of double-stranded viral DNA.

The viral DNA associates with various viral proteins, such as integrase, p17, and Vpr, forming the pre-integration complex (PIC) that is actively transported from the cytoplasm to the host cell's nucleus by Vpr. Once inside the nucleus, the viral DNA integrates into the host cell genome through the action of integrase, a crucial step that renders HIV-1 infection permanent. During this phase, the viral DNA undergoes modifications that result in the formation of the LTRs described above.

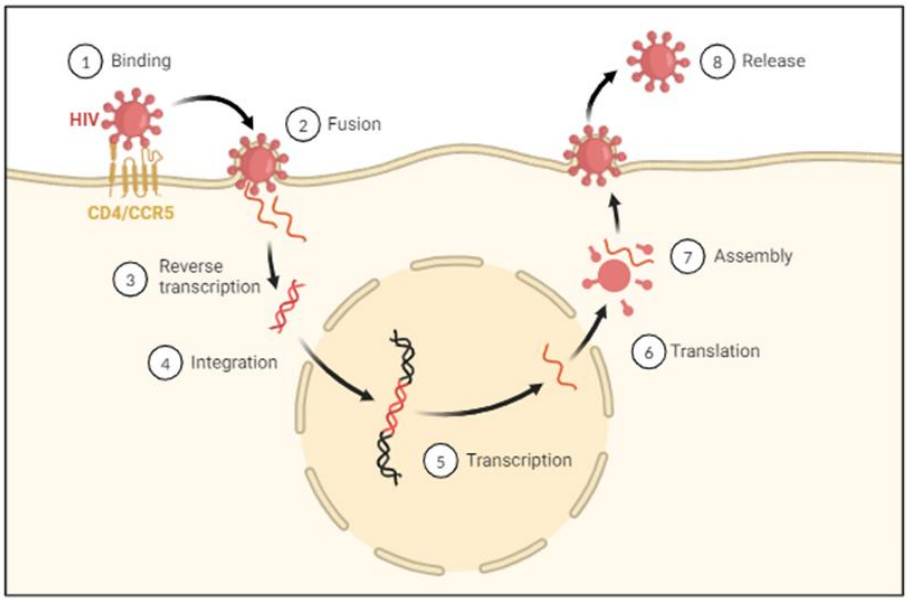

Fig. 4. HIV Mechanism of infection. *HIV infects host cells by binding to receptors such as CD4 and CCR5, followed by membrane fusion and reverse transcription of its RNA into DNA. The viral DNA integrates into the host genome, enabling transcription and translation of viral proteins. Newly assembled virions are then released to infect other cells.*

The viral DNA can remain into the host genome in a latent state or be transcribed alongside cellular DNA. In this latter case, the host transcriptional machinery produces new viral mRNA, which is exported to the cytoplasm for the synthesis of precursor viral proteins that, during the assembly phase, accumulate in the cytoplasm along with genomic RNA, with the consequent formation of the viral capsid and the inclusion of viral RNA and essential enzymes within the virion.

The final stage of the replication cycle is virion maturation and budding through the host cell membrane, where viral envelope proteins have already been incorporated. In this way, the virion acquires its lipid envelope, necessary for its ability to infect new cells. Mature viral particles are then released from the infected cell, completing the HIV-1 replication cycle. This release can irreversibly damage the cell, leading to its lysis, particularly in helper T lymphocytes, or allow the cell to survive, as in the case of macrophages and dendritic cells, which serve as viral reservoirs (Freed E.O., 2015).

2.1.4. Genetic Variability of HIV-1

The genetic variability of HIV-1 is one of the virus's most distinctive features, crucial for its ability to persist in the host organism and evade both immune defences and therapeutic interventions (Goodenow M. et al., 1989). It is evident not only between viruses isolated from

different individuals but also among viruses within the same person over infection (Li J.P. et al., 2010). The primary mechanisms driving this variability are mutation and genetic recombination that occur during the viral replication cycle (Wymant C. et al., 2018).

Mutations arise mainly due to the error-prone nature of HIV-1 RT, that introduces errors at a high rate during the synthesis of DNA from the viral RNA template. This leads to the frequent occurrence of point mutations, insertions, and deletions, which significantly alter the viral genome (Ji J.P. and Loeb L.A., 1992). Studies have shown that the in vivo mutation rate is approximately 3×10^-5 errors per base pair per replication cycle (Abram M.E. et al., 2010). The variability is particularly evident in env and nef, which exhibit a high degree of mutability (Simon-Loriere E. et al., 2009), while env contains both conserved regions (C1-C4) and hypervariable regions (V1-V5), allowing the virus to vary its antigenic surface and evade immune surveillance (Shankarappa et al., 1999). These features make the development of an effective vaccine extremely difficult.

In addition to mutation, genetic recombination represents another crucial mechanism of diversification that occurs during the synthesis of proviral DNA, when RT performs strand transfers. These transfers can be obligatory, as in the generation of LTRs, or optional, when the enzyme "jumps" between internal regions of co-packaged RNA, resulting in the formation of recombinant DNA molecules (Goldschmidt V. et al., 2002). This template switching occurs frequently throughout the HIV-1 genome and contributes to genetic diversification by mixing mutations within the viral quasispecies, thereby accelerating adaptation through natural selection (Levy D.N. et al., 2004).

The viral quasispecies are continuously shaped by mutations and recombinations and by selective pressure exerted by the host immune system (Liu Y. et al., 2020). Quasispecies complexity increases as infection progresses, making disease control and therapy increasingly challenging (Domingo E. et al., 2012). Studies have demonstrated that the genetic diversity of HIV-1 quasispecies in an individual in the advanced stages of infection is comparable to the global genetic variation of the influenza virus over an entire year (Korber B. et al., 2001), making HIV-1 genetic variability a key factor that contributes to the virus's ability to evade the immune system, develop drug resistance, and complicate the development of an effective vaccine (Mansky L.M. and Temin H.M., 1995). It derives that the comprehension of the molecular mechanisms underlying this variability is essential for addressing the challenges posed by HIV-1 infection and treatment (Simon-Loriere E. et al., 2011).

2.1.5. HIV-1 and Cancer

HIV-1 infection is associated with an increased risk of developing various types of cancer, mainly due to the immunodeficiency induced by the virus, which compromises the immune system's ability to monitor and control the proliferation of oncogenic cells. Kaposi’s sarcoma is one of the most clinically relevant cancers associated with HIV-1. It is a vascular tumor linked to infection with human herpesvirus type 8 and its onset is strongly correlated with the immunosuppression caused by HIV-1, which increases the risk of developing this neoplasm by thousands of times (Hernández-Ramírez R.U. et al., 2020). The introduction of highly active antiretroviral therapy (HAART) led to a significant reduction (68%) in the risk of Kaposi’s sarcoma (Bedimo R. et al., 2004).

Lymphomas are also closely linked to HIV-1 infection, with non-Hodgkin lymphoma (NHL) being of particular concern. It is characterized by uncontrolled proliferation of B cells and, indeed, HIV-1 leads to dysregulation of B cell control the most common subtypes of lymphoma in AIDS patients are primary brain lymphoma, immunoblastic large-cell lymphoma and Burkitt lymphoma. Primary brain lymphoma is thousands of times more common in AIDS patients than in the general population (Coté T.R. et al., 1996) while Immunoblastic large-cell lymphoma and Burkitt lymphoma are hundreds of times more common in AIDS patients compared to the general population (Engels E.A. et al., 2005; Engels E.A. et al., 2008). A rare subtype of lymphoma associated with HIV-1 is

primary effusion lymphoma, which presents as pleural or peritoneal effusions and is linked to human herpesvirus type 8 infection and Kaposi's sarcoma (Gibson T.M. et al., 2014).

The data on the effects of HAART of the incidence of HIV-1-associated lymphomas are variable: some studies reported a decline in non-Hodgkin lymphoma (Besson C. et al., 2017), while others have found that Burkitt lymphoma remained relatively unchanged (Engels E.A. et al., 2005).

Hodgkin lymphoma, though less common than non-Hodgkin lymphoma, is also associated with a tenfold increased risk in HIV + patients. is characterized by a higher prevalence of Epstein-Barr virus in HIV + patients compared to HIV-negative individuals (Shannon-Lowe C. et al., 2017). With the introduction of HAART, an overall increase in Hodgkin lymphoma rates has been observed, in contrast to the decline seen for non-Hodgkin lymphoma (Clifford G.M. et al., 2005).

HIV + patients also show an increased risk to develop solid cancers: HIV + women are at increased risk of developing cervical cancer, largely due to co-infection with human papillomavirus (HPV). HIV-1 infection facilitates the persistence of HPV and the progression to precancerous and cancerous lesions of the cervix. The risk of invasive cervical cancer remains high even in the HAART era (Dal Maso L. et al., 2009). Anal cancer is also associated with HIV-1 infection, and to HPV infection. The risk of anal cancer is particularly increased among HIV + homosexual that exhibit also very high rates of HPV infection (Patel K. et al., 2008). The introduction of HAART has not reduced significantly the incidence of anal cancer, with some studies even suggesting an increase in rates, possibly due to increased patient longevity, heightened screening and early diagnosis (Patel K. et al., 2008). Lung cancer incidence is also increased in HIV + patients (Grulich A.E. et al., 2007), likely due to a combination of factors such as the high prevalence of smoking in this population and HIV-1-induced immunosuppression (Dal Maso L. et al., 2003). Liver cancer, particularly hepatocellular carcinoma, is more common in HIV + patients, especially among those co-infected with hepatitis B or C viruses or hepatitis C viruses (Kramer J.R. et al., 2005). The risk of hepatocellular carcinoma is higher after HAART era compared to previous years (Hessol N.A. et al., 2007).

In summary, HIV-1 infection increases the risk of cancers by indirect mechanisms related to immunodeficiency and co-infection with oncogenic viruses. The risk remains elevated also after HAART, making cancers a main cause of mortality in HIV-1 patients (Engels E.A. et al., 2008).

2.1.6. HIV-1 Matrix Protein (p17)

P17 is a 132-amino acid (aa) protein composed of 5 α-helices and a highly basic platform of 3 β-strands, which together form a compact globular domain capped by a β-sheet. It is a critical component of the virus, serving both structural and regulatory functions that are pivotal for the virus's life cycle and its interactions with host cells (**Fig. 5**).

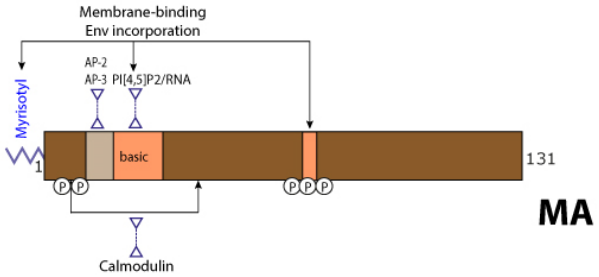

Fig. 5. Schematic representation of p17. *The matrix protein (p17) mediates membrane binding and Env incorporation via interactions with host factors such as AP-2, AP-3, PI(4,5)P2, RNA, and calmodulin. Myristoylation and basic regions contribute to its localization and function during viral assembly.*

2.1.6.1. Role of p17 in life cycle

In the early stages of the HIV-1 life cycle, p17 directs Gag polyprotein precursor to the plasma membrane (PM) of infected cells, where viral assembly occurs. This occurs via myristoylation of p17 at its N-terminal glycine residue, a modification that anchors the protein to PM. This association is further stabilized by a highly basic region in p17 that interacts with negatively charged phospholipids on the inner leaflet of the PM, specifically phosphatidylinositol-4,5-bisphosphate (PIP2) that not only anchors Gag to the PM but also induces a conformational change in p17 that exposes its myristyl group (Myr), promoting stable membrane binding. This interaction ensures that viral assembly occurs at the appropriate cellular site, and any disruption in this process, such as mutations in p17 or depletion of PI(4,5)P2, can result in the mis-localization of Gag to internal membranes, leading to the formation of non-infectious viral particles (Saad J.S. et al., 2006).

P17 is also critical for the incorporation of gp120 and gp41 into the budding virion, a process essential for the infectivity of the virus. P17 facilitates the incorporation of these glycoproteins by directly interacting with the cytoplasmic tail of gp41. Mutations in p17 that disrupt this interaction lead to the production of non-infectious virions (Freed E.O. et al., 1996). After assembly, p17 also facilitates bud off from the host cell of the assembled virion since, remaining associated with the inner leaflet of the viral membrane, it forms a protective shell that maintains the structural integrity of the virion as it exits the host cell and enters the extracellular environment.

P17 is also implicated in early post-entry events of HIV-1 infection, particularly in the nuclear import of PIC. After the virus enters a target cell, reverse transcription occurs in the cytoplasm, and the newly synthesized viral DNA must be transported into the nucleus for integration, a critical step for productive infection. A fraction of p17 is incorporated into the PIC, and, though its 2 nuclear localization signals, facilitates PIC nuclear import, particularly in non-dividing cells like macrophages, where the nuclear envelope remains intact and entry into the nucleus is more challenging (Bukrinsky M.I. et al., 1993). Conformational changes or post-translational modifications of p17 may modulate its involvement in nuclear import, allowing the virus to adapt to different cellular environments.

In macrophages, HIV-1 exhibits an alternative assembly pathway where viral particles are assembled in multivesicular bodies. P17 is integral to this process, interacting with the adaptor protein-3 (AP-3) complex, which is involved in the sorting of cargo proteins within the endosomal system. This interaction directs Gag to the multivesicular bodies, where viral assembly occurs. The assembled virions are then released from the cell through an exosomal pathway, a process that differs from the typical budding seen at the PM in other cell types. Disruption of the interaction between p17 and the AP-3 complex impairs viral assembly and release, highlighting the importance of p17 in facilitating the adaptation of HIV-1 to different cellular environments (Dong X. et al., 2005).

2.1.6.2. P17 receptors and effects on immune cells

P17 shares structural similarities with interferon-gamma (IFN-γ), although it does not replicate the immunomodulatory activities of IFN-γ nor bind to its receptors (Matthews S. et al., 1994). Despite this, the structural homology has prompted investigations into whether p17 might function similarly to a cytokine, a hypothesis supported by its interactions with specific cellular receptors.

The primary receptors for p17, known as p17 receptors (p17Rs), are constitutively expressed on B cells and are upregulated on T cells following stimulation with mitogens like IL-2. The of p17 to p17Rs is characterized by a high binding affinity, with a dissociation constant (Kd) of approximately 2.5×10^{-8} M (De Francesco M.A. et al., 2002). This strong interaction is crucial for p17's ability to enhance T cell activation and proliferation. While p17 alone does not significantly affect resting T cells, it acts synergistically with IL-2 to promote cell proliferation and the production of pro-inflammatory cytokines like IFN-γ and tumor necrosis factor (TNF)-α (De Francesco M.A. et al., 1998). Additionally, p17 can counteract the anti-inflammatory effects of IL-4, restoring the

production of IFN-γ and TNF-α in IL-2-stimulated cells (De Francesco M.A. et al., 2002). Taken together, these effects underscore the ability of p17 to establish a pro-inflammatory environment that supports HIV-1 replication and persistence.

P17 also influences the function of natural killer (NK) cells. When combined with IL-2, it enhances NK cell proliferation and the production of pro-inflammatory cytokines, although it does not affect NK cytotoxicity or induce apoptosis in target cells (Vitale M. et al., 2003). This selective modulation of NK cell activity suggests that p17 helps shape the immune response in a way that favours HIV-1 persistence without triggering excessive immune-mediated cell death.

In monocytes, p17 induces the secretion of monocyte chemoattractant protein-1 (MCP-1), a key player in recruiting monocytes themselves to sites of infection, thereby facilitating the establishment of an inflammatory microenvironment conducive to viral replication (Marini E. et al., 2008). Similarly, p17 induces B cell proliferation, contributing to the immune dysregulation seen in HIV-1 infection (Giagulli C. et al., 2011).

p17's impact on plasmocytoid dendritic cells (pDCs), that play a vital role in the antiviral immune response. Exposure to p17 leads to the upregulation of CCR7, a chemokine receptor essential for pDC migration to secondary lymphoid organs (Fiorentini S., Riboldi E. et al., 2008). However, this stimulation does not fully mature the pDCs, nor induces IFN-α production, a critical antiviral cytokine. Instead, p17-driven activation contributes to the mis-localization and depletion of pDCs in HIV-1-infected individuals, thereby facilitating viral persistence (Fiorentini S., Riboldi E. et al., 2008).

On neutrophils, monocytes and endothelial cells (ECs), p17 binds to CXCR1 and CXCR2. The binding of p17 to CXCR1 on monocytes triggers chemotactic migration, a process similar to that induced by the natural ligand IL-8, although p17 binds with a lower affinity (Giagulli C. et al., 2012). This chemotactic activity is crucial for recruiting monocytes to HIV-1 replication sites, where they can become infected and contribute to viral dissemination. By continuously attracting monocytes to tissues, p17 helps sustain chronic inflammation, a hallmark of HIV-1 infection. In ECs, p17's interaction with CXCR1 and CXCR2 promotes angiogenesis, a process involved the development of vascular complications associated with chronic HIV-1 infection (Caccuri F. et al., 2012). The presence of p17 in ECs of HIV-1-infected individuals, coupled with its ability to induce capillary formation, suggests that p17 may contribute to the increased prevalence of cardiovascular and lymphatic disorders in these patients (Caccuri F. et al., 2012).

P17 interacts also with heparan sulfate proteoglycans (HSPGs) on the surface of immune cells and ECs. The N-terminal region of p17 contains clusters of basic residues that are essential for binding to the negatively charged sulfate groups of HSPGs (Poiesi C. et al., 2008). This interaction not only increases the local concentration of p17 at the cell surface but also enhances its ability to interact with CXCR1 and p17Rs, being crucial for its pro-inflammatory and pro-angiogenic activities (Bugatti A. et al., 2013). The interaction with HSPGs also plays a role in chronic immune activation, as HSPGs can act as reservoirs for p17, maintaining its presence in the tissue microenvironment long after the initial viral infection has been controlled. In effect, significant amounts of p17 have been detected in patients' lymph nodes even in the absence of active viral replication, suggesting that p17 may continue to exert its biological effects long after the virus suppression by HAART (Popovic M. et al., 2005). This persistence underscores the potential role of p17 as a virokine, a viral protein that mimics host cytokines to modulate the immune response and promote viral persistence. By maintaining a pro-inflammatory microenvironment, p17 could contribute to the maintenance of viral reservoirs and complicating efforts to achieve a complete cure for HIV-1 infection.

In conclusion, p17 is a multifunctional protein that, through its interactions with cellular receptors (p17Rs, CXCR1, CXCR2 and HSPGs) plays a central role in the virus's replication and ability to induce host immune system dysregulation and chronic inflammation contributing to conditions

such as HIV-associated liver fibrosis and AIDS-related vascular disorders. Understanding the mechanisms by which p17 exerts these effects may offer new avenues for therapeutic intervention.

2.1.6.3. P17 as a therapeutic target

The functional epitope of p17 has been identified in the sequence GELDRWEK located within a partially unfolded α-helix forming a loop conserved across different HIV-1 strains, even those with significant mutations, and he neutralization of this epitope can effectively block the detrimental effects of p17 on the immune system (Fiorentini S., Marsico S. et al., 2008; Fiorentini S. et al., 2004).

Therapeutic strategies targeting p17 led to the development of a peptide-based vaccine focused on the AT20 sequence, a 20-aa segment that mimics the functional epitope of p17. The AT20 peptide elicits an immune response capable of neutralizing the p17/p17R interaction, thereby inhibiting the harmful biological activities of p17. Preclinical studies demonstrated that immunization with AT20 conjugated to the carrier protein Keyhole Limpet Hemocyanin (KLH) induced the production of neutralizing antibodies that effectively blocked p17 binding to its receptors, supporting its therapeutic potential (Fiorentini S., Marsico S. et al., 2008). AT20-KLH was further evaluated in a Phase I clinical trial, showing that it was safe, well-tolerated, and capable of inducing a robust anti-p17 antibody response in HAART-treated patients (Iaria M.L. et al., 2014). Anti-AT20 antibodies not only neutralize p17 but also protect human hepatic stellate cells from p17-induced profibrogenic effects, highlighting a potential clinical application in preventing HIV-1-associated liver complications (Renga B. et al., 2014).

Finally, the ability of anti-p17 antibodies to recognize both linear and conformational epitopes enhances their efficacy against HIV-1 variants, making AT20-KLH vaccine promising for inclusion in a multi-epitope therapeutic vaccine strategy aimed at achieving broader HIV-1 neutralization.

There are concerns regarding the possible emergence of mutants capable of evading neutralization by anti-AT20 antibodies, but the highly conserved polybasic region of p17 within the AT20 epitope suggests that mutations in this area would incur a significant fitness cost, limiting the virus's ability to evolve without compromising its functionality (Fiorentini S., Marsico S. et al., 2008).

Looking ahead, structural studies on p17 will allow the identification of specific motifs involved in CXCR1 and CXCR2 binding, which could inform the design of selective drugs that block receptor activation able to target p17 variants (vp17s) that promote B-cell growth and consequent lymphoma development and metastasis (Bugatti A. et al., 2013). HSPGs may also have future clinical applications. While heparin and HS are weak inhibitors of the p17-CXCR1 interaction, highly O-sulfated derivatives of Escherichia coli K5 polysaccharide have shown strong inhibitory activity against both p17/HSPGs and p17/CXCR1 interactions. These derivatives could serve as templates for developing multitarget anti-HIV drugs that act against different receptors and even HIV proteins, offering a novel therapeutic approach for HIV-1 disease management (Bugatti A. et al., 2013).

2.1.6.4. VP17s and Their Role in B-Cell Clonogenic Activity

Recent studies have highlighted that vp17s possess the capability to enhance B-cell clonogenic activity, a critical factor in the development of HIV-associated lymphomas (see paragraph 2.1.5. above. These vp17s, identified in both plasma and lymphoma tissue of HIV+ patients, exhibit specific aa insertions that seem crucial for the promotion of cell survival, proliferation, and oncogenesis (Caccuri F. et al., 2014).

The molecular mechanism underlying clonogenic activity involves destabilization of p17 that leads to conformational changes that enhance its interaction with cell receptors. The insertion of an Ala–Ala stretches at position 117–118 is particularly significant, being able to reproduce the clonogenic effects observed in natural vp17s (Caccuri F. et al., 2014). This destabilization unmasked a functional epitope within the N-terminal region of p17 that is critical for its B-cell growth-

promoting activity. This region interacts with protease-activated receptor 1 (PAR1), leading to the transactivation of epidermal growth factor receptor (EGFR) and subsequent activation of the PI3K/Akt signaling cascade, a pathway commonly implicated in cancer progression (Giagulli C. et al., 2017). The C-terminal region of p17 has been found to be dispensable for its clonogenic activity, but is essential for p17 interaction with PIP2 and its subsequent secretion (Caccuri F. et al., 2016).

vp17s secretion is similar to that of wild-type p17 (refp17). it occurs at the PM and involves cleavage by aspartyl proteases. vp17s are however secreted in an active form that promotes B-cell proliferation, angiogenesis and lymphangiogenesis, crucial for tumor progression (Caccuri F. et al., 2016). vp17s presence in lymphoid tissues suggests a persistent oncogenic stimulus even in the context of effective cART (Totonchy J. et al., 2016). Al, vp17s ability to bind HSPGs and to remain tethered to the cell surface further underscores their role in sustaining B-cell clonogenicity, potentially facilitating cell-to-cell spread of their oncogenic effects (Bugatti A. et al., 2013).

Given their significant role in driving B-cell proliferation and lymphoma development, therapeutic strategies targeting vp17s are crucial. Potential approaches include the development of neutralizing antibodies against the active epitopes of vp17s or the design of therapeutic vaccines that elicit a robust immune response against the (Fiorentini S. et al., 2010). Additionally, targeting the PIP2 binding pocket of vp17s or inhibiting the PAR1/EGFR/PI3K/Akt signaling pathway may offer novel avenues for preventing or treating HIV-related lymphomas (Giagulli C. et al., 2017).

The global prevalence of HIV-1 variants expressing vp17s has shown a significant increase. Vp17s have been identified through comprehensive sequencing studies, which revealed their higher frequency in patients with lymphoma, suggesting their potential involvement in lymphomagenesis. Notably, the most common insertions occur at positions 117-118 and 125-126 of the p17 protein, and their presence has been consistently observed across various lymphoma histotypes, including diffuse large B-cell lymphoma and immunoblastic/plasmablastic lymphoma.

The progressive rise in the frequency of vp17s over time supports the hypothesis that these mutants may be gradually replacing wild-type p17 due to a selective advantage. These vp17s could be indeed able to evade immune surveillance or to enhance viral replication and persistence. Furthermore, the identification of clusters of patients infected with HIV-1 strains expressing clonogenic vp17s underscores the ongoing evolution and adaptation of these viral variants, highlighting the importance of continued genomic surveillance to better understand their implications in HIV-1 pathogenesis and in the development of AIDS-related lymphomas. Early identification of such variants could be crucial for implementing targeted preventive therapeutic to mitigate their long-term impact on cancer risk in HIV-1–infected individuals.

2.1.6.5. Crystal structures of p17

The crystal structure of p17 in both monomeric and trimeric configurations have been resolved (**Fig. 6**). Monomeric p17 is an ordered globular structure stabilized by hydrophobic interactions. It consists of 5 α-helices (H1-H5), with H1 and H5 being at the N- and the C-terminus. H1 and 2 are involved in binding to the PM, facilitated by myristoylation of the N-terminal glycine residue. The hydrophobic nature of H1 mediates membrane insertion, while the basic residues of the H1-2 interface interact electrostatically with negatively charged phospholipids of the inner PM. Crystallographic studies and nuclear magnetic resonance (NMR) analyses revealed that monomeric p17 in solution resembles its crystallized form with only minor variations. H3 is centrally located and interacts hydrophobically with the other helices, particularly H4 and H5, to maintain p17 stability. The loops connecting these helices are flexible, which may be essential during viral assembly.

Trimeric p17, particularly relevant in the context of viral assembly, assembles into 3 different crystal lattices, each burying in a surface area of about 1880 Å² at the trimer interfaces that are highly hydrophilic, involving a network of intermolecular hydrogen (H)-bonds that stabilize the trimer. Trimerization may facilitate the interaction of p17 with the inner viral membrane, possibly

in cooperation with its N-terminal Myr group and the exposed basic residues of the matrix, forming a bipartite membrane-binding surface that is crucial for maintaining the structural integrity of the virus and is involved in organizing the envelope protein on the virion surface.

The fact that the trimeric interface involves residues of H1 and H2 of each monomer suggests that trimerization occurs early during Gag assembly on PM inner surface. Trimer might facilitate Gag lateral interactions promoting PM curvature necessary for virion budding and might help alignment and incorporation of gp120 and gp41 into budding virion, as p17 interacts with their cytoplasmic tails. Mutagenesis studies provided insights into the roles of specific p17's residues. T70E, E74K double substitution at the trimeric interface, such as the, severely impair viral particle production, highlighting their importance in maintaining trimer integrity. The introduction of larger side chains or charged residues at the interface likely disrupts the delicate balance of interactions required for stable trimer formation, leading to defects in viral assembly and a reduction in viral infectivity.

Mutations in C57 and C87, which are buried within the hydrophobic core of the monomeric structure, result in misfolding and aggregation of p17 and significant defects in virus production, indicating their critical role in maintaining the protein structural integrity and particle assembly.

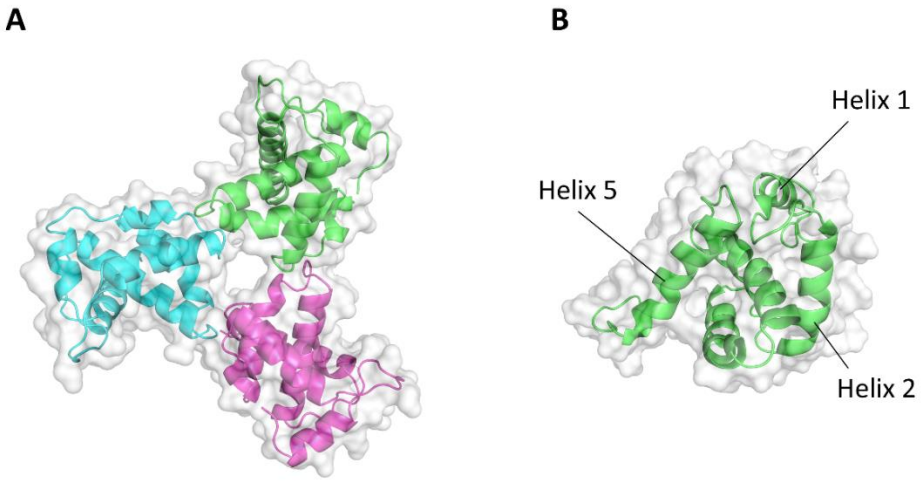

Fig. 6. Schematic representation of trimeric (PDB id: 1HIW) (A) and monomeric (PDB id: 1TAM) (B) HIV-1 p17 structure.

2.2 Computational methods in structural biology: from principles to HIV-1 applications

2.2.1. Computational methods on biological data

Computational methods have become indispensable in modern biological research, particularly in the study of complex diseases such as HIV-1. These approaches enable the exploration of biological systems at an atomic level, providing insights into molecular mechanisms that are crucial for understanding viral behavior and consequent drug discovery. In this context, one of the most powerful computational techniques is genome analysis, which has transformed our ability to study viral genomes, including HIV-1. By means of high-throughput sequencing technologies, researchers can rapidly sequence and compare viral genomes, identifying mutations responsible for drug resistance and revealing evolutionary trends that help predict how the virus may adapt to treatment strategies (Gayathiri E. et al., 2023). Genome-wide association studies and comparative genomics provide deep insights into the genetic factors influencing viral fitness and pathogenicity,

while also identifying conserved regions of the virus that could be targeted by broad-spectrum antivirals or vaccines (Cheifet B., 2019).

Molecular docking is a computational method that predicts the optimal binding mode of a small molecule (ligand) within the active site of a target macromolecule (protein). It is essential in drug discovery because it helps design more effective drugs and identifies new therapeutic candidates by offering detailed information about ligand-receptor interactions. It allows the researchers to virtually screen large compound libraries providing a ranking of the potential binding efficiency of drug-protein interactions thus allowing significantly accelerating the early stages of drug discovery by narrowing down candidates for experimental validation, thereby saving both time and resources (Van Drie J.H., 2007; Salmaso V., 2018).

Molecular Dynamics simulations (MDs) (see paragraph 2.2.1.2. below) is of importance in HIV-1 research since, in respect to molecular docking, that is static, they allow the study of atomic and molecular interactions over time, offering valuable information about the flexibility and conformational changes of viral proteins. In effect, by capturing the dynamic nature of molecular interactions, MDs provide detailed insights into how viral proteins behave in various environments and predict drug binding, interaction stability, and the impact of mutations on drug efficacy (Hollingsworth S.A. et al., 2018).

The integration of MDs with docking has further refined the prediction of protein-ligand interactions by accounting for protein flexibility, which is essential for accurately evaluating the efficacy of drug candidates (Salmaso V. et al., 2018). In the field of HIV-1, since integrase and protease can adopt multiple conformations, the integration of docking and MDs are used to identify the most druggable conformations and predict which drugs candidates are most likely to bind effectively to the 2 viral proteins (Lins R.D. et al., 1999; Pietrucci F. et al., 2015). Also, they have been instrumental in modeling how the active site of HIV-1 protease changes conformation when bound to different inhibitors, aiding in the optimization of drug design to improve binding affinity and reduce the potential for resistance (Perryman A.L. et al., 2004).

By combining genomic and structural biology data, researchers can predict how mutations in HIV-1 genome may affect protein function and drug binding, providing crucial insights for designing drugs that remain effective despite viral mutations (Candresse T. et al., 2014). These approaches complement time consuming experimental techniques, accelerating our ability to respond to the challenges posed by drug-resistant HIV-1 strains. As computational power and techniques continue to advance, the potential for further breakthroughs in HIV-1 research became immense.

2.2.1.1. Molecular docking

Molecular docking involves sampling and scoring. Sampling explores possible conformations of the ligand within the protein's binding site. Scoring ranks these conformations based on their predicted binding affinities. The goal is to find the conformation with the lowest free energy, which represents the most favourable interaction between the ligand and the protein. This is achieved using sophisticated search algorithms (e.g. Genetic algorithm, Monte Carlo algorithm) and scoring functions (e.g. Force-field-based Scoring, Empirical Scoring, Knowledge-based Scoring) that simulate the interaction between the molecules (Agu, P. C. et al., 2023).

Several types of docking approaches have been developed to address the inherent complexity of molecular interactions, each with its unique strengths and limitations. In rigid docking, both the binders (i.e. a ligand and a receptor) are treated as static, inflexible entities throughout the process. This approach simplifies the computational process, making it ideal for high-throughput virtual screening, where large libraries of compounds need to be evaluated quickly. However, the assumption of rigidity can be limiting, especially in biological systems where ligands and receptors often undergo conformational changes upon binding. Also, rigid docking is less effective in cases

where the interaction involves induced fit, as it does not account for the structural adjustments that can occur when a ligand binds to its target receptor (Meng X.Y. et al., 2011).

To address this limitation, flexible docking has been introduced that provides a more realistic representation of binding events by allowing conformational changes either in the ligand, the receptor, or both. In many cases, the ligand is allowed to explore different conformations by adjusting internal degrees of freedom such as bond rotations, while the receptor remains rigid, such as in the docking program AutoDock (Morris G.M. et al., 2009). On the other hand, receptor flexibility involves allowing certain regions of the receptor, particularly near the active site, to change conformation during docking. While this adds computational complexity, it results in a more accurate prediction of binding interactions. To this aim, programs such as GOLD (Verdonk M.L. et al., 2003) permit partial receptor flexibility.

Taking this concept further, the induced-fit docking models has been introduced, where both ligand and receptor undergo conformational changes to better accommodate each other. This approach mimics better the biological reality where binding events are often accompanied by structural adjustments on both sides. Induced-fit docking provides a highly accurate depiction of molecular interactions, though it is computationally expensive due to the need to account for the flexibility of both molecules.

In situations where the exact binding site of the ligand is unknown or multiple potential binding sites exist, blind docking becomes essential. It involves docking the ligand across the entire surface of the receptor, allowing the software to explore potential binding sites without any prior knowledge of where the ligand might bind. However, due to the larger conformational space that must be explored, blind docking requires significantly more computational resources compared to site-specific docking (M. Chaudhary and K. Tyagi, 2024).

Another form of docking is covalent docking (Scarpino A. et al., 2020), which is employed when the ligand forms a covalent bond with the receptor. This process requires different computational treatments because the covalent bond involves sharing electrons between atoms. Covalent docking is more computationally intensive and requires specialized algorithms that must account for the chemical reaction leading to bond formation and its effect on the overall stability of the ligand-receptor complex. GOLD program (Verdonk M.L. et al., 2003) can model these covalent interactions.

Lastly, reverse docking (Kharkar P.S. et al., 2014) focuses on predicting all the potential targets for a ligand. Instead of docking multiple ligands against a single target, it evaluates how one ligand interacts with various receptor structures. This approach is particularly useful in drug repositioning, where existing drugs are tested against new targets to uncover novel therapeutic uses. By docking known drugs against different proteins, researchers can identify previously unknown pathways or molecular mechanisms that the drug may influence, thus facilitating the discovery of new therapeutic applications for old drugs (Chaudhary M. et al., 2024).

Each docking method provides unique insights into molecular interactions, enabling researchers to explore different aspects of ligand-receptor binding and ultimately contribute to drug discovery and drug repositioning processes. Among the various docking tools available, AutoDock Vina (Eberhardt J. et al., 2021) stands out for its efficiency and accuracy in predicting binding affinities, particularly in flexible docking. Vina employs a hybrid scoring function that combines empirical data and knowledge-based elements to predict ligand binding. The optimization is performed using the Broyden-Fletcher-Goldfarb-Shanno algorithm (Shanno David F., 1970), which efficiently minimizes the energy of the system by finding the global minimum of the scoring function. The scoring function used by AutoDock Vina can be described as:

$$\Delta G_{binding} = \Delta G_{vdW} + \Delta G_{H-bond} + \Delta G_{eletrostatic} + \Delta G_{torsional}$$

where ΔG_{vdW} represents van der Waals interactions, $\Delta G_{H\text{-}bond}$ accounts for H-bonds, $\Delta G_{electrostatic}$ considers electrostatic interactions, and $\Delta G_{torsional}$ reflects the ligand torsional strain.

Vina optimizes both ligand and receptor conformations during docking to minimize the binding energy, providing the most favourable binding pose. This approach, which accounts for the flexibility of both molecules, allows accurate predictions of ligand binding in flexible receptors. AutoDock Vina is particularly popular indeed due to its combination of speed and precision, making it particularly effective for high-throughput virtual screening and structure-based drug design (Eberhardt J. et al., 2021) and to prioritize top candidates for experimental validation.

2.2.1.2. Molecular dynamics simulations (MDs)

Biomolecules, including proteins, sugars and nucleic acids are inherently flexible and dynamic. Their biological function is intricately linked to their ability to adopt multiple conformations and undergo rapid structure changes. For instance, enzymes often rely on their capacity to shift between different conformations to bind substrates and catalyze reactions. MDs play an essential role in capturing these structural transitions, providing insight into how molecules behave in their natural biological environments. MDs use computational models to predict the time evolution of atomic movements based on the physical forces acting on each atom in the system. By simulating the real-time dynamics of biomolecules, MDs help bridge the gap between static structural information, such as X-ray crystallography, NMR data and molecular docking itself and the dynamic, fluctuating nature of these molecules in living organisms.

The foundations of MDs lie in the principles of classical Newtonian mechanics: Newton's second law of motion governs atoms' behaviour in a molecular system. MDs' process begins with constructing a computational model of the system based on experimental data. Once the system is prepared, MDs calculate the time evolution of atomic positions by solving Newton's equations of motion for every atom in the system. The general form of Newton's second law for MDs is:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i}$$

where $r_i(t)$ represents the position vector of atom i at time t, $F_i(t)$ is the force acting on atom i at time t, and m_i is the mass of atom i. It shows that the acceleration of an atom is proportional to the force acting on it and inversely proportional to its mass. The forces $F_i(t)$ are derived from the potential energy function $V(r)$, which describes the interactions between atoms in the system. MDs divide time into small intervals called time steps (δt), typically in the order of femtoseconds (1 fs = 10^{-15} seconds). In the δt, atoms' positions and velocities are updated on the the forces acting on them. The integration of Newton's equations of motion is done using algorithms such as the Verlet algorithm, widely used due to its simplicity and stability in MDs. A commonly employed variant is the velocity-Verlet algorithm, which updates both position and velocity as follows:

$$r_i(t+\delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2$$

$$v_i(t+\delta t) = v_i(t) + \frac{1}{2}[a_i(t) + a_i(t+\delta t)]\delta t$$

where, $r_i(t)$ and $v_i(t)$ represent the position and velocity of atom i at time t, while $a_i(t)$ is the acceleration. Velocity-Verlet algorithm ensures both energy conservation and time-reversible dynamics, essential for maintaining the stability of long simulations.

Choosing the appropriate δt is critical in MDs. It must be short enough to capture the fastest atomic motions, such as bond vibrations that occur on the order of femtoseconds. If the δt is too

large, the simulation may introduce significant errors and lead to the accumulation of artefacts in the calculated trajectories. To mitigate this issue, constraints like SHAKE (Kräutler V. et al., 2001) and LINCS (LINear Constraint Solver algorithms (Hess B. et al., 1997) are often used to fix bond lengths involving hydrogen atoms, allowing for longer δt (2 fs or more) without sacrificing accuracy.

In MDs, the interactions between atoms are governed by force fields, which are sets of mathematical functions and parameters designed to approximate the potential energy of the system. The potential energy $V(r)$ is typically decomposed into bonded and non-bonded interactions. The total potential energy U_{tot} is expressed as:

$$U_{tot} = \sum_{bonds} \frac{k_l}{2}(l - l_{eq})^2 + \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{k_\chi}{2}\left[1 + \cos(n_\chi - \delta_\chi)\right] + \sum_{i=1}^{N-1}\sum_{i<1}^{N}\left\{4\,\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right\}$$

where the first 3 terms represent bonded interactions (bond stretching, angle bending, and dihedral torsions), where k_l, k_θ, and k_χ are the respective force constants and $l$eq, θ_{eq}, and δ_χ are the equilibrium bond lengths, angles, and dihedral angles. The non-bonded interactions include van der Waals forces, modeled by the Lennard-Jones potential, and electrostatic interactions, modeled by Coulomb's law. Here, ϵ_{ij} and σ_{ij} represent the Lennard-Jones well depth and atomic radii, respectively, while q_i and q_j are the partial atomic charges of atoms i and j, and r_{ij} is their distance.

Different force fields, such as CHARMM (Chemistry at Harvard Macromolecular Mechanics) (Best R.B. et al., 2012), OPLS-AA (Optimized Potentials for Liquid Simulations, All-Atom) (Robertson MJ et al., 2015), AMBER (Assisted Model Building and Energy Refinement) (Tian C. et al., 2020) and GROMOS (GROningen MOlecular Simulation) parameterize these interactions based on empirical data and quantum mechanical calculations. Each force field is tailored to specific biomolecules, providing a versatile platform for simulating various biological systems (Oostenbrink C. et al., 2004).

In most biological simulations, explicit solvent models are used to simulate the behaviour of water molecules surrounding the biomolecules. In the TIP3P and TIP4P water models each water molecule is represented by a set of atoms with fixed charges. These models are crucial for capturing solvation effects, such as H-bonding and the hydrophobic effect, essential for protein folding, stability, and molecular recognition (Hospital A. et al., 2015).

To account for the finite size of the simulation box, Periodic Boundary Conditions are employed that simulate an infinite system by replicating the simulation box in all directions. When a molecule crosses one boundary of the box, it reappears on the opposite side, effectively eliminating edge effects and providing a realistic system representation. Long-range electrostatic interactions are computed using methods such as the Particle Mesh Ewald technique, which separates the electrostatic potential into short-range and long-range components, improving the efficiency and accuracy of the simulation (Salmaso V., 2018). MDs are also often performed under specific thermodynamic ensembles to control properties such as temperature, pressure, and volume. The most common ensembles are the NVT ensemble (constant number of particles, volume, and temperature) and the NPT ensemble (constant number of particles, pressure, and temperature). These ensembles are maintained using thermostats (e.g., the Berendsen or Langevin thermostats) and barostats (e.g., the Parrinello-Rahman barostat) to simulate realistic experimental conditions. Thermostats control the system's temperature by adjusting the atomic velocities, while barostats maintain constant pressure by scaling the simulation box size (Friedrichs M.S. et al., 2009).

In conclusion, MDs have evolved into a powerful tool for studying the dynamic behaviour of biomolecules at atomic resolution. By using force fields to model interactions, time-stepping algorithms to integrate Newton's equations of motion, and explicit solvent models to capture the environment, MDs provide unparalleled insights into biological systems.

Finally, the more recent advancement of computational techniques, such as GPU acceleration and enhanced sampling methods, make possible to explore larger systems and longer timescales (Salmaso V. et al., 2018).

2.2.1.3. Enhanced sampling techniques in MDs

In MDs, the exploration of a biomolecule's conformational landscape is often limited by the presence of energy barriers that trap the system in local minima, preventing an adequate sampling of relevant conformational states within a feasible computational time. To overcome these limitations, several enhanced sampling techniques have been developed aimed at accelerating transitions between states by modifying the energy landscape or applying external forces, thus enabling the simulation to sample more conformational space efficiently:

Umbrella Sampling (Kästner, J. 2011) is a technique that improves sampling in systems where transitions between states are rare due to high energy barriers. It involves applying a biasing potential, called an "umbrella," to sample specific regions of phase space. By covering these regions with a set of overlapping bias potentials, one can compute the free energy landscape of the system. While effective, umbrella sampling requires prior knowledge of the states to be sampled and can be computationally expensive due to the need for multiple biased simulations.

Replica exchange MDs addresses the issue of sampling by running multiple copies (replicas) of the system at different temperatures simultaneously. Periodically, replicas change configurations based on a Metropolis criterion, allowing the system to escape from local energy minima at higher temperatures and refine the sampling at lower temperatures. This technique is highly effective for systems with rugged energy landscapes but requires significant computational resources as the number of replicas grows with system size (Qi R. et al., 2018).

Metadynamics enhance sampling by building a history-dependent potential that "fills in" regions of the energy landscape as the system explores them. This process encourages the system to escape local minima by gradually reducing the depth of sampled wells, thus facilitating transitions between states. Metadynamics is well-suited for exploring free energy surfaces but can be sensitive to the choice of parameters, such as the height and width of the bias potentials (Zhou R. et al., 2022).

Accelerated MDs (aMDs) (Hamelberg D. et al., 2004) is an advanced simulation technique that enhances the sampling of biomolecular systems by modifying the potential energy surface in real-time. It is designed to overcome the intrinsic limitation of classic MDs, where the system can become trapped in local energy minima for extended periods, resulting in poor sampling of the global conformational space. This limitation is especially problematic for biological systems, where large conformational changes, such as protein folding or ligand binding, often occur over timescales that are beyond the reach of conventional MDs.

The theoretical basis of aMDs involves adding a boost potential to flatten the energy landscape when the system resides in low-energy regions. This reduces the energy barriers between different states, enabling the system to more readily transition between conformational states. The boost potential only affects regions below a specified threshold energy E, while regions above E remain unmodified, preserving the natural high-energy barriers that define the system's overall behaviour. The modified potential energy function in aMDs can be expressed as follows:

$$V^*(r) = \begin{cases} V(r) + \Delta V(r) & if\ (V(r) < E) \\ V(r) & (V(r) \geq E) \end{cases}$$

where $V(r)$ is the system's original potential energy, $V*(r)$ is the modified potential and E is the energy threshold above which the boost potential is not applied. The boost potential $\Delta V(r)$ is defined as:

$$V^*(r) = \frac{\big(E - V(r)\big)^2}{\alpha + (E - V(r))}$$

where α is a tuning parameter that determines the sharpness of the boost potential and controls the degree of modification to the energy landscape. The parameter α controls the strength and smoothness of the boost potential. A well-chosen α ensures that the energy landscape is sufficiently modified to enhance transitions between states without distorting the system's physical properties. Larger α result in smoother potential modifications, while smaller values make the system more prone to artificial transitions. E is crucial because it defines the point below which the boost potential is applied and is typically set above the average potential energy of the system, but not too high, to ensure that only low-energy regions are boosted.

The key concept behind aMDs is that the system will experience a lowered potential energy surface in low-energy regions, which allows it to escape local minima more easily. These minima often correspond to stable, but not necessarily functionally relevant, conformations of the biomolecule. By dynamically lowering the barriers between these states, the system is more likely to explore alternative, potentially biologically relevant conformations.

Since aMDs modifies the natural energy landscape, it is necessary to apply reweighting techniques to recover the true thermodynamic properties of the system. Without reweighting, the modified energy landscape would introduce bias into the simulation results, leading to incorrect predictions of equilibrium properties. The reweighting factor is derived from the Boltzmann distribution and compensates for the artificial boost potential applied during the simulation. For each conformation sampled during the simulation, the reweighting factor is calculated as:

$$\omega(r) = e^{-\beta \Delta V(r)}$$

where β=1/kBT is the inverse thermal energy, kB is the Boltzmann constant, and T is the system temperature. The term $\Delta V(r)$ represents the boost potential applied at each configuration. By applying this reweighting factor to the simulation data, researchers can extract thermodynamically accurate ensemble averages and free energy landscapes, ensuring that the accelerated dynamics do not distort the underlying physical properties of the system.

One of the advantages of aMDs over the other techniques mentioned above is that it does not require prior knowledge of the energy barriers or transition pathways. Indeed, aMDs dynamically modifies the energy landscape as the simulation progresses, making it particularly well-suited for systems where the conformational landscape is unknow. Also, in respect to Replica exchange MDs, which requires running multiple replicas at different temperatures, aMDs modifies the energy landscape within a single simulation, reducing the computational overhead. aMDs has been applied successfully to a wide range of biological systems such as the study of protein conformational changes, protein-ligand binding, membrane dynamic and allosteric regulation in enzymes. In drug discovery, where classic MDs struggle to capture the full binding and unbinding events of a ligand to its target protein within feasible timescales, aMDs offer valuable insights into binding mechanisms, intermediate states, and the thermodynamics of ligand binding.

Despite its advantages, aMDs is not without challenges. The selection of appropriate parameters for the boost potential, particularly E and boost factor α, can be difficult, requiring trial and error. If these parameters are not chosen carefully, the system may experience artificial conformational transitions that do not correspond to real biological behaviour, requiring careful analysis and reweighting to ensure accurate results.

2.2.1.4. MDs studies on p17

MDs have already provided a deeper insight into the conformational flexibility and functional behaviour of p17, particularly for to its role in viral assembly and membrane interactions.

A 40 ns MDs of monomeric p17 revealed that p17's conformation in solution deviates from its crystallized form. The most notable structural change occurred in C-terminal H5, which showed partial unfolding during the simulation, losing 2 helical turns in the crystal structure. This suggests that p17 adopts a more flexible conformation in solution, especially in its C-terminal region, which could facilitate interactions with other viral proteins during virion assembly (Verli H. et al., 2007). When comparing NMR solution structure to X-ray structure, a 6 Å displacement in the short 310 helix was revealed that is functionally relevant for virion assembly. These structural rearrangements suggest that p17 undergoes conformational adjustments when transitioning from its monomeric state to its role in viral assembly. This is consistent with the observation of a reduced gyration radius during the simulation, that indicates how p17 becomes more compact by adopting different states. MDs also provided important insights into the stability of the hydrophobic core of p17 (H1–H4): it remained relatively stable while the partially exposed C-terminal H5 exhibited significant flexibility during the MD trajectory that is thought to contribute to p17's ability to interact with other components of the viral assembly machinery (Verli H. et al., 2007).

An all-atom MDs explored how p17 interacts with membranes containing phosphatidyl-serine (PS) and PIP2 (both abundant in PM and essential for viral assembly) revealing that p17 Myr tail insertion in PM is a key step that stabilizes p17-membrane interaction. Also, p17 Myr group dynamically re-enters the lipid bilayer, being strongly influenced by PIP2 and PS. This lipid reorganization may prepare the PM for Gag recruitment, facilitating viral assembly. MDs demonstrated that p17 remains bound to the membrane even without Myr insertion, due to strong electrostatic interactions between the highly basic region of p17 and the negatively charged phospholipids of the membrane and demonstrated that Myr insertion occurs only when p17 is in an "open" conformation, in which H1 and the highly basic region interact with the PM. In this conformation, the Myr tail exits the hydrophobic cavity of the protein and inserts into the lipid bilayer, aligning itself with the lipid tails in the membrane core (Monje-Galvan V. et al., 2020). Importantly, this insertion process was observed only in membranes that mimicked the inner leaflet of the PM. In contrast, in the "blocked" conformation of p17, the Myr tail is sequestered inside its hydrophobic cavity and its insertion did not occur, suggesting that p17 conformational state is crucial in regulating its interactions with PM, with the "open" state being the functional form.

The results of these MDs will help at understanding the role of p17 in viral assembly. Its ability to switch between different conformations may be essential for its binding to components of the viral assembly machinery (i.e. p24, gp120 and gp41), and also underscores the importance of PM composition in virion assembly and budding.

3. AIMS OF THE RESEARCH

The main goal of my PhD research is to exploit state of the art computational techniques to investigate the molecular mechanisms underlying the oncogenic activity of vp17s that have been implicated in the pathogenesis of HIV-related lymphomas. To this aim, the work has been organized in 3 phases (**Fig. 7**) and 2 primary objectives described here below:

Sequential Analysis of Polymerase Stalling and Mutation Hotspots. This first objective is aimed at elucidating how HIV-1 RT stumbles during replication at specific regions of the viral genome, particularly in the C-terminal domain of p17. This region has been already identified as a "hot spot" for mutations, due to the presence of repetitive elements, palindromic sequences, and a high concentration of A that induce polymerase slippage and pausing, thus leading to the introduction of nucleotide insertions or deletions that result in the emergence of vp17s. These variants have been demonstrated to be more prevalent in HIV-1-infected patients with lymphoma, suggesting a direct link between polymerase stalling and the development of oncogenic variants. By conducting detailed computational analyses, in this objective I will map the specific mutation patterns generated by RT stalling. This mapping will allow for a deeper understanding of how the viral genome exploits these mutational "hot spots" to increase genetic diversity, thereby promoting the evolution of vp17s with the emergence of those with oncogenic properties.

Structural Analysis of Conformational Changes in vp17s. This second objective is closely related to the first, being it focused on understanding the structural mechanisms that confer clonogenic properties to vp17s. Unlike the wild-type p17 protein, which maintains a stable and compact globular structure, vp17s undergo significant conformational changes that alter their biological function. These variants are characterized by destabilization of the hydrophobic core, particularly in the COOH-terminal region, that cause misfolding. This misfolding exposes previously masked functional epitopes, allowing the protein to interact with cellular receptors such as the PAR1, which in turn triggers B-cell proliferation. This second part of the work will employ computational modeling techniques such as MDs and aMDs to clarify how misfolded vp17s achieve their gain-of-function behaviour.

In conclusion, the appropriate integration of molecular biology and computational modeling here exploited is expected to provide a comprehensive understanding of how p17 clonogenic variants arise, evolve and function, paving the way for novel therapies targeting the clonogenic activity of vp17s in HIV-associated lymphomas, hopefully improving outcomes for patients.

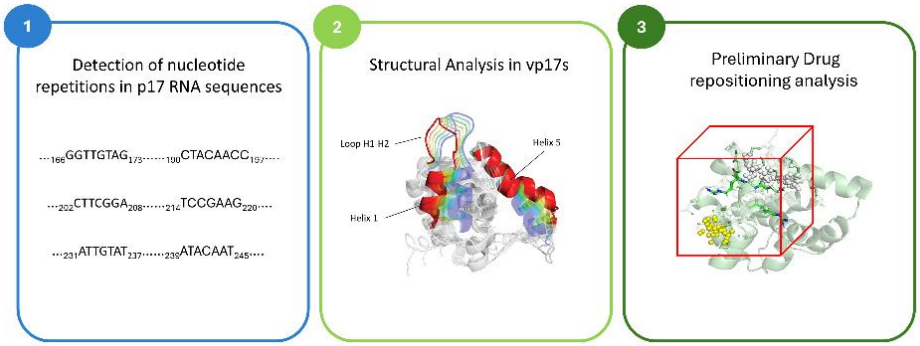

Fig. 7. Schematic representation of the activities addressed in my PhD project. *Starting with a sequential analysis aimed at identifying evidence of HIV RT stalling, the project then moves on to structural analysis and the effects of mutations in vp17s. Based on the obtained results, the project then focuses on drug repositioning to identify potential inhibitory molecules that bind to vp17s.*

4. RESULTS AND DISCUSSION

To pursue the aims of my doctorate research, I collaborated with the Section of Microbiology at the Department of Molecular and Translational Medicine of the University of Brescia setting up an integrated approach involving both computational and experimental studies to evaluate the role of polymerase stalling in generating mutations in p17 gene and the structural consequences of these alterations in vp17s. Sequential analysis of the polymerase slippage events coupled to next-generation sequencing, lead to the mapping of the mutation patterns in p17 C-terminal region, where most insertions occur. In parallel, the structural implications of these mutations on B-cell proliferation were explored by using MDs and aMDs to model how specific residues influence the stability and misfolding of vp17s. The computational work was complemented by in vitro experiments conducted to assess the consequences of these mutations on B-cell clonogenicity. The detailed findings of these studies are presented and discussed in the following chapters.

4.1. Mechanisms of Genomic Instability in Driving Evolution of HIV-1 vp17s

The study of the molecular mechanisms underlying the genomic instability of HIV-1 that drive the generation of vp17s offers valuable opportunities to understand the mutational and structural processes that contribute to their oncogenic potential. The identification of mutational "hot spots" is critical due to their known impact on genome stability (Zou X. et al., 2017) and the study of the conformational changes that drive the pathogenic behaviour of these variants are essential for developing targeted therapeutic strategies against HIV-associated lymphomas.

4.1.1. Identification and Analysis of Repetitive Sequences in vp17s

The results from this work have been collected and published in Zani A, Messali S, Bugatti A, Uggeri M, Rondina A, Sclavi L, Caccuri F, Caruso A. Molecular mechanisms behind the generation of pro-oncogenic HIV-1 matrix protein p17 variants. J Gen Virol. 2024 Apr;105(4).

Background: HIV-1 RT is error-prone and repetitive nucleotide sequences can exacerbate mutation rates. Repetitive elements (i.e. direct or inverted repeats and palindromic sequences) add to the genomic instability by promoting insertions and deletions. These sequences are critical factors that influence the evolution of the HIV-1 genome, especially in regions with high recombination rates. In p17 COOH-terminus, sequences rich in adenine (A) are known to induce polymerase slippage during reverse transcription, leading to high mutation rates.

Thus, the study of p17 gene is crucial to understand the molecular mechanisms driving vp17s generation. Mutations in p17 COOH-terminal region have been observed predominantly in patients with HIV-1 and lymphoma. This part of my research aims at the systematic identification repetitive sequences in the p17 gene and assess their potential role in generating the observed mutations and insertions in vp17.

Methods: a Custom Python script was developed to analyze repetitive sequences occurring within 50 base pairs of each other in refp17 (accession number: NC_001802) and vp17s included in a database generated from 153 sequences derived from a cohort of viremic HIV-1 subtype B–infected patients with or without lymphoma (GenBank Project: PRJEB53498).

Multiple sequence alignment was performed on the p17 sequences using the Geneious software (version 11.1.5, Biomatters Ltd, New Zealand). The MAFFT alignment method was employed with default parameters: G-INS-i algorithm, 200PAM/k=2 scoring matrix, gap open penalty set at 1.53 and an offset value of 0.123. refp17 was used as reference sequence to assess

nucleotide composition and characterize mutations across the 3 open reading frames of p17 gene. Mutation rate in nucleotide sequences was subsequently calculated using a custom in-house script, enabling the systematic characterization of mutations and insertions across the database of vp17s sequences, focusing particularly on regions prone to genomic instability and nucleotide changes.

Results and Conclusions: The analysis revealed 7 short direct repeats, 2 inverted repeats, and 4 palindromic sequences within refp17 (**Tab. 1**). Notably, 3 of the direct repeats, along with 2 inverted repeats and 3 palindromic sequences were located in the second half of the sequence, after nucleotide position 200. While the identified direct repeats are distributed throughout p17 sequence, inverted repeats, such as CTCTAT at 253–258 and ATAGAG at 274–279 and 310–315, are predominantly located closer to the COOH-terminal region.

Type of sequence	Nucleotide sequence	Position from-to	Nucleotide sequence	Position from-to
Direct repeats	GGGGGA	28-33	GGGGGA	70-75
	GGGAAA	47-52	GGGAAA	72-77
	GAAAAAAT	49-56	GAAAAAAT	78-85
	ATCAGAAG	159-166	ATCAGAAG	213-220
	GATAGAG	273-279	GATAGAG	309-315
	AGGAAG	293-298	AGGAAG	314-319
	CACAGCAA	344-351	CACAGCAA	370-377
Inverted repeats	CTCTAT	253-258	ATAGAG	274-279 & 310-315
	GCTGAC	358-363	GTCAGC	382-387
Palindrome	ATCGAT	41-46	/	/
	ATTATATAAT	231-240		
	AAGCTT	296-301		
	CAGCTG	356-361		

Tab. 1. Direct repeats, inverted repeats, and palindromic sequences found in refp17.

The A-rich segment spanning nucleotides 322–342, which contains 76% A (322CAAAACAAA-AGTAAGAAAAAA342) is particularly interesting because it is known from literature that similar sequences induce polymerase slippage, leading to errors during DNA replication or RNA transcription (Lovett S.T., 2004).

To further explore the impact of the repeated sequences, nucleotide sequences of the 153 vp17s were aligned and analyzed showing that insertions predominantly occur in p17 COOH-terminal region, mainly between nucleotides 342-385. Insertion hot spots were identified, with the highest frequencies at positions 351–352 (21.60%) and 375–376 (14.10%) (**Tab. 2**).

Also, 65 out of 72 insertions are associated with triplet nucleotide duplications, categorized as perfect, imperfect, or partial/split repetitions occurring before or after the insertion sites (**Tab. 3, 4 and 5**), suggesting that they facilitate genomic alterations.

Insertion position	% insertion
342-343	5.20% (8/153)
351-352	21.60% (33/153)
366-367	4.60% (7/153)
375-376	14.10% (19/153)
378-379	0.70% (1/153)
381-382	2.00% (3/153)
384-385	0.70% (1/153)

Tab. 2. Nucleotide position of insertions in the 153 p17 sequences.

Perfect duplications (i.e. GCACAGCAA) were found at positions 342–343, with similar patterns observed in other insertion types. Despite the presence of direct repeats in refp17 (see **Tab. 1** above), these repeats do not appear to play a direct role in generating insertions in vp17, as most insertions occurred outside these repeat regions. However, a significant proportion of inverted repeats and palindromic sequences in the COOH-terminal region (i.e. 358GCTGAC363 and 296AAGCTT301) showed high conservation across the sequences. Also, 44 novel inverted repeats and 31 novel palindromic sequences were identified that showed however lower conservation. The high conservation of certain sequences (i.e. the 55.60% recurrence of the GCTGAC repeat) highlights their potential in viral genome evolution or adaptation mechanisms (**Tab. 6**). Statistical analysis of mutations within these regions revealed a prevalence of A>G and G>A transitions, particularly in A-rich sequences (**Tab. 7**). Notably, A342 showed the highest mutation rate, with 58.00% of sequences exhibiting changes with the predominant substitution being A>G (98.00%). This pattern is consistent with previous findings in the HIV-1 genome, where A>G mutations in AA or TA have been linked to the formation of hairpin structures (Sun N. et al., 2022). Given the abundance of A-rich regions in refp17, this correlation likely contributes to the genomic instability observed in this part of the sequence.

Perfect Single Repetition					
Sample ID	**Insertion position**	**Inserted nucleotides**	**Inserted aa**	**p17 nucleotides**	**p17 aa**
C62	342-343	GCACAGCAA	AQQ	GCACAGCAA	AQQ
C03	342-343	GCACAGCAAGCA	AQQA	GCACAGCAAGCA	AQQA
L58	351-352	GCA	A	GCA	A
L67	351-352	GCA	A	GCA	A
L47	351-352	GCA	A	GCA	A
C10	351-352	GCA	A	GCA	A
C22	351-352	GCA	A	GCA	A
L36	351-352	GCAGCA	AA	GCAGCA	AA
L11	351-352	GCAGCA	AA	GCAGCA	AA
C36	351-352	GCAGCA	AA	GCAGCA	AA
L66	351-352	GCAGCA	AA	GCAGCA	AA
L50	351-352	GCAGCA	AA	GCAGCA	AA
L15	351-352	GCAGCAGCT	AAA	GCAGCAGCT	AAA
C57	351-352	GCAGCAGCT	AAA	GCAGCAGCT	AAA
L18	351-352	GCAGCAGCT	AAA	GCAGCAGCT	AAA
L21	351-352	GCAGCAGCT	AAA	GCAGCAGCT	AAA
L42	351-352	GCACAGCAA	AQQ	GCACAGCAA	AQQ
L20	375-376	AGC	A	AGC	A
L62	375-376	AGC	A	AGC	A
C26	384-385	GTC	V	GTC	V
Perfect Multiple Repetition					
Sample ID	**Insertion position**	**Inserted nucleotides**	**Inserted aa**	**p17 nucleotides**	**p17 aa**
C25	351-352	GCAGCA	AA	GCA	A
L77	351-352	GCAGCAGCA	AAA	GCA	A
L31	351-352	GCAGCAGCAGCA	AAAA	GCAGCA	AA
C60	351-352	GCAGCAGCAGCA	AAAA	GCAGCA	AA
L21	375-376	AGCAGC	SS	AGC	S
C61	375-376	AGCAGC	SS	AGC	S

Tab. 3. Perfect single or multiple repetition in refp17 sequences from our database.

Imperfect Repetition (1 mismatch)			
Sample ID	**Insertion position**	**Nucleotide sequence**	**aa sequence**
L14	342-343	GCACAAGCACAG	AQAQ
L74	342-343	GCACAGCAGGCACAGCAA	AQQAQQ
L64	342-343	GTACAGGCACAG	VQAQ
L28	351-352	GCAACAGCAGCA	ATAA
C61	351-352	GCGGCAGCAGCA	AAAA
C38	351-352	ACAGCAGCAGCA	TAAA
L73	351-352	GCAACAGCAGCA	ATAA
L10	351-352	GCGGCAGCAGCA	AAAA
L19	351-352	GCAGCAGCCGCAGCAGCT	AAAAAA
L02	351-352	GCAGCAGCAGCAGCAGCT	AAAAAA
L75	351-352	GCAGCAGCAGCAGCAGCT	AAAAAA
C13	351-352	GCAGCAGCTGGCACAGGAGCAGCAGCTGACACAGGA	AAAGTGAAAGTG
L23	366-367	GACACAGGCACA	DTDT
L47	366-367	GGAAACAGCGGACACAGC	GNSGHS
C42	375-376	ACAGGAAACAGCACAGGAAACAAC	TGNNTGNN
L41	378-379	AATAAA	NK
Imperfect Repetition (2 mismatches)			
Sample ID	**Insertion position**	**Nucleotide sequence**	**aa sequence**
C07	342-343	AAAAAAGAAAAG	KKEK
C02	342-343	AAAAAAGAAAAG	KKEK
C06	342-343	AAAAAAGAAAAG	KKEK
L71	375-376	AGCCAGGTCAATCAGGTC	SQVNQV
L61	375-376	AGCCAGGTCAATCAGGTC	SQVNQV
L45	375-376	AGCCAGGTCAATCAGGTC	SQVNQV
C55	375-376	AGTCAGGCCAGCAATCAGGTCAGC	SQASNQVS
L27	375-376	CACAGCGGCAGC	HSGS
L55	375-376	CACAGCAAAAGC	HSKS

***Tab. 4. Imperfect single or multiple repetition in refp17 sequences from our database**. The inserted nucleotides and aa are in red. The ones that have been duplicated are in black.*

Partial/Split Repetition			
Sample ID	**Insertion position**	**Nucleotide sequence**	**aa sequence**
C33	366-367	GGAAATAGCGGCCAAGGACACAGCAATCAG	GNSGQGHSNQ
L54	366-367	ACAGGAAACAGCAGCACAGGACACAGCAAT	TGNSSTGHSN
L82	351-352	CAAGCGCAAGCA	QAQA
C67	351-352	CAAGCAGCGCAAGCAGCA	QAAQAA
L33	351-352	GCAGCGGCAGCTGGCGCAGCAGCT	AAAAGAAA
C11	351-352	GCGGCCACTGGCGCAGCAGCTGAC	AATGAAAD
C55	351-352	CAAGCAGCAGCTGGCCAAGCAGCAGCTGAC	QAAAGQAAAD
L37	375-376	AGCCAGGCCAATCAGGTC	SQANQV
C46	375-376	AGCCAGGCCAGCAATCAGGTCAGC	SQASNQVS
C50	375-376	AGCAGCCAGGCCAGCAGCAATCAGGTCAGC	SSQASSNQVS
L65	375-376	ACAGGACACAGCAGCCAAGCAGGAAGTAGCAATCAG	TGHSSQAGSSNQ
L05	375-376	GGACACAGCAGCCAGGCCGGAAACAGCAATCAGGTC	GHSSQAGNSNQV
L23	381-382	CAGGTCAGCCAAAATCAGGTCAGCCAAAAT	QVSQNQVSQN
L49	381-382	AATCAGGCCAGCCAAAATAGCCAGGTCAGCCAAAAT	NQASQNSQVSQN

***Tab. 5. Partial/split repetition occurring in refp17 sequences from our database**. In red are highlighted the inserted nucleotides and aa, in black those that have been duplicated.*

Type of Sequence	refp17/ New	Sequence	Position	Sequence	Position	Total	%
Inverted	New	CGAATT	33-38	AATTCG	54-59	3	5.20
	New	TTATAT	37-42	ATATAA	84-89	1	1.70
	New	ACAATAT	81-87	ATATTGT	98-104	1	1.30
	New	AAACTA	88-93	TAGTTT	101-106	1	1.30
	New	AAACTAAA	88-95	TTTAGTTT	99-106	1	1.30
	New	GATTAA	89-94	TTAATC	137-142	5	6.50
	New	ACTAAA	90-95	TTTAGT	99-104	2	2.60
	New	ATTAAA	90-95	TTTAAT	99-104	1	1.30
	New	TCTAATA	99-105	TATTAGA	149-155	1	1.30
	New	AGCAGG	112-117	CCTGCT	147-152	2	2.60
	New	ATTTGC	129-134	GCAAAT	174-179	1	1.30
	New	CTATTA	148-153	TAATAG	179-184	1	1.30
	New	CCAGAA	160-165	TTCTGG	179-184	1	1.30
	New	GAAGGCT	163-169	AGCCTTC	199-205	3	3.80
	New	GAAGGG	163-168	CCCTTC	200-205	1	1.30
	New	AGGGCTG	164-170	CAGCCCT	197-203	2	2.50
	New	GGCTGTA	166-172	TACAGCC	191-197	2	2.50
	New	GGCTGTAG	166-173	CTACAGCC	190-197	10	12.50
	New	GGCTGTAGA	166-174	TCTACAGCC	189-197	2	2.50
	New	GGTTGTAG	166-173	CTACAACC	190-197	4	5.00
	New	GCTGTA	167-172	TACAGC	185-190	1	1.30
	New	AAATAAT	176-182	ATTATTT	231-237	1	1.20
	New	AAATAATG	176-183	CATTATTT	230-237	1	1.20
	New	ATAATG	178-183	CATTAT	230-235	1	1.20
	New	ATACTG	178-183	CAGTAT	230-235	1	1.20
	New	ATCCTG	178-183	CAGGAT	209-214	1	1.20
	New	CTTCAGA	202-208	TCTGAAG	214-220	1	1.20
	New	AGGATC	210-215	GATCCT	227-232	1	1.20
	New	AGAGGA	216-221	TCCTCT	251-256	1	1.20
	New	AAGAACT	218-224	AGTTCTT	249-255	1	1.20
	New	TTAAAT	224-229	ATTTAA	234-239	18	22.20
	New	ATCCTT	228-233	AAGGAT	270-275	1	1.20
	New	CATTAT	230-235	ATAATG	236-241	4	4.90
	New	TTACAT	232-237	ATGTAA	278-283	1	1.20
	New	TGTTTA	233-238	TAAACA	267-272	1	1.20
	New	ATAATA	236-241	TATTAT	256-261	1	1.20
	New	ACAATAG	241-247	CTATTGT	255-261	15	18.50
	New	CAATAG	242-247	CTATTG	255-260	4	4.90
	refp17-like	CCTCTAT	252-258	ATAGAGG	274-280	3	3.70
	refp17-like	CCTCTAT	252-258	ATAGAGG	274-280 & 310-316	10	12.30
	refp17-like	CCTCTATT	252-259	AATAGAGG	273-280 & 309-316	1	1.20
	refp17-like	CTCTAT	253-258	ATAGAG	274-279	1	1.20
	refp17	**CTCTAT**	**253-258**	**ATAGAG**	**274-279 & 310-315**	**11**	**13.60**
	refp17-like	CTCTATT	253-259	AATAGAG	273-279	3	3.70
	refp17-like	TCTATT	254-259	AATAGA	273-278	2	2.50
	refp17-like	TCTATT	254-259	AATAGA	273-278 & 309-314	3	3.70
	New	TCTTTT	254-259	AAAAGA	282-287	1	1.20
	New	TACATC	263-268	GATGTA	277-282	13	16.00
	New	GCTGCT	355-360	AGCAGC	370-375	2	2.50
	New	ACCTGAC	357-363	GTCAGGT	377-383	1	1.20

	New	GGCTGAC	357-363	GTCAGCC	382-388	1	1.20
	refp17-like	ACTGAC	358-363	GTCAGT	382-387	1	1.20
	refp17	**GCTGAC**	**358-363**	**GTCAGC**	382-387	**45**	**55.60**
	refp17-like	GCTGACA	358-364	TGTCAGC	381-387	2	2.50
	refp17-like	GCTGACC	358-364	GGTCAGC	381-387	1	1.20
	New	GCTGCTAC	358-365	GTAGCAGC	371-378	1	1.20
	refp17-like	GCTGGC	358-363	GCCAGC	382-387	2	2.50
Palin-drome	New	AATATT	18-23	/	/	5	10.40
	New	AATATT	33-38			1	1.70
	refp17	**ATCGAT**	**41-46**			**7**	**11.10**
	refp17-like	ATGCAT	41-46			1	1.60
	New	GAATTC	53-58			6	9.00
	New	CCCGGG	67-72			1	1.40
	New	GATATC	83-88			1	1.30
	New	ACATGT	96-101			3	3.90
	New	TTGCAA	131-136			1	1.30
	New	ATTAAT	136-141			3	3.80
	New	GTTAAC	136-141			12	15.20
	New	GCTAGC	142-147			1	1.30
	New	CAGCTG	158-163			1	1.30
	New	CTGCAG	168-173			5	6.30
	New	TGTATACA	169-176			1	1.30
	New	AAATTT	176-181			1	1.20
	New	AATATT	177-182			16	19.80
	New	TACTAGTA	179-186			1	1.20
	New	TTGCAA	181-186			1	1.20
	New	CTGCAG	190-195			1	1.20
	New	CTTAAG	202-207			3	3.70
	New	GGATCC	211-216			1	1.20
	New	ACTTAAGT	222-229			2	2.50
	New	CTTAAG	223-228			1	1.20
	refp17-like	CATTATATAATG	230-241			4	4.90
	refp17	**ATTATATAAT**	**231-240**			**39**	**48.10**
	refp17-like	TTATATAA	232-239			3	3.70
	refp17-like	TATATA	233-238			4	4.90
	New	AGTACT	249-254			1	1.20
	New	TGTACA	261-266			28	34.60
	refp17	**AAGCTT**	**296-301**			**73**	**90.10**
	New	TCTAGA	300-305			1	1.20
	New	ATATAT	303-308			1	1.20
	New	CTTAAG	331-336			2	2.50
	New	AGCAGCTGCT	351-360			2	2.50
	New	AGCAGCTGCT	354-363			1	1.20
	refp17	**CAGCTG**	**356-361**			**62**	**76.50**
	New	AAAATTTT	389-396			2	2.50
	New	AAATTT	390-395			3	3.70

Tab. 6. Inverted repeats and palindromic sequences found in the 81 p17s without insertion from our database*. Each position is referred to refp17 sequence. The data related to refp17 are highlighted in bold.*

Nucleotide	% of point mutations	% of mutation X to A	% of mutation X to T	% of mutation X to G	% of mutation X to C
A	6.9%	/	0.74%	5.15%	0.99%
T	4.9%	0.42%	/	0.75%	3.78%
G	7.8%	6.10%	1.00%	/	0.66%
C	8.0%	3.85%	3.53%	0.62%	/

Tab. 7. Nucleotide substitution analysis of the 153 sequences of p17 taken in consideration.

These results suggest that repeated sequences, especially those rich in A, play a significant role in promoting the high mutation rates seen in p17, potentially facilitating structural changes impacting virus's evolution and adaptability.

4.1.2. Genomic Instability and Recombination Events in the Generation of Oncogenic HIV-1 vp17s

The results from this work have been collected and published in Zani A, Messali S, Bugatti A, Uggeri M, Rondina A, Sclavi L, Caccuri F, Caruso A. Molecular mechanisms behind the generation of pro-oncogenic HIV-1 matrix protein p17 variants. J Gen Virol. 2024 Apr;105(4).

Background: In my initial research I've characterized repetitive sequences and their role in creating mutation hotspots within p17 COOH-terminal region. Since these mutation-prone areas have been linked to the development of oncogenic vp17, in this part we tested experimentally how these mutations and recombinations occur in a cellular context and their consequences.

Methods: recombinant plasmids based on a patient-derived HIV-1 Gag sequence (pNL4.3C63) and the refp17 construct were designed to introduce specific nucleotide changes in the highly variable p17 COOH-terminus (114-130). RNA was synthesized from both the refp17 and variant constructs, and reverse transcription was performed by HIV-1 or Moloney murine leukemia virus RT to compare their fidelity and mutation-inducing capacities. Then, 3 different human cell lines [H9 (T-cell), U937 (monocyte), and THP1 (macrophage-differentiated)] were nucleofected with refp17, pNL4.3C63 variant, or a combination of both plasmids. Cells were then allowed to undergo viral replication, and viral products were harvested to assess the occurrence of recombination events. Next-generation sequencing was conducted on the viral products to analyze recombination events, particularly at known mutation hotspots in vp17s (e.g., 321 G>A, 342 G>A, 355 G>A, 374 A>G, 387 C>T).

Results and Conclusions: The experimental results confirmed that recombination between refp17 and pNL4.3C63 constructs occurred in all cell lines, leading to of vp17s generation. Chimeric molecules were observed in 18.6% of sequencing reads, that contained a mix of refp17 and variant segments. Recombination events were most observed at specific insertion points at aa 125-126, a characteristic feature of vp17s associated with lymphoma in patients. Moreover, the analysis showed that both HIV-1 RT and Moloney murine leukemia virus RT induced errors at the same A-rich mutation hotspots, confirming that the COOH-terminal region's genomic instability is not solely a consequence of HIV-1 RT's errors also driven by intrinsic sequence features. The presence of inverted repeats and palindromic structures was identified as key contributors to RT stalling, facilitating the recombination events that give rise to pro-oncogenic vp17s.

4.2. Molecular Insights into p17 Variant Evolution: Hydrophobic Core and Functional Shifts

4.2.1 Structural Assessment of vp17s

The results from this work have been collected and published in D'Ursi P, Rondina A, Zani A, Uggeri M, Messali S, Caruso A, Caccuri F. Molecular Mechanisms Involved in the B Cell Growth and Clonogenic Activity of HIV-1 Matrix Protein p17 Variants. Viruses. 2024 Jun 28;16(7):1048.

Background: Structural stability and functionality of refp17, particularly in its variants, are linked to alterations in the hydrophobic core. On the other hand, vp17s, especially those from HIV+ patients with lymphomas, exhibit structural changes due to insertions and mutations, particularly in their C-terminus that can drastically affect the protein's overall stability. Previous studies demonstrated that refp17 misfolding is necessary to expose a functional epitope involved in B-cell clonogenicity. Misfolding is facilitated by changes in secondary-structure, particularly within the hydrophobic core, that influence vp17s capacity of to promote interactions with host cell receptors, contributing to altered cellular activity. Structural modeling and computational simulations can provide useful insights into how mutations and insertions drive functional changes in these variants.

Methods: The 3D structure of refp17 was examined to identify clusters of hydrophobic residues, in particular isoleucine, leucine, and valine (ILV clusters), known to stabilize protein structures. Three ILV clusters forming refp17 hydrophobic core were visualized using molecular visualization software. To quantify hydrophobicity of refp17 and vp17s, Kyte and Doolittle hydropathy index was employed to measure hydrophobicity of individual aa. A comparative analysis of refp17 and vp17s was performed focusing on the H3 and H4 helices and the C-terminal regions. The potential impact of aa insertions on the hydrophobicity of these regions was assessed to identify any structural alterations that could correlate with the functional differences observed in the vp17.

Results and Conclusions: Upon examining refp17, 3 distinct ILV clusters were identified that serve as hydrophobic core of the protein, supporting its structural integrity (**Tab. 8**).

Cluster 1			
refp17	**NHL-a101**	**NHL-a104**	**vp17c2**
Leu8	/	/	/
Leu13	/	/	/
Leu31	/	/	/
Ile34	/	/	/
Val35	/	/	/
Leu85	/	/	/
Cluster 2			
refp17	**NHL-a101**	**NHL-a104**	**vp17c2**
Ile19	/	/	/
Val88	/	/	/
Ile92	/	/	/
Ile94	Val94	/	/
Cluster 3			
refp17	**NHL-a101**	**NHL-a104**	**vp17c2**
Leu41	/	/	/
Leu50	/	/	/
Ile60	/	/	/
Leu61	/	/	Ile61
Leu64	/	/	/
Leu68	/	/	/
Leu75	/	/	/
Leu78	/	/	/
Val82	Ile82	/	Met82

Ile104	/	/	/

Tab. 8. Mapping Hydrophobic Clusters*. Residues belonging to each HC (cluster 1, cluster 2, and cluster 3) in refp17 and in vp17s are shown. Mutated residues in vp17s are also indicated.*

We then extended this analysis to vp17s that, in addition to other aa mutations, carry specific insertions. These insertions include a glutamic acid (E) and lysine (K) insertion at 114–115 in vp17c2, and a double alanine (Ala-ala) insertion at 117–118 in variants NHL-a101 and NHL-a104. The results indicated that these mutations and insertions had no significant effect on the ILV clusters since these regions remained largely unchanged (**Tab. 8**).

We then performed a more detailed comparison of the hydrophobic profiles between refp17 and vp17s considering both mutations and insertions: refp17 exhibited a slight increase in hydrophobicity at the beginning of the H3 helix in respect to vp17s NHL-a101 and NHL-a104. Also, refp17 showed a lower hydrophobicity in the terminal portion of the H4 helix when compared to both NHLa-104 and vp17c2, further confirming that the overall impact of the mutations in vp17s is minimal. A more pronounced increase in the hydrophobic profile was observed in the C-terminal region of NHL-a101 and NHL-a104, where vp17c2 exhibited instead a marked decrease. This hydrophobic shift appears to be directly influenced by the double aa insertions found in these vp17s (**Fig. 8**).

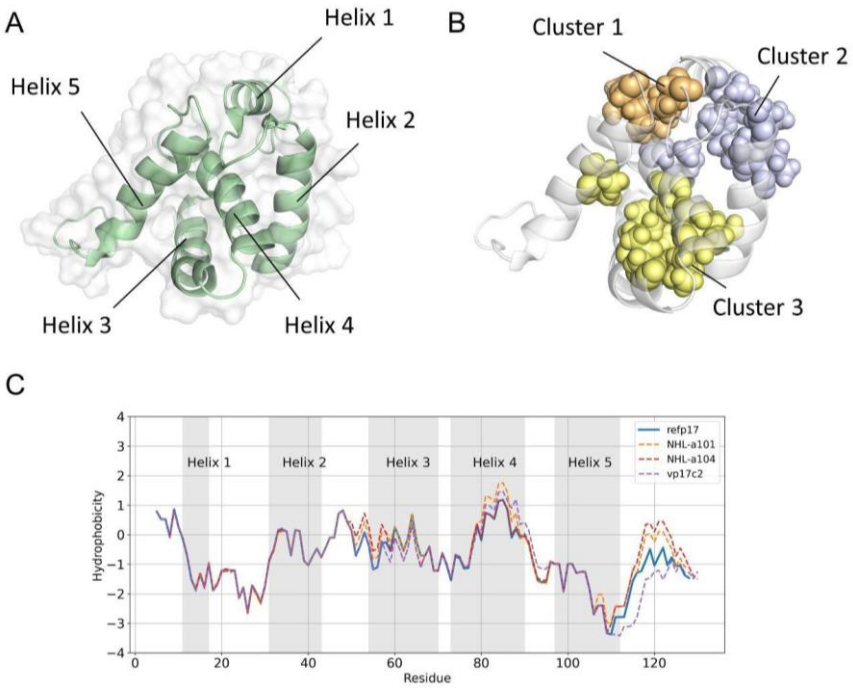

Fig. 8. Hydrophobic profile and structural features of refp17 and vp17s. *(A) refp17 structure showing protein folding (green cartoon and white surface). (B) refp17 ILV clusters 1, 2 and 3 in orange, purple and yellow. The residues forming the clusters are shown as a Corey–Pauling-Koltun model. (C) Hydrophobic profiles of refp17 and vp17s.*

4.2.2 Exploring the structural dynamics in refp17 and vp17s

The results from this work have been collected and published in D'Ursi P, Rondina A, Zani A, Uggeri M, Messali S, Caruso A, Caccuri F. Molecular Mechanisms Involved in the B Cell Growth and Clonogenic Activity of HIV-1 Matrix Protein p17 Variants. Viruses. 2024 Jun 28;16(7):1048.

Background: p17s found in patients with HIV-associated lymphomas exhibit mutation along the whole sequence and insertions in the COOH-terminus, particularly in the H5 helix, that, along with ILV clusters, is crucial for stabilizing the hydrophobic core, maintaining the folding of the protein. Alterations of the hydrophobic core lead to the exposure of the functional epitope involved in B-cell proliferation and located in the N-terminus of the p17. This epitope has been shown to interact with the protease-activated receptor 1 (PAR-1) promoting B-cell growth, suggesting that p17 conformational changes drive the acquisition of its oncogenic potential. Since computational studies can provide insights on how oncogenic-driving structural changes occur in p17 at the atomic level, we employed a combination of structural modeling and MDs to assess the structural consequences of specific mutations and insertions in vp17s, to elucidate the molecular mechanisms driving vp17s gaining of B-cell clonogenic properties.

Methods: Crystallographic structures 1TAM and 2HMX of p17 were from PDB database (Berman H.M. et al., 2000) and used as templates for modeling refp17, vp17s NHL-a101, NHL-a104 from HIV+ patients, vp17c2 (Dolcetti R. et al., 2015) and refp17 carrying W16A and Y29A mutations ($p17^{W16A+Y29A}$). Protein modeling was performed by MODELLER program (version 10.5) (Sali A.A. et al., 1993). Complete structures of p17s were built by comparative multi-template modeling. Models' assessment was done by Ramachandran plots (PROCHECK software version 3.5.4) (Laskowski R.A. et al., 1993) and qualitative model energy analysis by QMEANDisCo (Studer G. et al., 2020). The model showing the highest structural quality was selected to build the system.

MDs were carried out using Amber18 MDs package with the ff14SB force fields (Case D.A. et al., 2018). 3D structures were solvated using TIP3P water models in octahedral boxes extending 10 Å from the protein surface. The systems were neutralized with counterions (150 mM) (Machado M.R. et al., 2020). Energy minimization process was conducted in 4 consecutive steps, each consisting of 5000 steps and employing the steepest descent method, gradually decreasing restraint to allow the system to stabilize. 1st minimization step: all atoms (except hydrogens) were restrained at 2.0 Kcal/mol; 2nd step: the restraints were reduced to 1.0 Kcal/mol and removed for water molecules; 3rd step: the restraints were further lowered to 0.5 Kcal/mol; final minimization step: all restraints were removed. The system was heated from 0 to 300 K at a constant volume using a Langevin thermostat, with protein atoms restrained at 1.0 Kcal/mol for 200,000 steps. Afterward, the system was equilibrated at constant pressure for 3 ns. Each system was then subjected to a MD simulation of 100 ns without restraints, with a cut-off for non-bonded interactions set at 8 Å.

To analyze the simulation results, MDs data analysis was performed using CPPTRAJ from the AMBER18 package. H-bond network was evaluated by considering only those H-bond interactions present for more than 75% of the simulation (Durrant J.D. et al., 2011).

aMDs were performed to enhance conformational sampling of macro and micro molecules through the addition of a non-negative boost potential. Average potential energy and dihedral energies were used as references to calculate the appropriate boost factor for the simulations (Hamelberg D. et al., 2004). A "dual-boost" aMDs approach was applied to better explore the conformational dynamics of the proteins. Average Ep and Ed values for each system were obtained from the initial 100 ns of MDs and used to calculate the EthreshD, EthreshP, αD and αP values necessary for aMDs (**Tab. 9**) (Pierce L.C. et al., 2012). The systems were simulated for 700 ns, with the simulation of NHL-a101 extended to 1 μs to further explore the energy landscape and allow the system to reach a state where observable conformational changes could occur.

To investigate whether Trp16 and Tyr29, 2 residues of the hydrophobic core of p17, contribute to the impaired activity of the variants, the values of the dihedral angles Φ (phi) and ψ (psi), defined on the backbone atoms, and χn (chi-n), defined on the side chain atoms, were calculated. Atoms of angles Φ (C-N-CA-C), ψ (N-CA-C-N), and chi1(C-CA-CB-CG) of the Trp16 and Tyr29 and chi2 (CA-CB-CG-CD) of Trp16 were selected for this analysis (**Fig. 9**).

System	Nres	Natoms	MDs (ns)	aMDs (ns)	EthreshD (kcal/mol)	αD (kcal/mol)	EthreshP (kcal/mol)	αP (kcal/mol)
refp17	132	64717	100	700	2099	92.4	-195907	12943
NHL-a101	134	52856	100	1000	2130	93.8	-159308	10571
NHL-a104	134	45108	100	700	2126	93.8	-135564	9022
vp17c2	134	52735	100	700	2159	93.8	-158800	10547
p17W16A+Y29A	132	49936	100	700	2095	92.4	-150684	9987

***Tab. 9**: **Model Systems simulated by aMDs**. N_{res} and N_{atoms} are the number of residues and of atoms in the system. Etreshd, αD, Etreshp, αP are the average dihedral energy threshold, the inverse strength boost factor for the dihedral energy, the average total potential energy threshold, and the inverse strength boost factor for the total potential energy, respectively.*

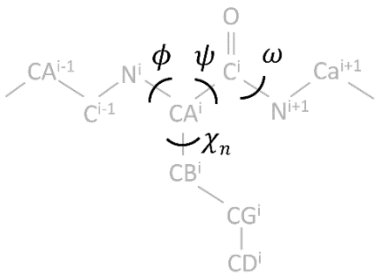

***Fig. 9. Protein dihedral angles**: for the protein backbone, phi is defined by the sequence C^{i-1}-N^i-CA^i-C^i, Ψ is defined by N^i-CA^i-Ci-N^{i-1}, and ω is defined by CA^i-C^i-N^{i+1}-CA^{i+1}; for the protein side chains, χn (chi-n) is defined by C^i-CA^i-CB^i-CG^i.*

Principal component analysis (PCA) was also performed to reduce the number of dimensions needed systematically to reveal the most important motions in proteins via a decomposition process that filters the observed motions from the largest to the smallest spatial scales. First, the covariance matrix was calculated from simulation data. Rotational and translational motions of the proteins during the MDs trajectory were removed by performing the best coordinates fit on a reference and by calculating their RMSD. Principal Components (PCs) were then obtained by diagonalizing the covariance matrix, thus allowing for the determination of the eigenvalues and eigenvectors of each PC. The analysis was performed on the backbone of the protein to provide a more detailed characterization of essential spatial motions.

Results and Conclusions: The comparison of the MDs trajectories for both refp17 and vp17s revealed notable distinctions in their behaviour that were quantified analyzing residue-specific fluctuations by RMSF. The data indicated an overall increase in flexibility in vp17s when compared to refp17, with the greatest changes found in the region between residues 8–32 in the vp17 NHL-a101 and NHL-a104. Notably, this is the area containing the functional epitope associated with B cell clonogenicity (Giagulli C. et al., 2017).

Another region displaying significant flexibility was the C-terminal portion of the protein. In particular, residues 111–120 in vp17c2 (carrying an EK insertion) and residues 121–132 in NHL-a101 and NHL-a104 (both with Ala-Ala insertions) showed increased flexibility compared to refp17 (**Fig. 10**). These findings suggest that the structural alterations in vp17s may influence their functional roles, particularly in their interaction with immune cells.

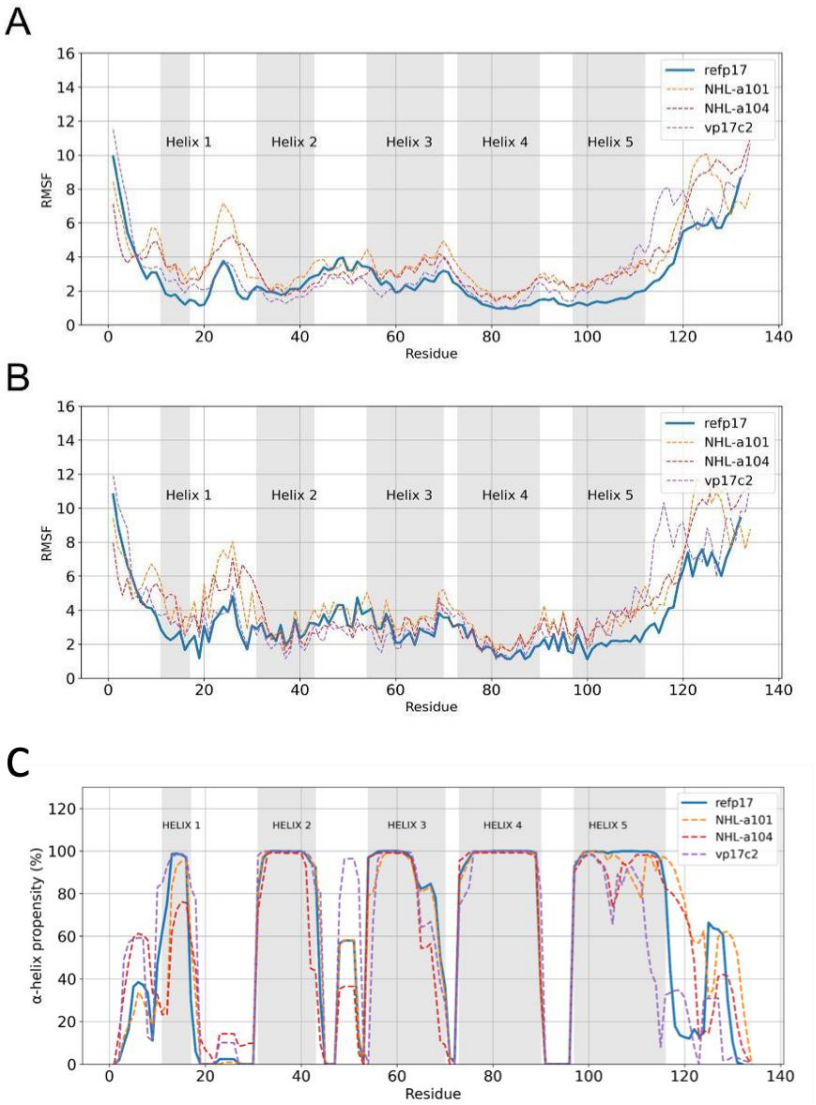

Fig. 10. Protein residue fluctuations analysis and α-helix propensity comparison*. RMSF calculation based on backbone (A) and side chains (B) of the proteins. Grey regions correspond to the residues of the α-helices. RMSF values of refp17 and vp17s are shown in solid and dashed lines, respectively. (C) Average of α-helix propensity profile of refp17 and vp17s (solid and dashed lines, respectively) calculated from aMDs (300 ns).*

The causes of this increased flexibility were further investigated by analysis of the secondary structure of refp17 and vp17s. The results highlighted differences in the helical propensity, particularly in the final portion of the H3 helix in NHL-a104 and vp17c2 and in the H1 helix in NHL-a104 (**Fig. 10C**), where the helices were less stable than those of refp17. Also, the H5 helix, which extended from residues 97–116 in refp17, displayed consistent α-helical stability after 300 ns of simulation, a stability that was not observed in vp17s. Specifically, In the NHL-a101 variant, a short loop initially formed between residues 112–114 early in the simulation (100–300 ns), which later extended to include residues 107–111 during the intermediate phase (500–600 ns). In contrast, the NHL-a104 variant exhibited a loop formation between residues 104–107 early in the simulation (100–400 ns), and a subsequent shortening of H5 up to residue 114 in the last 100 ns. Vp17c2 showed instability in H5, that extends only to residue 103 by the end of the simulation (**Fig. 11**).

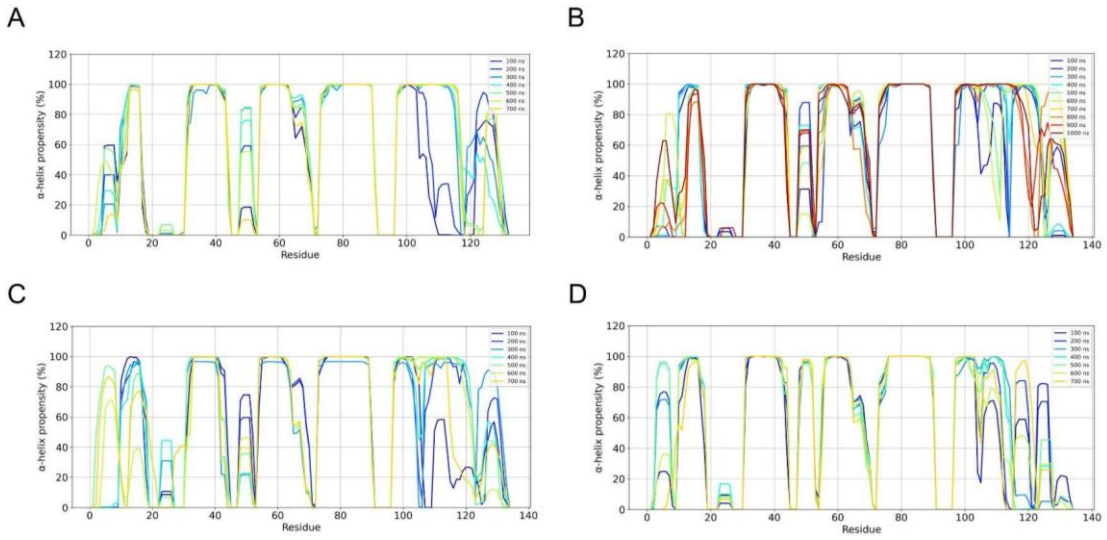

***Fig. 11**. **Time series of α-helix propensity**. (A) α-helix propensity of refp17, (B) NHL-a101, (C) NHL-a104, and (D) vp17c2 are shown for every 100 ns of aMDs.*

The alterations in α-helical stability in vp17s may be linked to their B cell clonogenic activity. It is known that changes in a protein's secondary structure have an impact on its function, particularly when specific interactions that confer functional advantages are gained (Yue P. et al., 2006). This is further supported by comparisons with the simian immunodeficiency virus (SIV) MA, which is known to exhibit B cell clonogenic activity (Feichtinger H. et al., 1990). Notably, SIV p17 has a truncated H5 helix compared to HIV-1. Also, the C-terminus of SIV p17 folds into a β-sheet structure (Rao Z. et al., 1995). The instability of H5 in vp17s may reflect a functional adaptation like the truncated H5 observed in SIV, suggesting a possible correlation between structural changes and functional properties of these proteins. Finally, an analysis focusing on H-bonds and hydrophobic interactions that revealed how vp17s, including NHL-a101, NHL-a104, and vp17c2 showed a reduction of 10, 9, and 12 H-bonds, respectively, with each bond showing less than 75% of its original lifetime (**Tab. 10**). Five disrupted interactions were common among vp17s (**Fig. 12A,D**) and four of these are part of a H-bond network in refp17 that involves residues Trp16, Ile19, Arg20, Leu21, Lys27, Tyr29, Lys95 and Thr97 (**Fig. 12E,H**).

H-bond interactions					
Acceptor	**Donor**	**refp17**	**NHL-a101**	**NHL-a104**	**vp17c2**
THR 53/O	CYS 57/N	91.73%	73.02%	88.28%	1.12%
LYS 27/O	LEU 21/N	90.45%	62.28%	73.08%	71.13%
SER 54/O	ARG 58/N	89.84%	67.65%	90.08% (M)	28.29%
LYS 95/O	ARG 20/N	89.74%	67.80%	5.38%	10.72%
TYR 79/O	ALA 83/N	89.73%	88.58%	92.41%	56.52%
LEU 13/O	GLU 17/N	88.94%	84.29%	61.21%	76.66%
GLU 12/O	TRP 16/N	85.32%	66.03%	59.99% (M)	86.36% (M)
GLU 40/O	PHE 44/N	85.07%	79.75%	86.60%	59.24%
TRP 16/O	ILE 19/N	82.80%	56.92%	30.71%	13.27%
GLU 73/O	SER 77/N	80.67%	72.09%	81.32%	66.32%
SER 72/O	ARG 76/N	80.39%	77.12% (M)	89.31% (M)	69.07% (M)
TRP 36/O	GLU 40/N	79.97%	87.27%	64.05%	74.75%
LEU 41/O	VAL 46/N	79.13%	58.11%	23.95%	7.72%
ILE 19/O	TYR 29/N	76.17%	43.79%	23.58%	2.83%
GLU 99/O	LYS 103/N	75.78%	69.32%	66.71%	85.02%

***Tab. 10**. **H-bond frequencies of refp17 and vp17s during aMDs**. Residues involved in H-bond and the corresponding fraction of time that each interaction lasts are reported. (M): mutated residues.*

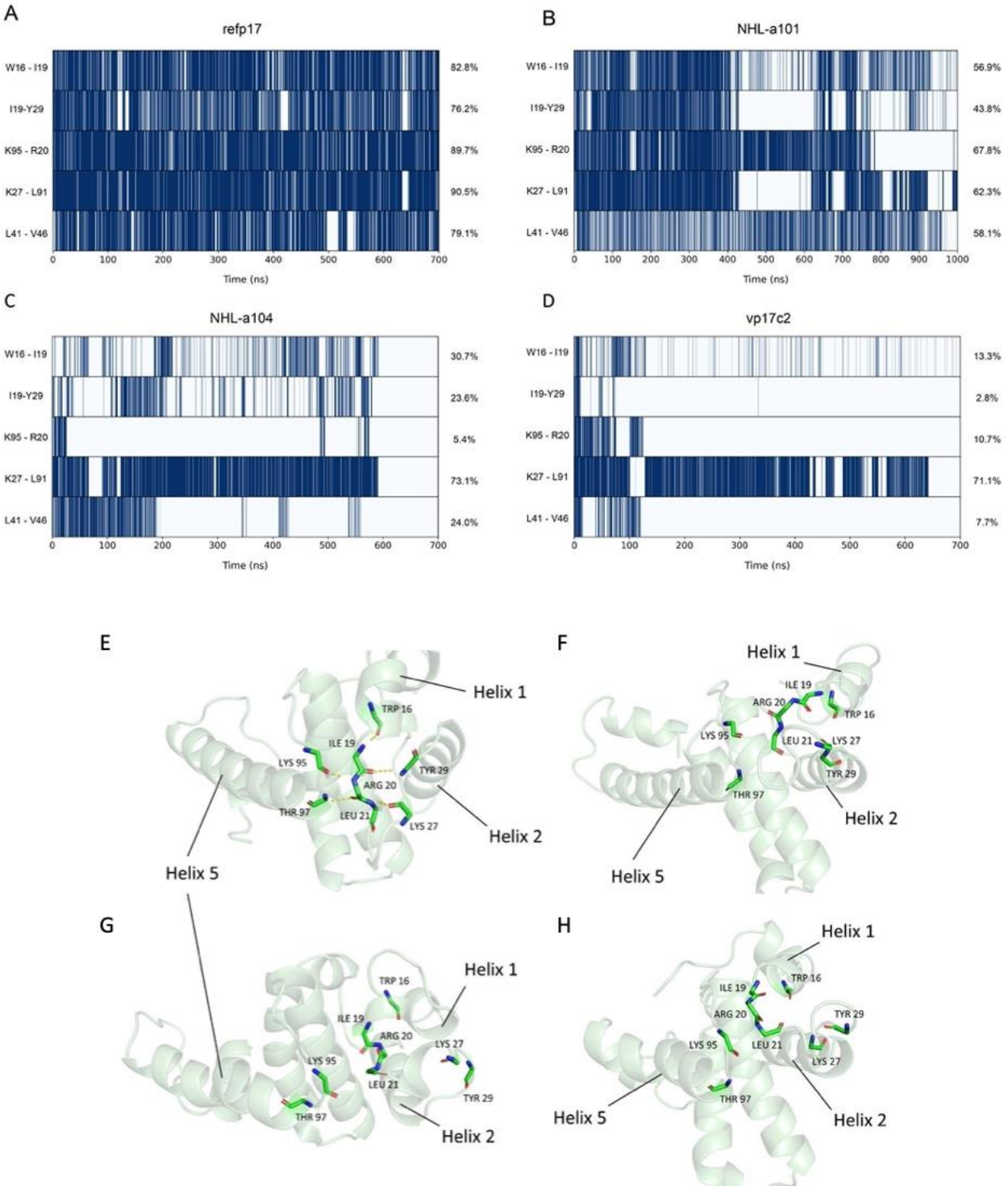

Fig. 12. H-bond analysis*. (A) H-bond occupancy in refp17, (B) NHL-a101, (C) NHL-a104, and (D) vp17c2 during the aMDs. Y-axis: residues and atoms of each H- bond. Dark blue and light blue indicate the presence or absence of the interaction. (E–H) Residues involved in the H-bond network in refp17 and vp17s. (E) Stable H-bond network in refp17 observed along aMDs. In contrast, in (F) NHL-a101, (G) NHL-a104, and (H) vp17c2, the H-bond network is disrupted.*

This H-bond network stabilizes a mixed 3-stranded β-sheet that contributes to maintain the connection between helices H1, H2 and H5, thus preserving the overall tertiary structure of this region in refp17. Interestingly, this disrupted network in the vp17s also involves residues located in the N-terminal region, where the functional epitope responsible for B cell clonogenicity residues.

Next, we focused on the analysis of hydrophobic interactions, specifically those with a lifespan of less than 60% in the vp17s. Our attention was drawn to the region termed AT20, which encompasses the functional epitope and where the H-bond network had previously been shown to be disrupted. By visually examining this region, we identified a hydrophobic cluster (HC) termed HC-AT20, comprising residues Leu8, Leu13, Trp16, Ile19, Leu21, Tyr29, Ile34, Ala37, Leu41, Thr81, Thr84, Leu85, Val88, and Thr97 (see also **Fig. 15** below). The analysis revealed that Trp16 and Tyr29 are central to this cluster, as the hydrophobic interactions are organized around these core residues.

To assess the contribution of Trp16 and Tyr29 in maintaining refp17 globular structure, we conducted a detailed analysis of residues' behaviour during aMDs, focusing on their dihedral angles. In refp17, the χ2 dihedral angle of Trp16 fluctuated between 0° and −100°, whereas in vp17s, this angle shifted more dramatically, often exceeding 100° (**Fig. 13**). A similar pattern was observed for Tyr29, where the χ1 dihedral angle, which remained close to −100° in refp17, exhibited larger variations between −200° and 200° in vp17s (**Fig. 14**). These differences underscore the altered side chain orientations of Trp16 and Tyr29 in vp17s compared to refp17 during the simulations.

The prominent dihedral angle movements of Trp16 and Tyr29 in vp17s were further confirmed by structural analysis of vp17s by aMDs. Both residues were consistently observed to detach from the HC-AT20 region in the B cell clonogenic vp17s. This disengagement led to the breakdown of the HC-AT20, a hydrophobic cluster critical for stabilizing the protein's structure (**Fig. 16**).

In refp17, Trp16 and Tyr29 are critical residues that help to maintain the integrity of HC-AT20 (see **Fig. 15** above). However, in vp17s such as NHL-a101, NHL-a104, and vp17c2, the loss of H-bonds destabilizes this region, increasing its flexibility and leading to the eventual disruption of HC-AT20. This loss of structure near the functional epitope AT20 is illustrated by the disengagement of Trp16 and Tyr29 in vp17s, contributing to partial misfolding in this region (**Fig. 16**). In the B cell clonogenic vp17s, this disruption of HC-AT20 is a critical event that explains the observed conformational changes.

We applied PCA to examine the conformational motions that mainly represent the dynamics of refp17 and vp17s. PCA is a widely employed method for studying the collective dynamics of proteins, and it is of primary interest when investigating large conformational changes (Galindo-Murillo R. et al., 2015). Refp17 showed a stable H5, with no significant movement of the N-terminal portion of the protein, where the active epitope is located (**Fig. 17A**). Conversely, all vp17s displayed a loss of compactness in the loop between H1 and H2 (**Fig. 17B,D**), which is less constrained to the globular portion of the protein. Furthermore, the movement of H1 and the instability of H5 in NHL-a101 and vp17c2 were also evident. This result highlights the conformational changes occurring in vp17s as compared to refp17 and allows us to hypothesize that the AT20 epitope of vp17s may likely be more exposed and suitable for fitting with PAR-1 than the one in the refp17.

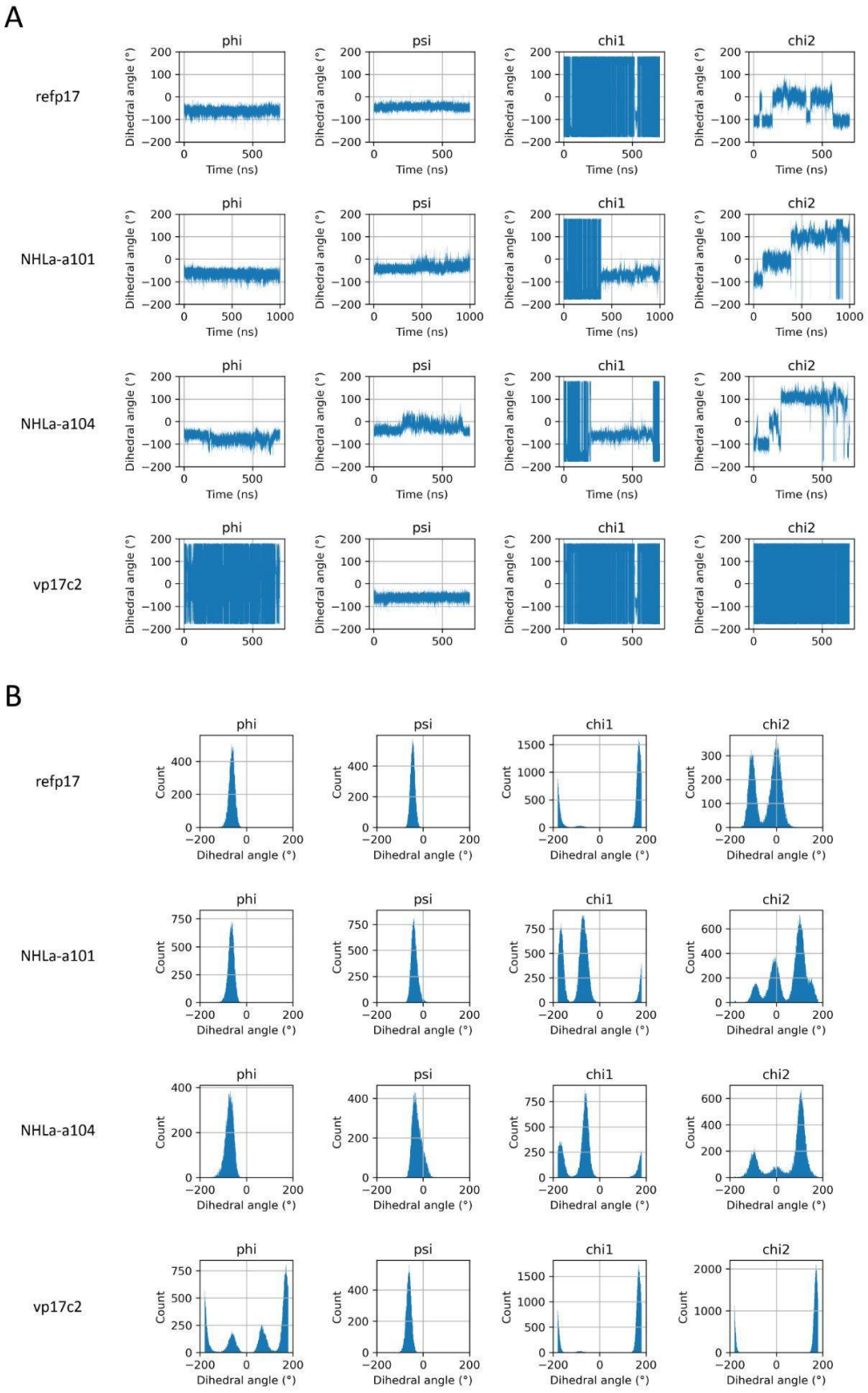

Fig. 13. ***Comparative Dihedral Angle Distribution Analysis****: (A) Dihedral angle distribution along aMDs of Φ, ψ, χ1, and χ2 dihedral angles of Trp16 in refp17 and vp17s. (B) Density histogram of dihedral angle distribution along aMDs of Φ, ψ, χ1, and χ2 dihedral angles of Trp16 in refp17 and vp17s.*

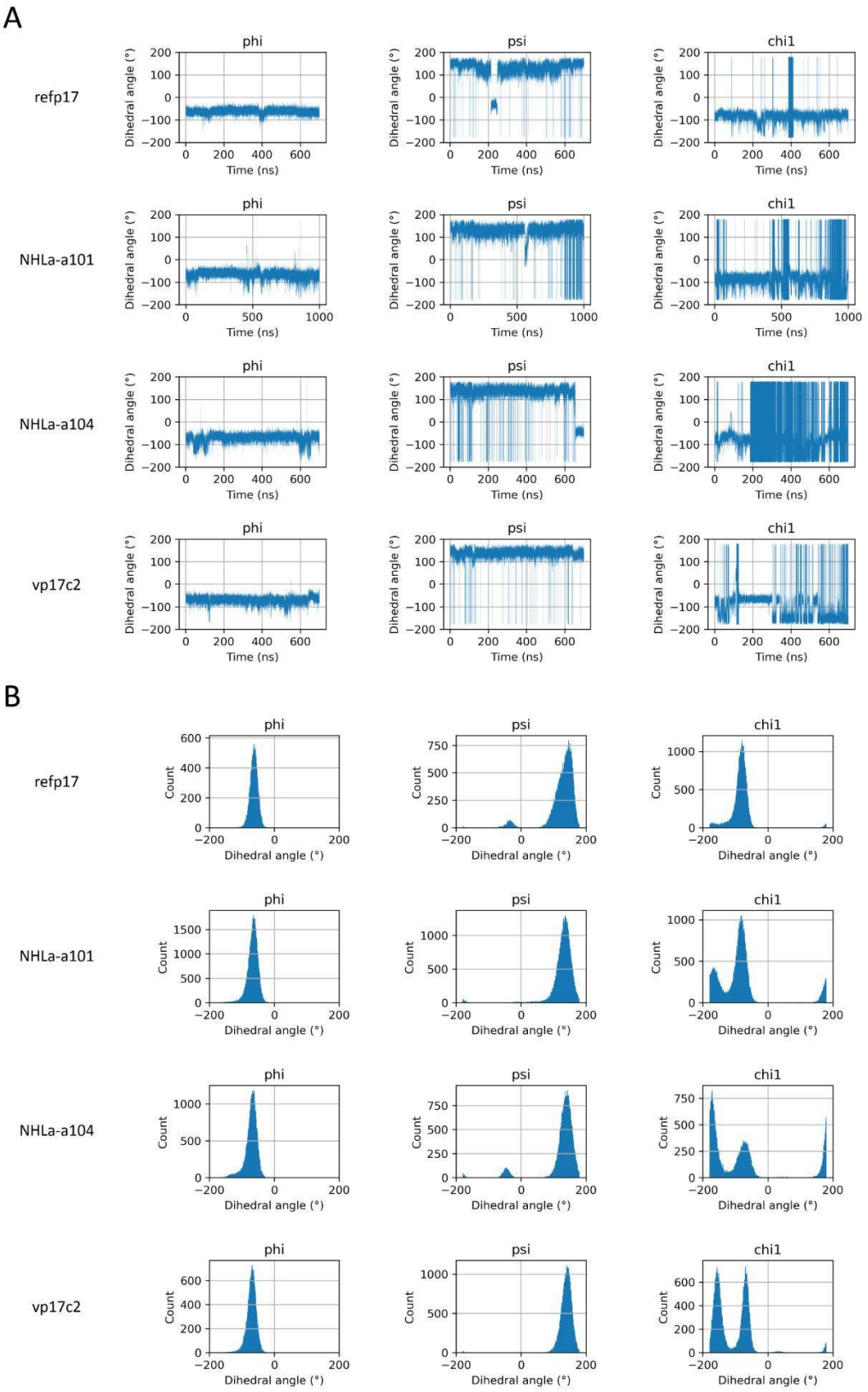

Fig. 14. Comparative Dihedral Angle Distribution Analysis. *(A) Dihedral angle distribution along aMDs of Φ, ψ, and χ dihedral angles of Tyr29 in refp17 and vp17s. (B) Density histogram of dihedral angle distribution along aMDs of Φ, ψ, χ1 dihedral angles of Tyr29 in refp17 and vp17s.*

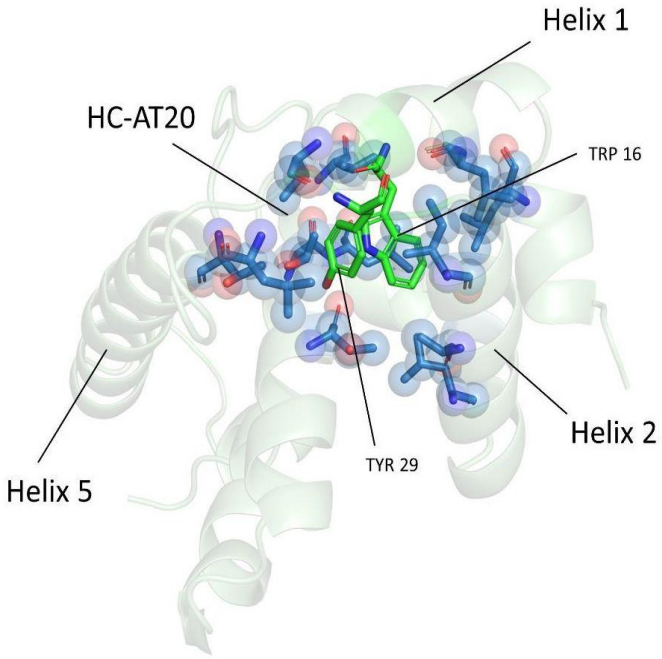

Fig. 15. Representation of the hydrophobic cluster near the AT20 region (HC-AT20)*. The refp17 protein is represented as a cartoon; Trp16 and Tyr29 are denoted by green sticks, while the other residues forming the cluster in blue sticks.*

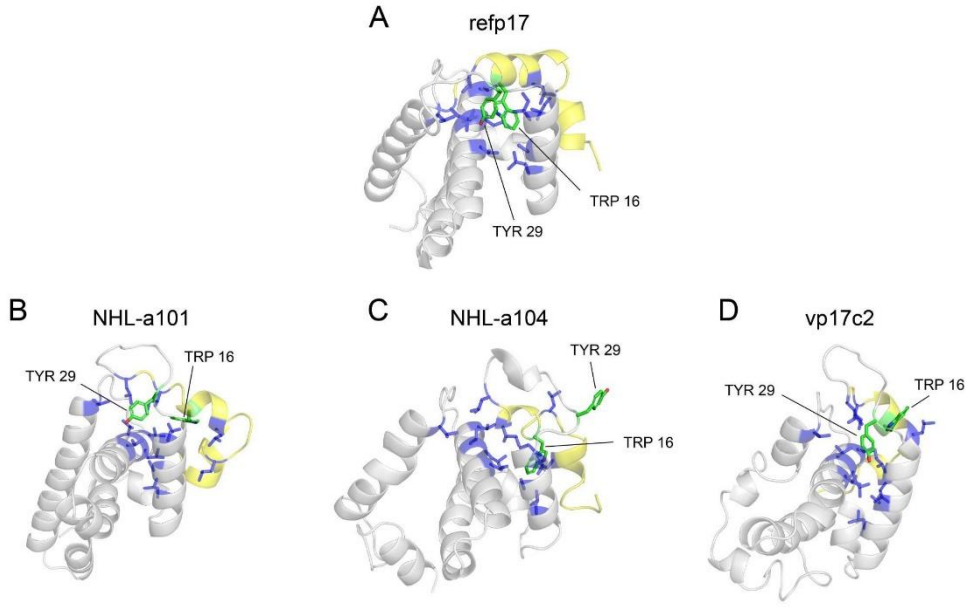

Fig. 16. Disengagement of Trp16 and Tyr29 from HC-AT20 in B cell clonogenic vp17s*. Starting from a stable HC-AT20 in refp17, the loss of H-bonds makes the region of the functional epitope AT20 (yellow cartoon) more flexible and unstable, ultimately leading to disruption of the HC-AT20 in all vp17s. (A) refp17; (B) NHL-a101; (C) NHL-a104; and (D) vp17c2. Trp16 and Tyr29 are indicated with green sticks, and HC-AT20 is indicated via a blue cartoon and sticks.*

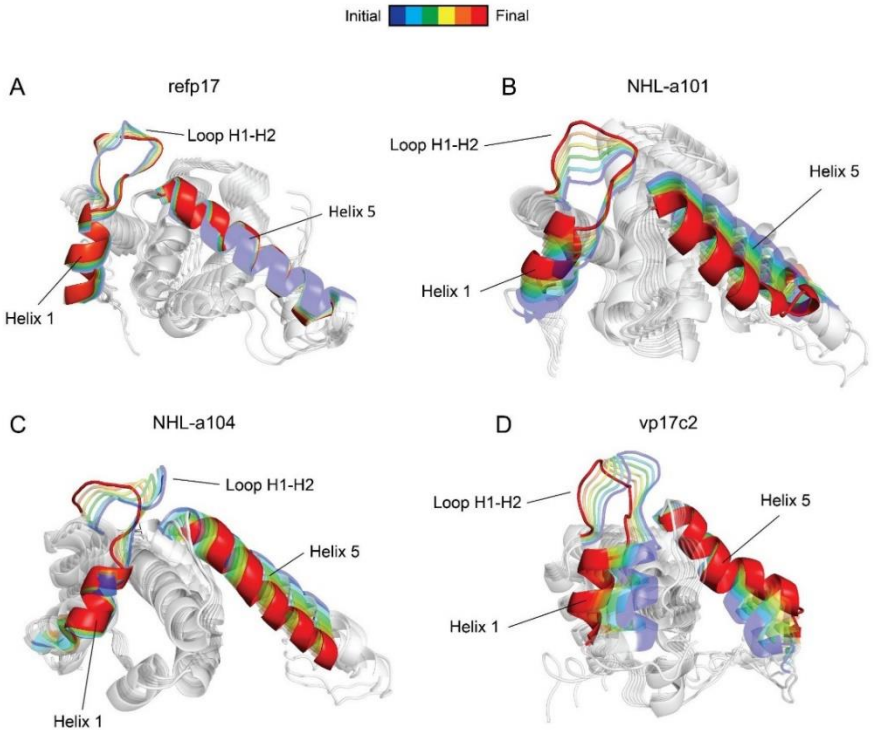

Fig. 17. Motion across the first principal component*. The viral proteins (A) refp17, (B) NHL- a101, (C) NHL-a104, and (D) vp17c2 are represented as cartoons. H1, H5, and the H1-H2 loop are highlighted in colors from the start (blue) to the end (red) of the motion.*

4.3. Dual Approach Validation of Trp16 and Tyr29 in Protein Stability and B Cell Clonogenic Activity

4.3.1 In Silico Validation of Alanine scanning analysis of Trp16 and Tyr29

The results from this work have been collected and published in D'Ursi P, Rondina A, Zani A, Uggeri M, Messali S, Caruso A, Caccuri F. Molecular Mechanisms Involved in the B Cell Growth and Clonogenic Activity of HIV-1 Matrix Protein p17 Variants. Viruses. 2024 Jun 28;16(7):1048.

Background: In silico validation is a proven technique for assessing the structural and functional consequences of specific protein mutations before experimental testing. Alanine scanning, a key approach in this context, enables the evaluation of the contributions of individual residues to protein stability dynamic behaviour, interaction networks and function. Here, alanine scanning has been exploited to understand the role of residues Trp16 and Tyr29 of p17 in protein structural integrity and to complement experimental data about clonogenic activity of vp17s.

Methods: The modeling and system setup for MDs follows what outlined in chapter 4.2.2.

Results and Conclusions: Trp16 and Tyr29 were mutated to alanine in refp17, and the obtained $p17^{W16A+Y29A}$ 3D model was subjected to aMDs. Trajectory analysis revealed no significant increases in fluctuations of $p17^{W16A+Y29A}$ in comparison to refp17 (**Fig. 18**). However, H5 suffered a total loss of secondary structure in the region corresponding to residues 107 to 109. This led to a segmentation of the H5, similar to what was observed in vp17s (**Fig. 19**). Also, the same H-bonds lost in vp17s were missing in $p17^{W16A+Y29A}$, as well as the loss of hydrophobic interactions and disruption of the HC-AT20 (**Fig. 20** and **Tab. 11**). Overall, this result proves the pivotal role played by Trp16 and Tyr29 in maintaining the refp17 globular structure near the AT20 domain. Ultimately, it

is also conceivable that p17$^{W16A+Y29A}$ may acquire a B cell clonogenic phenotype similar to that of NHL-a101, NHL-a104, and vp17c2 vp17s.

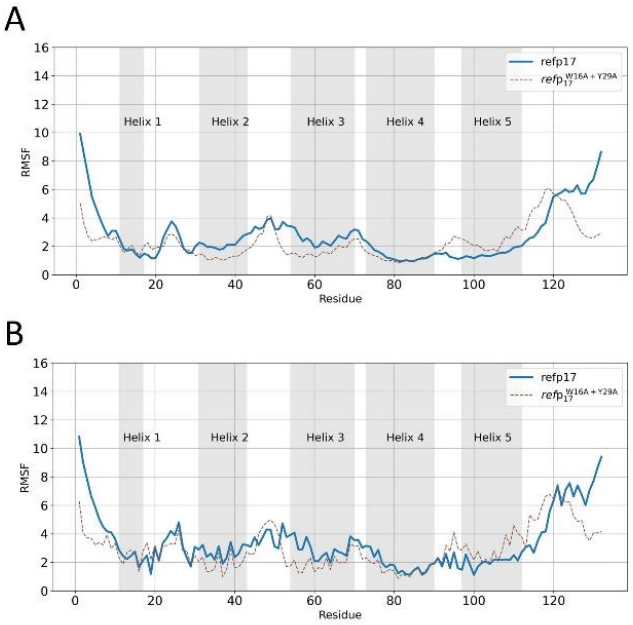

Fig. 18. Protein Residue Fluctuations analysis. *RMSF analysis of protein backbone (A) and side chains (B). Gray regions correspond to the residues forming refp17 α-helices. Solid blue lines: fluctuations of refp17 residues. Dashed brown lines: p17$^{W16A+Y29A}$.*

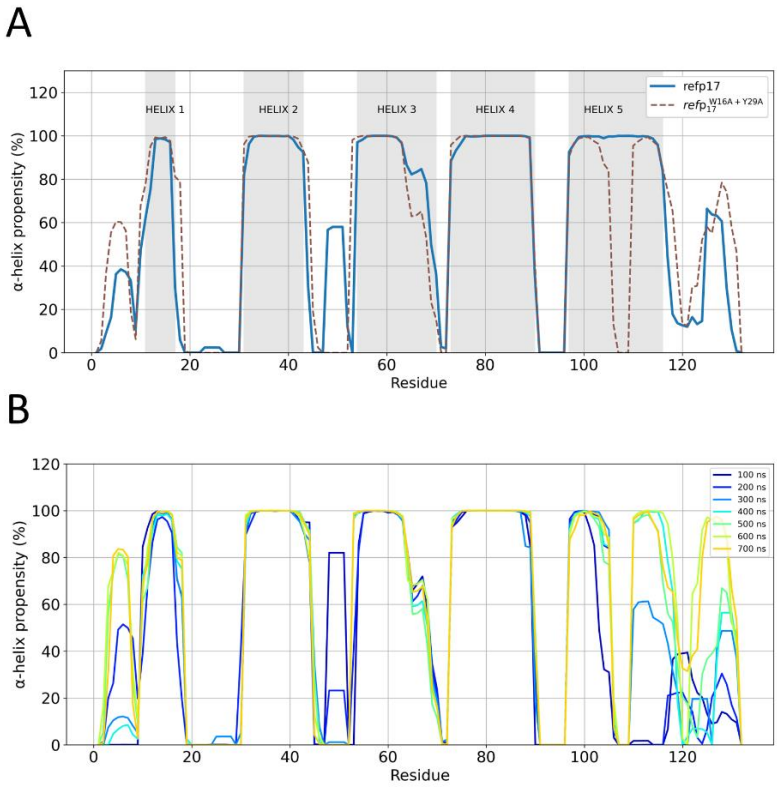

Fig. 19. **α *helix propensity comparison****. (A) Average α-helix propensity profile of p17$^{W16A+Y29A}$ compared to refp17 from 300ns of aMDs. (B) Time series of the p17$^{W16A+Y29A}$ α-helix propensity every 100 ns step of aMDs.*

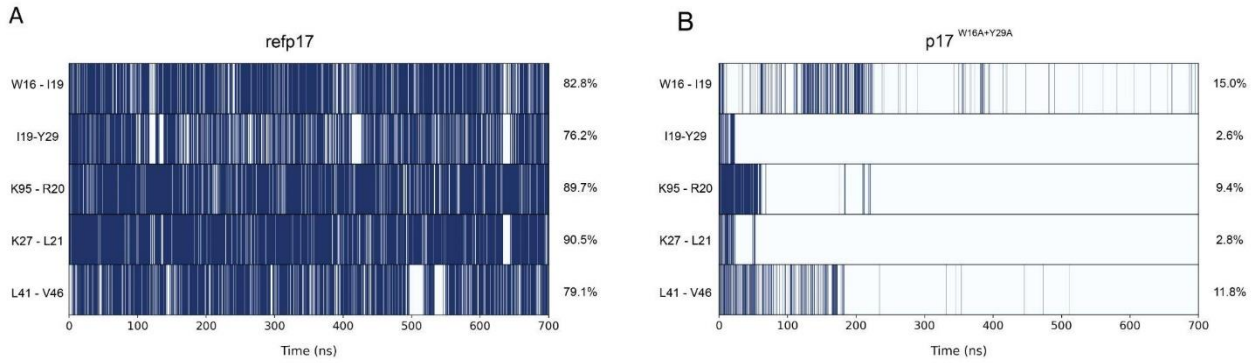

Fig. 20. H-bond time series. *The analysis finds and tracks H-bonds throughout the aMDs. Colors are set according to the presence (Dark blue line) or absence (Light blue line) of the interaction. The analysis compares the p17$^{W16A+Y29A}$ H-bonds with those of refp17.*

Hydrophobic interactions			
Residue 1	**Residue 2**	**refp17**	**p17$^{W16A+Y29A}$**
LEU 8	LEU 13	74.29%	74.29%
LEU 8	TRP 16	62.86%	8.57% (M)
LEU 8	LEU 85	60.00%	34.29%
LEU 13	TRP 16	71.43%	42.86% (M)
LEU 13	LEU 31	80.00%	51.43%
LEU 13	ILE 34	71.43%	82.86%
TRP 16	TYR 29	100.00%	5.71% (M)
TRP 16	ALA 37	62.86%	0.00% (M)
TRP 16	VAL 88	88.57%	88.57% (M)
ILE 19	TYR 29	97.14%	2.86% (M)
LEU 31	VAL 35	71.43%	82.86%
TRP 36	ALA 37	62.86%	34.29%
LEU 41	VAL 46	97.14%	11.43%
LEU 41	LEU 85	71.43%	54.29%
PHE 44	VAL 46	88.57%	11.43%
PHE 44	LEU 75	85.71%	2.86%
PHE 44	LEU 78	88.57%	5.71%
LEU 50	ILE 60	91.43%	0.00%
LEU 51	VAL 82	82.86%	0.00%
LEU 51	LEU 85	97.14%	2.86%
LEU 64	LEU 78	94.29%	94.29%
LEU 75	LEU 78	77.14%	65.71%
TYR 79	ILE 104	71.43%	0.00%
LEU 85	VAL 88	60.00%	42.86%
ILE 92	ILE 94	97.14%	82.86%
LEU 101	ILE 104	68.57%	0.00%

Tab. 11. Hydrophobic interactions analysis*. Comparison of hydrophobic interactions between refp17 and p17$^{W16A+Y29A}$. Residues involved in the hydrophobic interaction, as well as the corresponding fraction of time that each interaction lasts are shown. M, mutated residues.*

4.3.2. Experimental Validation of the Clonogenic Activity of B Cells in point mutation proteins

The results from this work have been collected and published in D'Ursi P, Rondina A, Zani A, Uggeri M, Messali S, Caruso A, Caccuri F. Molecular Mechanisms Involved in the B Cell Growth and Clonogenic Activity of HIV-1 Matrix Protein p17 Variants. Viruses. 2024 Jun 28;16(7):1048.

Background: vp17s have been more frequently observed in HIV-1+ patients with lymphoma, suggesting their potential role in B cell clonogenicity. This gain of function is linked to conformational changes that alter the protein's structural stability, specifically involving the hydrophobic core and the exposure of functional epitopes that promote B cell proliferation.

Refp17 maintains a stable conformation with no significant clonogenic effects on B cells. However, Trp16 and Tyr29 replacement with alanine causes protein destabilization by disrupting the hydrophobic core and the exposure of a functional epitope in the AT20 domain involved in B cell clonogenic activity (see chapter 4.3.1). Such conformational shifts mimic vp17s structural features.

On these bases, here we approached the experimental validation of our computational predictions by evaluating the effects of refp17 mutations at Trp16 and Tyr29 in a clonogenic assay.

Methods: Human B lymphoma cell line Raji was obtained from American Type Culture Collection (ATCC, Manassas, VA, USA) and cultured in RPMI-1640 medium with 10% (v/v) fetal bovine serum (FBS), 1 mM L-glutamine, and 1 mM sodium pyruvate. Cells were maintained at 37 °C in a humidified atmosphere of 5% CO2. Recombinant endotoxin-free viral proteins were produced as previously described (De Francesco M.A. et al., 2002). Briefly, the coding sequence of refp17 (aa 1–132) from the HIV-1 BH10 isolate was amplified by PCR and cloned into the BamHI site of the pGEX-2T prokaryotic expression vector (GE Healthcare, Chicago, IL, USA). Similarly, the vp17 NHL-a101, derived from a lymphoma biopsy of an HIV-1-infected patient, was cloned into BamHI and XhoI sites of the same vector. Specific alanine mutations were introduced into the refp17 sequence at W16A and Y29A, either individually or in combination ($P17^{W16A+Y29A}$, $p17^{W16A}$, and $p17^{Y29A}$, respectively) using the Quick-Change Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA). All recombinant proteins were purified (>98%) using reverse-phase fast protein liquid chromatography (FPLC), and endotoxin contamination was verified to be less than 0.25 endotoxin units/mL using the Limulus amoebocyte lysate assay (Associates of Cape Cod, East Falmouth, MA, USA). The monoclonal antibody (mAb) to p17, designated MBS-3, was produced and purified in-house as previously described (Caccuri, F. et al., 2012). The B cell colony formation assay was performed by seeding Raji cells at 0.5 cells/well in 96-well plates, incubating them for 8 days in RPMI medium with 10% FBS in the presence or absence of 0.01 μg/mL of the various p17 proteins.

Colony areas were measured using Leica Qwin image analysis software, and propidium iodide (PI)-stained cells were analyzed for viability via flow cytometry. Absolute cell counts were obtained using the MACSQuant® Analyzer (Miltenyi Biotec, Bergish Gladbach, Germany). For selected experiments, $p17^{W16A+Y29A}$ was pre-incubated with control or neutralizing mAb MBS-3 for 30 min at 37 °C. Statistical analysis was performed using chi-square tests for trend and one-way ANOVA with Bonferroni's post hoc test, considering p-values < 0.05 as significant. All analyses were conducted using GraphPad Prism 8 (GraphPad, San Diego, CA, USA).

Results and Conclusions: At day 8 of culture, Raji cells formed a visible single colony in >60% of seeded wells, attesting to active cell proliferation. $P17^{W16A+Y29A}$, $p17^{W16A}$, and $p17^{Y29A}$, differently from their natural counterpart refp17, promoted the formation of colonies with a drastically larger size as compared to cells cultured in medium alone (**Fig. 21.A,B**). Similar results were obtained with the vp17 NHL-a101 used as a positive control (Dolcetti, R. et al., 2015). When the number of cells from an equal number of colonies was evaluated via PI staining and flow cytometry, $P17^{W16A+Y29A}$, $p17^{W16A}$, $p17^{Y29A}$ or NHL-a101 showed an increased B cell proliferation as compared with cells treated with refp17 or NT (**Fig. 21C**).

In order to confirm that the gain of B cell clonogenic function showed by refp17 arises from the exposure of the AT20 epitope following the W16A and Y29A double mutations, the single-cell cloning assay was performed in the presence or absence of an unrelated mAb (Ctrl) or of the p17 neutralizing mAb, named MBS-3, which recognizes the AT20 epitope of refp17, included between aa 9 and 22 (De Francesco, M.A. et al., 2002). As expected, in the presence of 0.01 μg/mL of $p17^{W16A+Y29A}$, colonies showed a significantly ($p < 0.001$) larger size than colonies formed by NT cells.

Interestingly, the presence of the mAb MBS-3 (1 µg/mL) completely inhibited the clonogenic activity of $p17^{W16A+Y29A}$, whereas the Ctrl mAb (1 µg/mL) proved to be ineffective (**Fig. 21A**). $P17^{W16A+Y29A}$-triggered cell proliferation was also significantly inhibited by the presence of the mAb MBS-3 but not by that of Ctrl mAb (see **Fig. 19B** above).

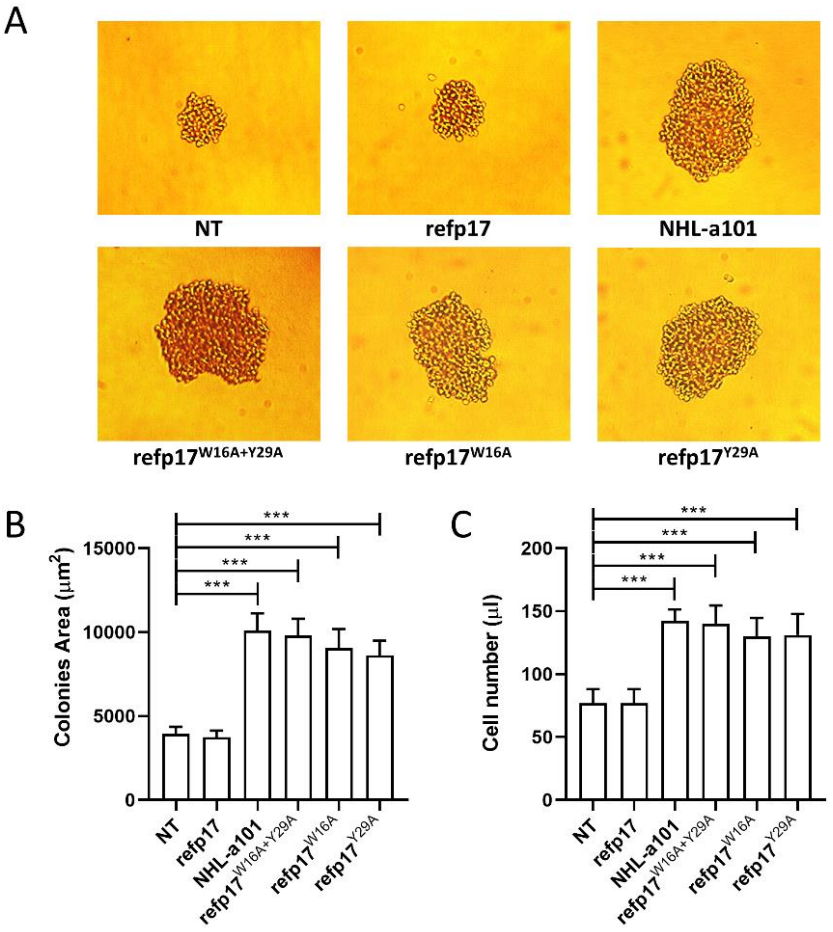

Fig. 21. Effect of $p17^{W16A+Y29A}$, $p17^{W16A}$, and $p17^{Y29A}$ on B cell growth and clonogenicity*. (A) Bright-field images: morphology of colonies (original magnification: ×40). (B) Colony area. (C) Cell counts were obtained by MACSQuant Analyzer. Bars represent the means ± SD of 3 independent experiments. NT, not treated cells. *** p < 0.001.*

These results demonstrate that the clonogenic activity gained by Trp16 and Tyr29 mutations can be neutralized by the specific MBS-3 mAb targeting p17 N-terminal region.

During the course of HIV-1 infection, Ab response to p17 is rare (Fiorentini, S. et al., 2010). Indeed, serum from HIV+ patients does not neutralize the biological activity of the viral protein despite the presence of high-titer anti-p17 abs (Iaria M.L. et al., 2014). Interestingly, during a therapeutic phase-I clinical trial, HIV+ patients immunized with a 20-aa-long synthetic peptide, named AT20, representative of the p17 functional region, developed high titers of p17 neutralizing Abs (Focà E. et al., 2015). Here, we evaluated the efficacy of a serum derived from an HIV+ patient immunized with AT20-KLH at inhibiting $p17^{W16A+Y29A}$ clonogenic activity. A single-cell cloning assay was performed by pre- incubating $p17^{W16A+Y29A}$ with a 1:10 dilution of the serum collected 56 days post-immunization (serum 56). A serum derived from the same patient but collected before immunization was used as a control. Serum 56 significantly reduced the size of the colonies and inhibited B cell proliferation induced by $p17^{W16A+Y29A}$ (**Fig. 22B,C**). As expected, serum Ctrl was unable to neutralize the growth- promoting effects of $p17^{W16A+Y29A}$ (**Fig. 22C,D**). These results suggest that vaccination with the AT20-KLH is able to redirect the Ab response against a specific

epitope by hampering the interaction between vp17s and their receptor, thereby blocking the vp17s clonogenic activity on B cells.

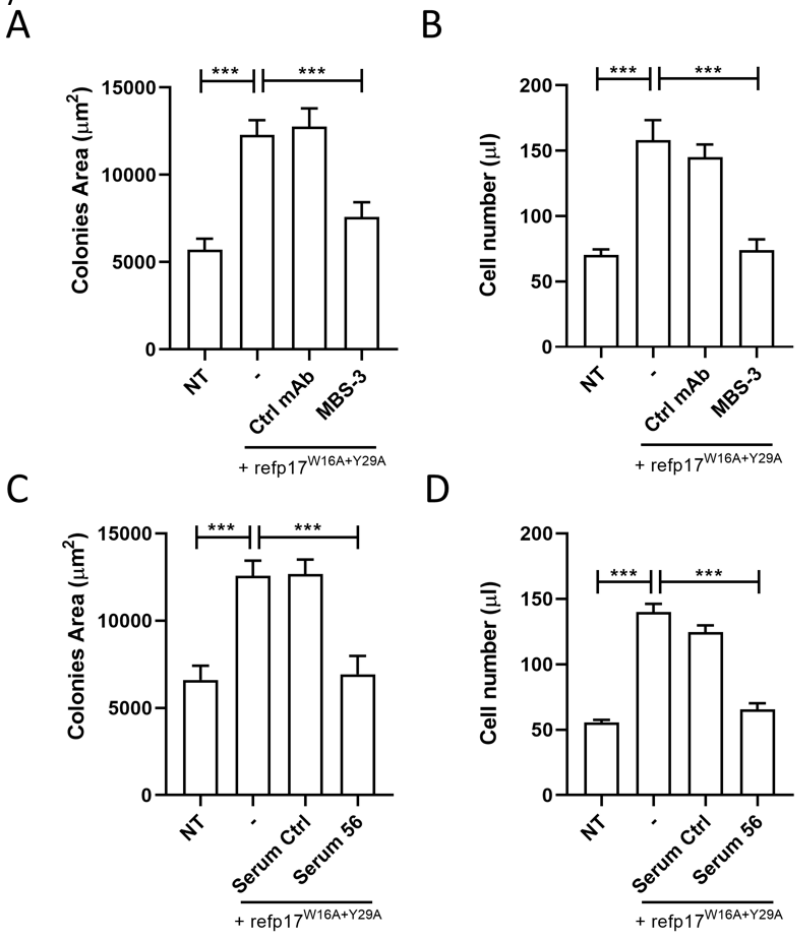

Fig. 22. Neutralization of p17$^{W16A+Y29A}$ clonogenicity using the anti-p17 mAb MBS-3 and a serum derived from an HIV+ patient immunized with AT20-KLH*. For some experimental conditions, the viral protein was pre-incubated for 30 min with 1 µg/mL of mAb MBS-3 or of a Ctrl (A,B), or with a 1:10 dilution of serum 56 or of serum Ctrl (C,D). Bars represent the means ± SD of 3 independent experiments performed in triplicate. NT, not treated cells. *** p < 0.001.*

4.4 Targeting vp17s: preliminary results on drug repurposing

4.4.1. Drug repositioning on vp17s

Background: In this last part of my doctorate research, I've evaluated the possibility to apply a drug repositioning approach to tackle clonogenic activity of vp17s.

Drug repurposing (or repositioning) consists in the identification of new therapeutic uses for existing drugs already approved for other pathologies that has recently gained significant attention. In contrast to classical drug development, that requires 10-15 years and billions of dollars in investment, drug repositioning offers a more cost-effective and faster alternative. Repositioned drugs have already undergone safety testing, making it easier to bring them to market for new indications. For instance, thalidomide, once infamous for its teratogenic effects, was repurposed for use in multiple myeloma and leprosy. More recently, a systematic approach to drug repositioning has emerged, facilitated by computational tools which allow researchers to quickly predict potential interactions between existing drugs and new biological targets (Jourdan J. P. et al., 2020).

Here, molecular docking has been employed to screen libraries of existing drugs to identify candidates that interact with vp17s' AT20 epitope, thus potentially inhibiting its clonogenic activity.

Methods: MDs conditions were consistent with those previously described (section 4.2.2) with the simulation's time extended to 2.5 microseconds to allow for a more detailed characterization of the functional site. The most representative conformations along the simulation trajectory were subsequently analysed and used for docking studies. The software Fpocket (Le Guilloux, V. et al., 2009) was used for the identification and characterization of binding pockets (BPs) that were then employed for drug repositioning through docking with drugs contained in the AIFA ligand dataset that I contributed to modify (Fossa P. et al., 2022). Briefly, starting from the approved list of 1,130 of the original AIFA drugs (https://medicinali.aifa.gov.it), non-relevant molecules were excluded, reducing the dataset to 846 small-molecule drug to facilitate drug repositioning. For each molecule in the 3D conformers of the AIFA dataset, H atoms and Gasteiger charges (Gasteiger J. et al., 1978) were added to vp17 conformations, and pdbqt files were generated as input for Autodock Vina docking studies (Trott O. et al., 2010).

Semi-flexible docking was used to allow ligands to explore different conformations while keeping p17 rigid. Default parameters were applied except for exhaustiveness set to 24. P17 docking site encompassed the predicted BPs, with the grid box centred on the geometric centre of these BPs. Docking results were evaluated by clustering the binding poses based on their distance from the identified BPs using a custom Python script. The best binding pose of each drug was selected and further filtered based on molecular weight (<500 Da), binding within specific sub-regions of the BPs and known pharmacological effects (**Fig. 23**).

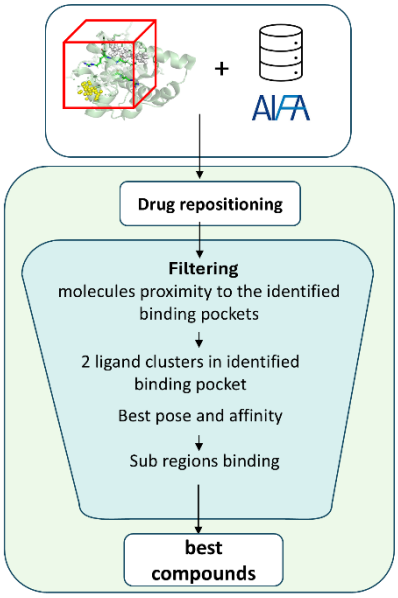

Fig. 23. Schematic representation of the drug repositioning pipeline.

Results and Conclusions: Among the vp17s studied, NHLa104 demonstrated the highest stability within the binding region, while NHLa101 and vp17c2 exhibited destabilizing fluctuations. In particular, vp17c2 showed significant fluctuations at 1.5 μs, where the largest RMSD peak was observed (**Fig. 24.B,D**). The fluctuations were accompanied by high RMSF values compared to earlier steady state (700 ns). On the other hand, NHLa101 stabilizes only at 1.5 μs (**Fig. 24.A,C**), although the fluctuations remained too marked to consider the binding region stable. Instead, NHLa104 preserved more stable RMSD throughout the simulation.

On these bases, a preliminary investigation of the putative binding site was conducted on NHLa104 (that exhibited the most stable binding region) to characterize the chemical and physical properties necessary for successful ligand docking.

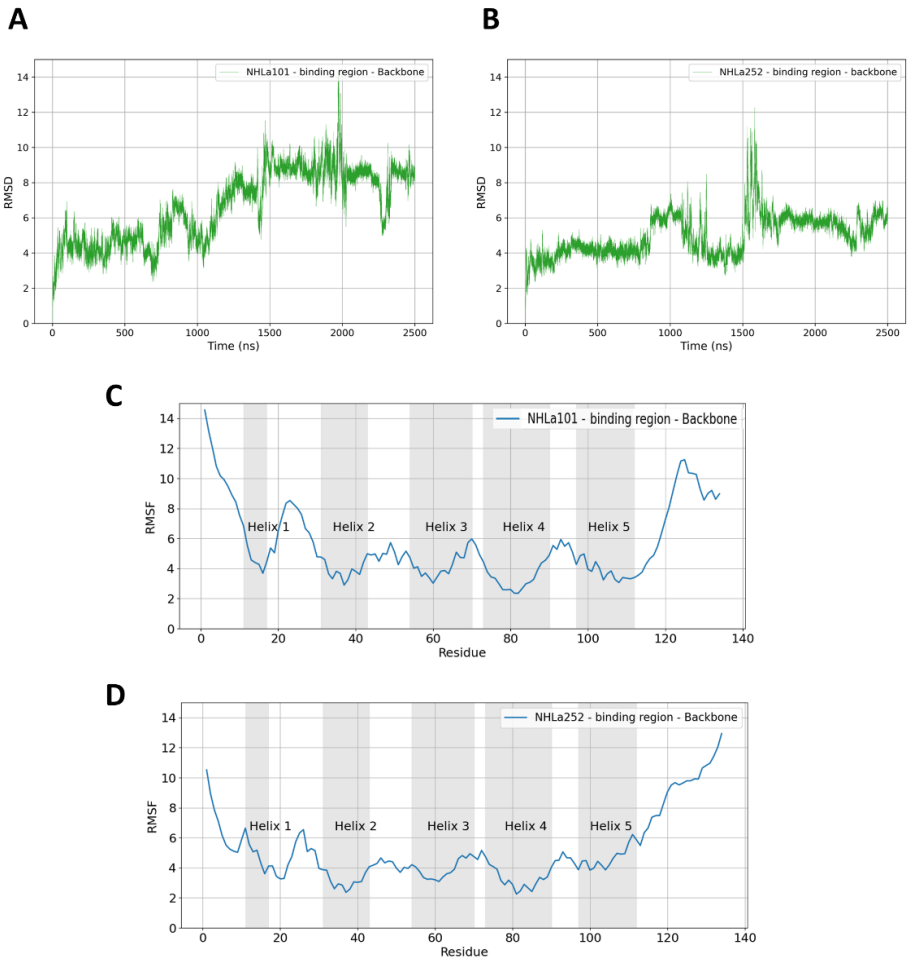

Fig. 24. RMSD and RMSF analysis in vp17s at 2.5 µs. *The figure shows the RMSD levels in NHLa-101 (A) and vp17c2 (B), along with the RMSF levels of the same variants (C and D, respectively).*

Cluster analysis was performed on the most representative structures of NHLa104. This variant's stability was confirmed (**Fig. 25A**), with the most representative cluster covering 76% of MDs. This structure was then used as a reference to identify potential BPs: two possible BPs were identified near AT20, referred to as BP1 and BP2 (**Fig. 25B**). The results indicated a strong potential of BP2 as a viable binding site: it exhibited a good balance between polar and apolar interactions and its hydrophobic and polar interactions points are well-distributed for ligand binding. Instead, BP1 showed a lower druggability index due to its inability to support strong hydrophobic interactions. Also, it is positioned closer to the functional residues Arg15, Lys18 and Arg20, suggesting its capacity to interfere with receptor binding. On these bases, BP2 was selected for further docking studies, as it may block vp17s clonogenic activity.

A ligand library containing active ingredients approved by AIFA was then used for repositioning studies on BP2. Docking results were analysed by clustering the binding poses based on their proximity to the identified BPs using a custom Python script.

The repositioned hits primarily clustered at two ends of BP2, with one cluster closer to the N-terminal region and the other between the loop between helix 1 and helix 2, extending to the start

of helix 5 (**Fig. 26**). Molecules in this latter region exhibited higher affinity, likely due to the increased number of interaction points compared to the N-terminal cluster, which is more hydrophobic.

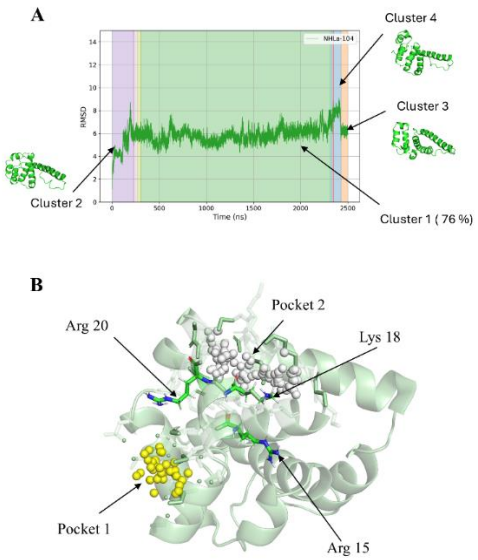

Fig. 25. BP analysis. *Predominant conformations (A) from the simulation with the corresponding structures of the identified clusters shown alongside. Most representative cluster was then used to identify BPs (B). Yellow spheres (BP 1) and white spheres (BP 2) indicate the cavities of the BPs.*

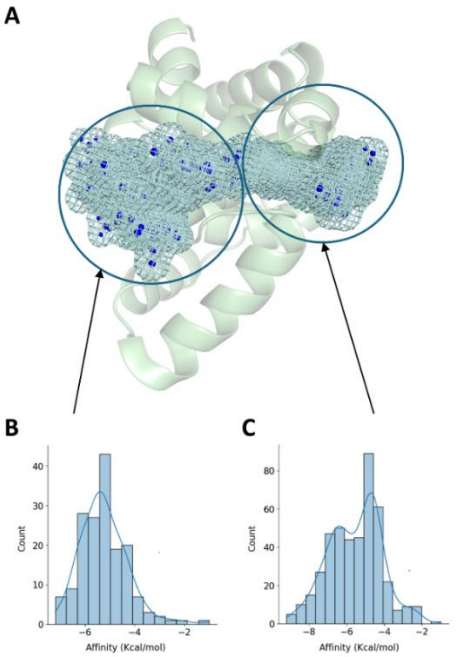

Fig. 26. Distribution of docking poses within the BPs. *Figure B and C illustrate the distribution of hits based on their binding affinity in each region.*

Few molecules spanned both regions, which is favourable for strong, stable binding across the entire BP. The best binding pose of each drug was selected and further filtered based on molecular weight (<500 Da), binding within specific subregions, and known pharmacological effects. Ultimately, 13 small molecules drugs were identified as potential hits (**Tab. 12**).

Molecule	Binding Energy (Kcal/mol)
Molecule 1	-7.7
Molecule 2	-7.2
Molecule 3	-6.9
Molecule 4	-6.9
Molecule 5	-6.8
Molecule 6	-6.6
Molecule 7	-5.8
Molecule 8	-5.6
Molecule 9	-5.5
Molecule 10	-5.5
Molecule 11	-5.4
Molecule 12	-5.3
Molecule 13	-5.1

Tab. 12. Molecular docking analysis of AIFA ligand dataset. *The table presents the binding affinity levels of each molecule as predicted by docking analysis using Autodock Vina.*

From molecular descriptor analysis, several recurring features and trends were identified that could influence the binding potential of the selected hits. H-bonding capabilities also differed across the ligands, with most exhibiting 3 to 6 H bond acceptors and donors, suggesting the potential for strong polar interactions. Notably, hit # 3 had the highest number of acceptors (13) and donors (11), which could lead to extensive interactions with polar regions of vp17s binding site. Structurally, the hits displayed a mix of aromatic and aliphatic rings, indicating a good structural stability when binding to hydrophobic BPs. Lastly, the presence of heteroatoms and heteroaromatic rings was prominent in many hits, suggesting potential for electrostatic interactions or H-bonding. In particular, hits # 10 and # 12 demonstrated these features, indicating potential enhanced binding affinity through polar interactions. finally, hits such as # 3 and # 11, which possess balanced features of aromatic and aliphatic rings and strong H-bonding capabilities, are expected to interact effectively with BP2 endowed with mixed hydrophobic and polar properties.

4.4.2. MDs validation on best candidates

Background: Molecular docking predicts potential binding modes of ligands with proteins based on static structures but overlook the flexibility and conformational dynamics of the binders. MDs are then used to provide a dynamic view of the interaction and the validation of docking poses, ensuring protein-ligand complex stability over time, an important indicator of binding affinity and efficacy of a drug in a mimicked real biological environment (Pavan et al., 2022).

Thus, MDs within a solvated environment over several ns were performed to validate the static predictions obtained for the vp17s, evaluating whether the hits identified in the docking stage remain stably bound to p17 under physiological conditions.

Methods: The topology and coordinate files for p17/hit complexes were generated by the Antechamber package included in AmberTools. Atomic point charges were calculated with the AM1-BCC charge model and the presence of all necessary parameters with the parmchk utility that, in case, generated a frcmod file containing the missing parameters. The 3D models of vp17 were processed with the pdb4amber tool included in the Amber software suite (AMBER Software, University of California, San Francisco, CA, USA). MDs were performed using the Amber18 MD package with the ff14SB force field as described in section 4.2.2.

Results and Conclusions: Each hit was validated for its binding affinity to NHLa-104 using MDs (100 ns simulation). Of the 13 hits, only three (# 3, 5, and 10) maintained binding stability throughout the entire simulation (**Fig. 27**). Hit # 9 exhibited an initial binding interaction with p17, but as the

simulation progressed, the hit's fit within the BP deteriorated, leading to loss of interaction with the functional residues and displacement, being it discarded from further analysis.

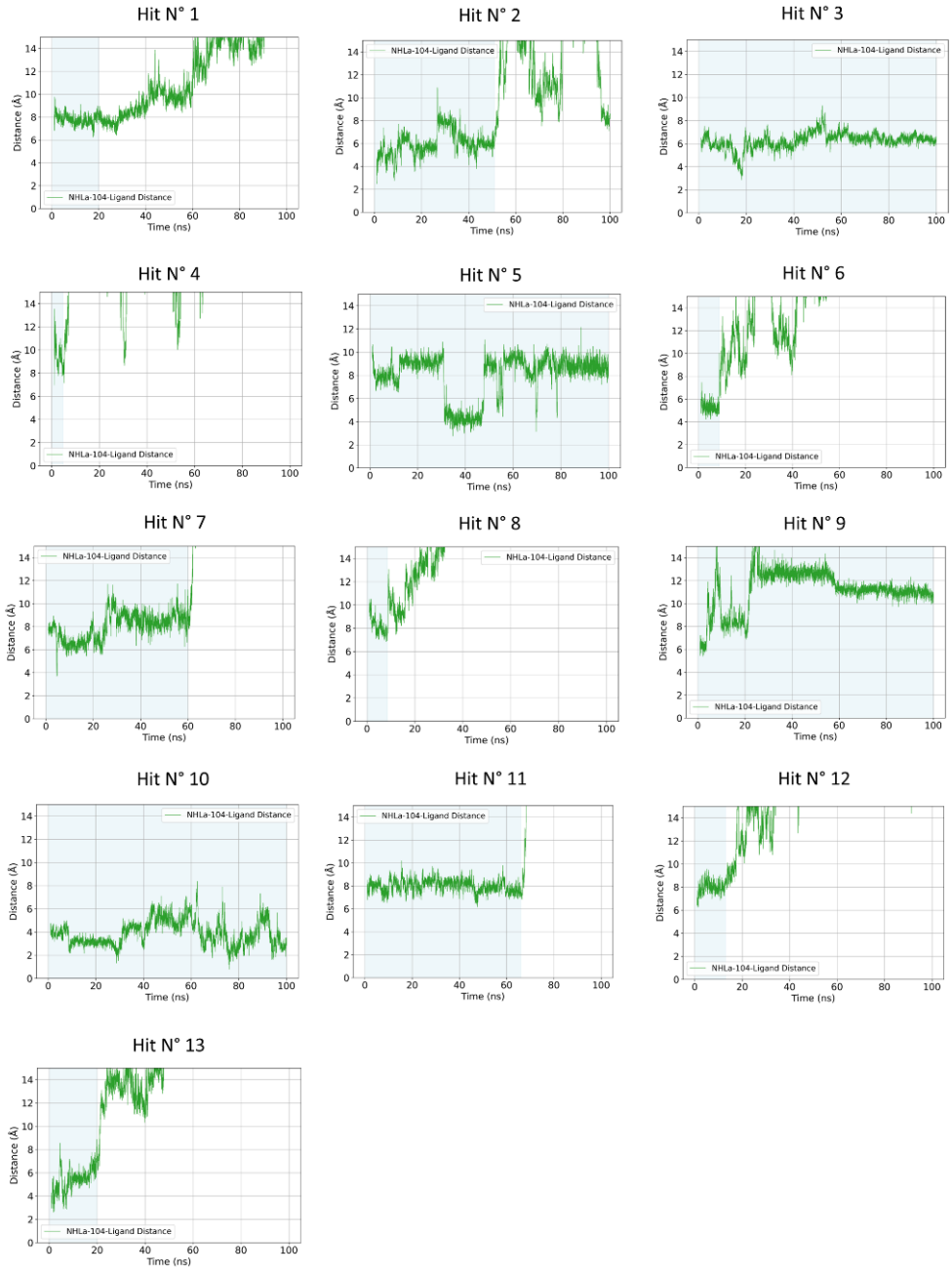

Fig. 27. Distance analysis. *The figure shows the distance of each molecule from the binding pocket throughout the entire simulation. The highlighted portion of the graph represents the period during which the binding interaction is maintained.*

Among the remaining 3 hits, # 3 demonstrated the most stabilizing effect on p17. Indeed, hit # 3 fluctuations decreased during the simulation until stabilization. Also, NHLa-104 itself remained stable during the simulation, indicating that the binding of hit # 3 did not induce structural destabilization of the protein (**Fig. 28**). H-bond analysis revealed that hit # 3 interacted via H-bonding

with the functional residue Arg15 for 56% of the simulation, highlighting its potential as an inhibitor that could interfere with the binding of vp17 to its receptors.

Focusing on the physicochemical properties of the 3 validated hits, they share a similar number of H-bond donors and acceptors, moderate flexibility due to the presence of rotatable bonds and ring structures that likely facilitate hydrophobic interactions (see **Fig. 25C** above) and make them favourable candidates for stable interactions with the protein target.

In conclusion, these data provide valuable guidelines for future drug repositioning studies on larger drug libraries, such as those approved by the FDA. The structural features here identified will aid in grouping molecules with similar properties. Additionally, a consensus analysis incorporating the remaining two protein variants will enable the identification of a common molecule capable of accommodating the diverse features presented by the different mutations of vp17s.

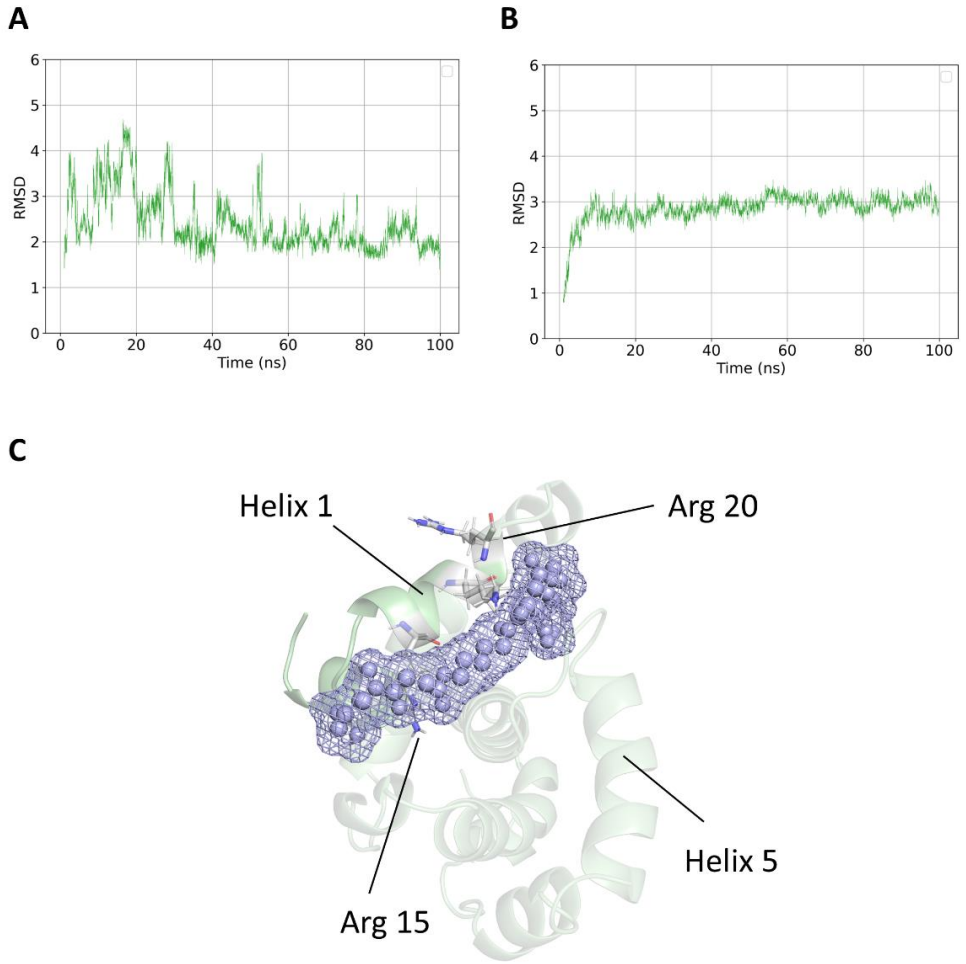

Fig. 28. Characterization of hit #3 interaction within the BP. *RMSD levels of the protein (A) and the ligand (B) throughout the simulation. (C): schematic representation of the binding pose of molecule 3 within the pocket.*

5. CONCLUDING REMARKS

HIV was first identified in the early 1980s, marking the beginning of a global health crisis, starting countless researched aimed at exploring the virus's replication mechanisms, genomic variability, and interactions with host cells.

Although ART has practically transformed HIV infection from a fatal disease into a manageable chronic condition, HIV's ability to persist in latent reservoirs and to evolve originating variants endowed with different pathological potential, continues to pose significant medical challenges. In particular, HIV oncogenic potential (i.e. non-Hodgkin lymphomas) continue to be a significant concern.

P17 is a critical component of the virus, serving both structural and regulatory functions pivotal for the virus's life cycle, its interactions with host cells and the arise of AIDS-associated tumors. In particular, vp17s has been identified that promote B-cell growth and consequent lymphoma development and metastasis. The understanding of this process supports the need for targeted therapeutic interventions to inhibit vp17s pathogenic potential and advance novel strategies against HIV-associated lymphoma. The research conducted in this thesis has given a contribution to the understanding of the implications of HIV-1 vp17s in HIV-associated lymphomagenesis and the strategies to identify new dedicated drugs.

Unlike refp17, vp17s are characterized by specific aa insertions at their COOH-terminal, which confer enhanced B-cell growth and clonogenic activity. These variants were predominantly found in HIV-1 positive patients with lymphomas, suggesting a strong association between vp17s and cancer progression in the context of HIV. The presence of vp17s has been linked to a higher frequency of non-Hodgkin lymphomas, underscoring their role as a contributing factor to oncogenesis and the necessity for a more in-depth understanding of the molecular mechanisms that contribute to the development of these malignancies.

In this thesis, the various studies were conducted with a sequential approach. Initial investigations focused on understanding how vp17s emerge from genomic alterations within the HIV-1 genome. Extensive sequence analyses identified mutation hotspots within the Gag gene, particularly in the MA domain. The presence of palindromic sequences and a high adenine content was noted as factors that may contribute to the error-prone nature of HIV reverse transcriptase, facilitating mutations and recombination events. This genomic instability was hypothesized to drive the generation of vp17s, a hypothesis that was experimentally confirmed by detecting frequent recombination events at these sites.

Following the genomic findings, MD simulations were performed to explore how these mutations affected the structure of vp17s. Key residues, such as Trp16 and Tyr29, were found to undergo significant conformational changes, destabilizing the hydrophobic core and leading to the exposure of functional epitopes. This exposure is crucial, as it enables vp17s to interact with and activate proliferative signaling pathways in B cells, providing a mechanistic explanation for their enhanced clonogenic activity observed in earlier studies. To validate these structural predictions, site-specific mutations were introduced in the refp17 to mimic the structural changes observed in vp17s. Experimental data confirmed that these mutations replicated the clonogenic effects, reinforcing the importance of the identified key residues. This logical progression, from genomic analysis to functional validation, established a clear link between specific mutations, structural instability, and the clonogenic potential of vp17s.

Based on these molecular insights, the research then focused on therapeutic interventions, exploring drug repositioning as a strategic approach to inhibit the biological function of vp17s. Given the structural vulnerabilities of vp17s, molecular docking and dynamics simulations were employed to screen existing drug libraries for compounds that could bind to key residues exposed by the

conformational changes in vp17s. Promising drug candidates were validated for their binding, offering a potential pathway for therapeutic intervention.

In conclusion, this thesis has provided a comprehensive exploration of the structural and functional dynamics of HIV-1 vp17s, identifying key vulnerabilities that can be considered druggable targets for novel therapeutic intervention. By an appropriate integration of genomic, structural, and computational insights, this work has elucidated the molecular mechanisms underlying vp17 evolution and function, offering new perspectives on the role of these variants in HIV-associated cancer.

The findings represent a baseline for the development of novel drug repositioning strategies that could improve the clinical management of HIV-related lymphomas. Future research should aim to refine these therapeutic candidates, optimize AI-based models for other viral proteins, and expand the integration of computational and experimental techniques to accelerate the discovery of effective treatments against virus-induced malignancies.

Relevant to the future developments of these studies, I spent a research period at the University of Cambridge, focusing on integrating MD simulations with advanced AI methodologies to enhance the predictive capabilities of protein structure modeling and their impact in pathogenesis and drug development-oriented studies.

6. BIBLIOGRAPHY

Gonda MA, Wong-Staal F, Gallo RC, Clements JE, Narayan O, Gilden RV. Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. Science. 1985;227(4683):173-177. doi:10.1126/science.2981428.

Myers, G., Pavlakis, G.N. (1992). Evolutionary Potential of Complex Retroviruses. In: Levy, J.A. (eds) The Retroviridae. The Viruses. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-3372-6_3

Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. Strand transfer events during HIV-1 reverse transcription. Virus Res. 2008;134(1-2):19-38. doi:10.1016/j.virusres.2007.12.017.

La Placa M. Principi di microbiologia medica, 524-548; 561-569; 665-679, ed. Esculapio.

Re M.C. Il laboratorio virologico nella diagnosi di infezione da HIV e nel monitoraggio del paziente infetto (disponibile su: http//www.bioetica-vssp.it/temi.html).

Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem. 1998; 67:1-25. doi:10.1146/annurev.biochem.67.1.1

Macher A. M. (1988). The pathology of AIDS. Public health reports (Washington, D.C. : 1974), 103(3), 246–254.

Seddon J, Bhagani S. Antimicrobial therapy for the treatment of opportunistic infections in HIV/AIDS patients: a critical appraisal. HIV AIDS (Auckl). 2011;3:19-33. doi:10.2147/HIV.S9274.

Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE, Quinn TC, Chadwick K, Margolick J, Brookmeyer R, Gallant J, Markowitz M, Ho DD, Richman DD, Siliciano RF. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science. 1997 Nov 14;278(5341):1295-300. doi: 10.1126/science.278.5341.1295.

Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841. doi:10.1101/cshperspect.a006841

Cohen J. Virology. Early AIDS virus may have ridden Africa's rails. Science. 2014;346(6205):21-22. doi:10.1126/science.346.6205.21.

Freed EO, Martin MA. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol. 1996;70(1):341-351. doi:10.1128/JVI.70.1.341-351.1996

Zhu P, Liu J, Bess J Jr, Chertova E, Lifson JD, Grisé H, Ofek GA, Taylor KA, Roux KH. Distribution and three-dimensional structure of AIDS virus envelope spikes. Nature. 2006 Jun 15;441(7095):847-52. doi: 10.1038/nature04817.

Ganser-Pornillos BK, Yeager M, Sundquist WI. The structural biology of HIV assembly. Curr Opin Struct Biol. 2008;18(2):203-217. doi:10.1016/j.sbi.2008.02.001

Freed EO. HIV-1 assembly, release and maturation. Nat Rev Microbiol. 2015;13(8):484-496. doi:10.1038/nrmicro3490.

Muriaux D, Darlix JL. Properties and functions of the nucleocapsid protein in virus assembly. RNA Biol. 2010;7(6):744-753. doi:10.4161/rna.7.6.14065

Peng C, Ho BK, Chang TW, Chang NT. Role of human immunodeficiency virus type 1-specific protease in core protein maturation and viral infectivity. J Virol. 1989;63(6):2550-2556. doi:10.1128/JVI.63.6.2550-2556.1989

Ohishi M, Nakano T, Sakuragi S, Shioda T, Sano K, Sakuragi J. The relationship between HIV-1 genome RNA dimerization, virion maturation and infectivity. Nucleic Acids Res. 2011;39(8):3404-3417. doi:10.1093/nar/gkq1314

Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. Nat Rev Genet. 2004;5(1):52-61. doi:10.1038/nrg1246

Quinn TC. Population migration and the spread of types 1 and 2 human immunodeficiency viruses. Proc Natl Acad Sci U S A. 1994;91(7):2407-2414. doi:10.1073/pnas.91.7.2407

Hahn BH, Shaw GM, De Cock KM, Sharp PM. AIDS as a zoonosis: scientific and public health implications. Science. 2000;287(5453):607-614. doi:10.1126/science.287.5453.607

Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med. 2011;1(1):a006841. doi:10.1101/cshperspect.a006841

Plantier JC, Leoz M, Dickerson JE, De Oliveira F, Cordonnier F, Lemée V, Damond F, Robertson DL, Simon F. A new human immunodeficiency virus derived from gorillas. Nat Med. 2009 Aug;15(8):871-2. doi: 10.1038/nm.2016.

Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. Annu Rev Biochem. 1998;67:1-25. doi:10.1146/annurev.biochem.67.1.1

Basu VP, Song M, Gao L, Rigby ST, Hanson MN, Bambara RA. Strand transfer events during HIV-1 reverse transcription. Virus Res. 2008;134(1-2):19-38. doi:10.1016/j.virusres.2007.12.017

Wilusz J. Putting an 'End' to HIV mRNAs: capping and polyadenylation as potential therapeutic targets. AIDS Res Ther. 2013;10(1):31. Published 2013 Dec 13. doi:10.1186/1742-6405-10-31

Ocwieja KE, Sherrill-Mix S, Mukherjee R, Custers-Allen R, David P, Brown M, Wang S, Link DR, Olson J, Travers K, Schadt E, Bushman FD. Dynamic regulation of HIV-1 mRNA populations analyzed by single-molecule enrichment and long-read sequencing. Nucleic Acids Res. 2012 Nov 1;40(20):10345-55. doi: 10.1093/nar/gks753.

Van Heuverswyn F, Li Y, Neel C, Bailes E, Keele BF, Liu W, Loul S, Butel C, Liegeois F, Bienvenue Y, Ngolle EM, Sharp PM, Shaw GM, Delaporte E, Hahn BH, Peeters M. Human immunodeficiency viruses: SIV infection in wild gorillas. Nature. 2006 Nov 9;444(7116):164. doi: 10.1038/444164a.

Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull. 2001;58:19-42. doi:10.1093/bmb/58.1.19

Goodenow M, Huet T, Saurin W, Kwok S, Sninsky J, Wain-Hobson S. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. J Acquir Immune Defic Syndr (1988). 1989;2(4):344-352

Li H, Bar KJ, Wang S, Decker JM, Chen Y, Sun C, Salazar-Gonzalez JF, Salazar MG, Learn GH, Morgan CJ, Schumacher JE, Hraber P, Giorgi EE, Bhattacharya T, Korber BT, Perelson AS, Eron JJ, Cohen MS, Hicks CB, Haynes BF, Markowitz M, Keele BF, Hahn BH, Shaw GM. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. PLoS Pathog. 2010 May 13;6(5):e1000890. doi: 10.1371/journal.ppat.1000890

Wymant C, Blanquart F, Golubchik T, Gall A, Bakker M, Bezemer D, Croucher NJ, Hall M, Hillebregt M, Ong SH, Ratmann O, Albert J, Bannert N, Fellay J, Fransen K, Gourlay A, Grabowski MK, Gunsenheimer-Bartmeyer B, Günthard HF, Kivelä P, Kouyos R, Laeyendecker O, Liitsola K, Meyer L, Porter K, Ristola M, van Sighem A, Berkhout B, Cornelissen M, Kellam P, Reiss P, Fraser C; BEEHIVE Collaboration. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. Virus Evol. 2018 May 18;4(1):vey007. doi: 10.1093/ve/vey007

Ji JP, Loeb LA. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochemistry. 1992;31(4):954-958. doi:10.1021/bi00119a002

Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. J Virol. 2010;84(19):9864-9878. doi:10.1128/JVI.00915-10

Simon-Loriere E, Rossolillo P, Negroni M. RNA structures, genomic organization and selection of recombinant HIV. RNA Biol. 2011;8(2):280-286. doi:10.4161/rna.8.2.15193

Shankarappa R, Margolick JB, Gange SJ, Rodrigo AG, Upchurch D, Farzadegan H, Gupta P, Rinaldo CR, Learn GH, He X, Huang XL, Mullins JI. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. J Virol. 1999 Dec;73(12):10489-502. doi: 10.1128/JVI.73.12.10489-10502.1999.

Goldschmidt V, Rigourd M, Ehresmann C, Le Grice SF, Ehresmann B, Marquet R. Direct and indirect contributions of RNA secondary structure elements to the initiation of HIV-1 reverse transcription. J Biol Chem. 2002;277(45):43233-43242. doi:10.1074/jbc.M205295200

Liu Y, Jia L, Su B, Li H, Li Z, Han J, Zhang Y, Zhang T, Li T, Wu H, Li J, Li L. The Genetic Diversity of HIV-1 Quasispecies Within Primary Infected Individuals. AIDS Res Hum Retroviruses. 2020 May;36(5):440-449. doi: 10.1089/AID.2019.0242.

Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. Microbiol Mol Biol Rev. 2012;76(2):159-216. doi:10.1128/MMBR.05023-11

Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull. 2001;58:19-42. doi:10.1093/bmb/58.1.19

Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. J Virol. 1995;69(8):5087-5094. doi:10.1128/JVI.69.8.5087-5094.1995

Hernández-Ramírez, R. U., Qin, L., Lin, H., Leyden, W., Neugebauer, R. S., Althoff, K. N., Hessol, N. A., Achenbach, C. J., Brooks, J. T., Gill, M. J., Grover, S., Horberg, M. A., Li, J., Mathews, W. C., Mayor, A. M., Patel, P., Rabkin, C. S., Rachlis, A., Justice, A. C., Moore, R. D., … North American AIDS Cohort Collaboration on Research and Design of the International Epidemiologic Databases to Evaluate AIDS (2020). Association of Immunosuppression and Human Immunodeficiency Virus (HIV) Viremia With Anal Cancer Risk in Persons Living With HIV in the United States and Canada. Clin Infect Dis. 2020;70(6):1176-1185. doi:10.1093/cid/ciz329

Bedimo, R., Chen, R. Y., Accortt, N. A., Raper, J. L., Linn, C., Allison, J. J., Dubay, J., Saag, M. S., & Hoesley, C. J. (2004). Trends in AIDS-defining and non-AIDS-defining malignancies among HIV-infected patients: 1989-2002. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 39(9), 1380–1384. https://doi.org/10.1086/424883

Coté TR, Biggar RJ, Rosenberg PS, Devesa SS, Percy C, Yellin FJ, Lemp G, Hardy C, Geodert JJ, Blattner WA. Non-Hodgkin's lymphoma among people with AIDS: incidence, presentation and public health burden. AIDS/Cancer Study Group. Int J Cancer. 1997 Nov 27;73(5):645-50. doi: 10.1002/(sici)1097-0215(19971127)73:5<645::aid-ijc6>3.0.co;2-x

Engels EA, Goedert JJ. Human immunodeficiency virus/acquired immunodeficiency syndrome and cancer: past, present, and future. J Natl Cancer Inst. 2005 Mar 16;97(6):407-9. doi: 10.1093/jnci/dji085.

Engels EA, Biggar RJ, Hall HI, Cross H, Crutchfield A, Finch JL, Grigg R, Hylton T, Pawlish KS, McNeel TS, Goedert JJ. Cancer risk in people infected with human immunodeficiency virus in the United States. Int J Cancer. 2008 Jul 1;123(1):187-94. doi: 10.1002/ijc.23487

Gibson TM, Morton LM, Shiels MS, Clarke CA, Engels EA. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. AIDS. 2014 Sep 24;28(15):2313-8. doi: 10.1097/QAD.0000000000000428. PMID: 25111081; PMCID: PMC4260326.

Besson C, Lancar R, Prevot S, Algarte-Genin M, Delobel P, Bonnet F, Meyohas MC, Partisani M, Oberic L, Gabarre J, Goujard C, Boue F, Coppo P, Costello R, Hendel-Chavez H, Mekerri N, Dos Santos G, Recher C, Delarue R, Casasnovas RO, Taoufik Y, Mounier N, Costagliola D; ANRS-CO16 LYMPHOVIR Cohort. Outcomes for HIV-associated diffuse large B-cell lymphoma in the modern combined antiretroviral therapy era. AIDS. 2017 Nov 28;31(18):2493-2501. doi: 10.1097/QAD.0000000000001652

Shannon-Lowe C, Rickinson AB, Bell AI. Epstein-Barr virus-associated lymphomas. Philos Trans R Soc Lond B Biol Sci. 2017 Oct 19;372(1732):20160271. doi: 10.1098/rstb.2016.0271.

Clifford GM, Polesel J, Rickenbach M, Dal Maso L, Keiser O, Kofler A, Rapiti E, Levi F, Jundt G, Fisch T, Bordoni A, De Weck D, Franceschi S; Swiss HIV Cohort. Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy. J Natl Cancer Inst. 2005 Mar 16;97(6):425-32. doi: 10.1093/jnci/dji072

Dal Maso L, Polesel J, Serraino D, Lise M, Piselli P, Falcini F, Russo A, Intrieri T, Vercelli M, Zambon P, Tagliabue G, Zanetti R, Federico M, Limina RM, Mangone L, De Lisi V, Stracci F, Ferretti S, Piffer S, Budroni M, Donato A, Giacomin A, Bellù F, Fusco M, Madeddu A, Vitarelli S, Tessandori R, Tumino R, Suligoi B, Franceschi S; Cancer and AIDS Registries Linkage (CARL) Study. Pattern of cancer risk in persons with AIDS in Italy in the HAART era. Br J Cancer. 2009 Mar 10;100(5):840-7. doi: 10.1038/sj.bjc.6604923

Patel K, Hernán MA, Williams PL, Seeger JD, McIntosh K, Van Dyke RB, Seage GR 3rd; Pediatric AIDS Clinical Trials Group 219/219C Study Team. Long-term effectiveness of highly active antiretroviral therapy on the survival of children and adolescents with HIV infection: a 10-year follow-up study. Clin Infect Dis. 2008 Feb 15;46(4):507-15. doi: 10.1086/526524

Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet. 2007 Jul 7;370(9581):59-67. doi: 10.1016/S0140-6736(07)61050-2. PMID: 17617273.

Dal Maso L, Franceschi S, Polesel J, Braga C, Piselli P, Crocetti E, Falcini F, Guzzinati S, Zanetti R, Vercelli M, Rezza G; Cancer and AIDS Registy Linkage Study. Risk of cancer in persons with AIDS in Italy, 1985-1998. Br J Cancer. 2003 Jul 7;89(1):94-100. doi: 10.1038/sj.bjc.6601017

Kramer JR, Giordano TP, Souchek J, Richardson P, Hwang LY, El-Serag HB. The effect of HIV coinfection on the risk of cirrhosis and hepatocellular carcinoma in U.S. veterans with hepatitis C. Am J Gastroenterol. 2005 Jan;100(1):56-63. doi: 10.1111/j.1572-0241.2005.40670.x

Hessol NA, Pipkin S, Schwarcz S, Cress RD, Bacchetti P, Scheer S. The impact of highly active antiretroviral therapy on non-AIDS-defining cancers among adults with AIDS. Am J Epidemiol. 2007 May 15;165(10):1143-53. doi: 10.1093/aje/kwm017

Saad JS, Miller J, Tai J, Kim A, Ghanam RH, Summers MF. Structural basis for targeting HIV-1 Gag proteins to the plasma membrane for virus assembly. Proc Natl Acad Sci U S A. 2006 Jul 25;103(30):11364-9. doi: 10.1073/pnas.0602818103

Freed EO, Martin MA. Domains of the human immunodeficiency virus type 1 matrix and gp41 cytoplasmic tail required for envelope incorporation into virions. J Virol. 1996 Jan;70(1):341-51. doi: 10.1128/JVI.70.1.341-351.1996

Bukrinsky MI, Haggerty S, Dempsey MP, Sharova N, Adzhubel A, Spitz L, Lewis P, Goldfarb D, Emerman M, Stevenson M. A nuclear localization signal within HIV-1 matrix protein that governs infection of non-dividing cells. Nature. 1993 Oct 14;365(6447):666-9. doi: 10.1038/365666a0

Dong X, Li H, Derdowski A, Ding L, Burnett A, Chen X, Peters TR, Dermody TS, Woodruff E, Wang JJ, Spearman P. AP-3 directs the intracellular trafficking of HIV-1 Gag and plays a key role in particle assembly. Cell. 2005 Mar 11;120(5):663-74. doi: 10.1016/j.cell.2004.12.023

Matthews S, Barlow P, Boyd J, Barton G, Russell R, Mills H, Cunningham M, Meyers N, Burns N, Clark N, et al. Structural similarity between the p17 matrix protein of HIV-1 and interferon-gamma. Nature. 1994 Aug 25;370(6491):666-8. doi: 10.1038/370666a0.

De Francesco MA, Baronio M, Fiorentini S, Signorini C, Bonfanti C, Poiesi C, Popovic M, Grassi M, Garrafa E, Bozzo L, Lewis GK, Licenziati S, Gallo RC, Caruso A. HIV-1 matrix protein p17 increases the production of proinflammatory cytokines and counteracts IL-4 activity by binding to a cellular receptor. Proc Natl Acad Sci U S A. 2002 Jul 23;99(15):9972-7. doi: 10.1073/pnas.142274699

De Francesco MA, Caruso A, Fallacara F, Canaris AD, Dima F, Poiesi C, Licenziati S, Corulli M, Martinelli F, Fiorentini S, Turano A. HIV p17 enhances lymphocyte proliferation and HIV-1 replication after binding to a human serum factor. AIDS. 1998 Feb 12;12(3):245-52. doi: 10.1097/00002030-199803000-00001

Vitale M, Caruso A, De Francesco MA, Rodella L, Bozzo L, Garrafa E, Grassi M, Gobbi G, Cacchioli A, Fiorentini S. HIV-1 matrix protein p17 enhances the proliferative activity of natural killer cells and increases their ability to secrete proinflammatory cytokines. Br J Haematol. 2003 Jan;120(2):337-43. doi: 10.1046/j.1365-2141.2003.04053.x.

Marini E, Tiberio L, Caracciolo S, Tosti G, Guzman CA, Schiaffonati L, Fiorentini S, Caruso A. HIV-1 matrix protein p17 binds to monocytes and selectively stimulates MCP-1 secretion: role of transcriptional factor AP-1. Cell Microbiol. 2008 Mar;10(3):655-66. doi: 10.1111/j.1462-5822.2007.01073.x.

Giagulli C, Marsico S, Magiera AK, Bruno R, Caccuri F, Barone I, Fiorentini S, Andò S, Caruso A. Opposite effects of HIV-1 p17 variants on PTEN activation and cell growth in B cells. PLoS One. 2011 Mar 14;6(3):e17831. doi: 10.1371/journal.pone.0017831.

Fiorentini S, Riboldi E, Facchetti F, Avolio M, Fabbri M, Tosti G, Becker PD, Guzman CA, Sozzani S, Caruso A. HIV-1 matrix protein p17 induces human plasmacytoid dendritic cells to acquire a migratory immature cell phenotype. Proc Natl Acad Sci U S A. 2008 Mar 11;105(10):3867-72. doi: 10.1073/pnas.0800370105.

Giagulli C, Magiera AK, Bugatti A, Caccuri F, Marsico S, Rusnati M, Vermi W, Fiorentini S, Caruso A. HIV-1 matrix protein p17 binds to the IL-8 receptor CXCR1 and shows IL-8-like chemokine activity on monocytes through Rho/ROCK activation. Blood. 2012 Mar 8;119(10):2274-83. doi: 10.1182/blood-2011-06-364083.

Caccuri F, Giagulli C, Bugatti A, Benetti A, Alessandri G, Ribatti D, Marsico S, Apostoli P, Slevin MA, Rusnati M, Guzman CA, Fiorentini S, Caruso A. HIV-1 matrix protein p17 promotes angiogenesis via chemokine receptors CXCR1 and CXCR2. Proc Natl Acad Sci U S A. 2012 Sep 4;109(36):14580-5. doi: 10.1073/pnas.1206605109.

Poiesi C, De Francesco MA, Baronio M, Manca N. HIV-1 p17 binds heparan sulfate proteoglycans to activated CD4(+) T cells. Virus Res. 2008 Mar;132(1-2):25-32. doi: 10.1016/j.virusres.2007.10.006.

Bugatti A, Giagulli C, Urbinati C, Caccuri F, Chiodelli P, Oreste P, Fiorentini S, Orro A, Milanesi L, D'Ursi P, Caruso A, Rusnati M. Molecular interaction studies of HIV-1 matrix protein p17 and heparin: identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists. J Biol Chem. 2013 Jan 11;288(2):1150-61. doi: 10.1074/jbc.M112.400077

Popovic M, Tenner-Racz K, Pelser C, Stellbrink HJ, van Lunzen J, Lewis G, Kalyanaraman VS, Gallo RC, Racz P. Persistence of HIV-1 structural proteins and glycoproteins in lymph nodes of patients under highly active antiretroviral therapy. Proc Natl Acad Sci U S A. 2005 Oct 11;102(41):14807-12. doi: 10.1073/pnas.0506857102.

Fiorentini S, Marsico S, Becker PD, Iaria ML, Bruno R, Guzmán CA, Caruso A. Synthetic peptide AT20 coupled to KLH elicits antibodies against a conserved conformational epitope from a major functional area of the HIV-1 matrix protein p17. Vaccine. 2008 Aug 26;26(36):4758-65. doi: 10.1016/j.vaccine.2008.06.082

Fiorentini S, Marini E, Bozzo L, Trainini L, Saadoune L, Avolio M, Pontillo A, Bonfanti C, Sarmientos P, Caruso A. Preclinical studies on immunogenicity of the HIV-1 p17-based synthetic peptide AT20-KLH. Biopolymers. 2004;76(4):334-43. doi: 10.1002/bip.20130.

Iaria ML, Fiorentini S, Focà E, Zicari S, Giagulli C, Caccuri F, Francisci D, Di Perri G, Castelli F, Baldelli F, Caruso A. Synthetic HIV-1 matrix protein p17-based AT20-KLH therapeutic immunization in HIV-1-infected patients receiving antiretroviral treatment: A phase I safety and immunogenicity study. Vaccine. 2014 Feb 19;32(9):1072-8. doi: 10.1016/j.vaccine.2013.12.051

Renga B, Francisci D, Schiaroli E, Carino A, Cipriani S, D'Amore C, Sidoni A, Sordo RD, Ferri I, Lucattelli M, Lunghi B, Baldelli F, Fiorucci S. The HIV matrix protein p17 promotes the activation of human hepatic stellate cells through interactions with CXCR2 and Syndecan-2. PLoS One. 2014 Apr 15;9(4):e94798. doi: 10.1371/journal.pone.0094798

Bugatti A, Giagulli C, Urbinati C, Caccuri F, Chiodelli P, Oreste P, Fiorentini S, Orro A, Milanesi L, D'Ursi P, Caruso A, Rusnati M. Molecular interaction studies of HIV-1 matrix protein p17 and heparin: identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists. J Biol Chem. 2013 Jan 11;288(2):1150-61. doi: 10.1074/jbc.M112.400077

Caccuri F, Giagulli C, Reichelt J, Martorelli D, Marsico S, Bugatti A, Barone I, Rusnati M, Guzman CA, Dolcetti R, Caruso A. Simian immunodeficiency virus and human immunodeficiency virus type 1 matrix proteins specify different capabilities to modulate B cell growth. J Virol. 2014 May;88(10):5706-17. doi: 10.1128/JVI.03142-13

Giagulli C, D'Ursi P, He W, Zorzan S, Caccuri F, Varney K, Orro A, Marsico S, Otjacques B, Laudanna C, Milanesi L, Dolcetti R, Fiorentini S, Lu W, Caruso A. A single amino acid substitution confers B-cell clonogenic activity to the HIV-1 matrix protein p17. Sci Rep. 2017 Jul 26;7(1):6555. doi: 10.1038/s41598-017-06848-y

Caccuri F, Iaria ML, Campilongo F, Varney K, Rossi A, Mitola S, Schiarea S, Bugatti A, Mazzuca P, Giagulli C, Fiorentini S, Lu W, Salmona M, Caruso A. Cellular aspartyl proteases promote the unconventional secretion of biologically active HIV-1 matrix protein p17. Sci Rep. 2016 Dec 1;6:38027. doi: 10.1038/srep38027

Totonchy J, Cesarman E. Does persistent HIV replication explain continued lymphoma incidence in the era of effective antiretroviral therapy? Curr Opin Virol. 2016 Oct;20:71-77. doi: 10.1016/j.coviro.2016.09.001

Bugatti A, Giagulli C, Urbinati C, Caccuri F, Chiodelli P, Oreste P, Fiorentini S, Orro A, Milanesi L, D'Ursi P, Caruso A, Rusnati M. Molecular interaction studies of HIV-1 matrix protein p17 and heparin:

identification of the heparin-binding motif of p17 as a target for the development of multitarget antagonists. J Biol Chem. 2013 Jan 11;288(2):1150-61. doi: 10.1074/jbc.M112.400077

Fiorentini S, Giagulli C, Caccuri F, Magiera AK, Caruso A. HIV-1 matrix protein p17: a candidate antigen for therapeutic vaccines against AIDS. Pharmacol Ther. 2010 Dec;128(3):433-44. doi: 10.1016/j.pharmthera.2010.08.005

Gayathiri E, Prakash P, Kumaravel P, Jayaprakash J, Ragunathan MG, Sankar S, Pandiaraj S, Thirumalaivasan N, Thiruvengadam M, Govindasamy R. Computational approaches for modeling and structural design of biological systems: A comprehensive review. Prog Biophys Mol Biol. 2023 Dec;185:17-32. doi: 10.1016/j.pbiomolbio.2023.08.002

Cheifet B. Where is genomics going next? Genome Biol. 2019 Jan 22;20(1):17. doi: 10.1186/s13059-019-1626-2

Van Drie JH. Computer-aided drug design: the next 20 years. J Comput Aided Mol Des. 2007 Oct-Nov;21(10-11):591-601. doi: 10.1007/s10822-007-9142-y

Salmaso V, Sturlese M, Cuzzolin A, Moro S. Combining self- and cross-docking as benchmark tools: the performance of DockBench in the D3R Grand Challenge 2. J Comput Aided Mol Des. 2018 Jan;32(1):251-264. doi: 10.1007/s10822-017-0051-4

Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. Neuron. 2018 Sep 19;99(6):1129-1143. doi: 10.1016/j.neuron.2018.08.011

Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. Molecular dynamics studies on the HIV-1 integrase catalytic domain. Biophys J. 1999 Jun;76(6):2999-3011. doi: 10.1016/s0006-3495(99)77453-9

Pietrucci F, Vargiu AV, Kranjc A. HIV-1 Protease Dimerization Dynamics Reveals a Transient Druggable Binding Pocket at the Interface. Sci Rep. 2015 Dec 22;5:18555. doi: 10.1038/srep18555

Salmaso V, Moro S. Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. Front Pharmacol. 2018 Aug 22;9:923. doi: 10.3389/fphar.2018.00923

Perryman AL, Lin JH, McCammon JA. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci. 2004 Apr;13(4):1108-23. doi: 10.1110/ps.03468904

Candresse T, Filloux D, Muhire B, Julian C, Galzi S, Fort G, Bernardo P, Daugrois JH, Fernandez E, Martin DP, Varsani A, Roumagnac P. Appearances can be deceptive: revealing a hidden viral infection with deep sequencing in a plant quarantine context. PLoS One. 2014 Jul 25;9(7):e102945. doi: 10.1371/journal.pone.0102945

Agu PC, Afiukwa CA, Orji OU, Ezeh EM, Ofoke IH, Ogbu CO, Ugwuja EI, Aja PM. Molecular docking as a tool for the discovery of molecular targets of nutraceuticals in diseases management. Sci Rep. 2023 Aug 17;13(1):13398. doi: 10.1038/s41598-023-40160-2

Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. 2011 Jun;7(2):146-57. doi: 10.2174/157340911795677602

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem. 2009 Dec;30(16):2785-91. doi: 10.1002/jcc.21256

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD. Improved protein-ligand docking using GOLD. Proteins. 2003 Sep 1;52(4):609-23. doi: 10.1002/prot.10465

Kharkar PS, Warrier S, Gaud RS. Reverse docking: a powerful tool for drug repositioning and drug rescue. Future Med Chem. 2014 Mar;6(3):333-42. doi: 10.4155/fmc.13.207

Chaudhary M, and Tyagi K. (2024); A REVIEW ON MOLECULAR DOCKING AND ITS APPLICATION Int. J. of Adv. Res. (Mar). 1141-1153] (ISSN 2320-5407). doi: 10.21474/IJAR01/18505

Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. J Chem Inf Model. 2021 Aug 23;61(8):3891-3898. doi: 10.1021/acs.jcim.1c00203

Shanno D.F. Conditioning of Quasi-Newton Methods for Function Minimization. Mathematics of Computation, 1970, 24, 647-656. https://doi.org/10.1090/S0025-5718-1970-0274029-X

Kräutler, V., Gunsteren, W.F., & Hünenberger, P.H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. 2001, Journal of Computational Chemistry, 22.

Hess, B., Bekker, H., Berendsen, H.J., & Fraaije, J.G. LINCS: A linear constraint solver for molecular simulations. 1997, Journal of Computational Chemistry, 18.

Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain χ(1) and χ(2) dihedral angles. J Chem Theory Comput. 2012 Sep 11;8(9):3257-3273. doi: 10.1021/ct300400x.

Robertson MJ, Tirado-Rives J, Jorgensen WL. Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field. J Chem Theory Comput. 2015 Jul 14;11(7):3499-509. doi: 10.1021/acs.jctc.5b00356

Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migues AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. J Chem Theory Comput. 2020 Jan 14;16(1):528-552. doi: 10.1021/acs.jctc.9b00591

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem. 2004 Oct;25(13):1656-76. doi: 10.1002/jcc.20090

Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. Adv Appl Bioinform Chem. 2015 Nov 19;8:37-47. doi: 10.2147/AABC.S70333

Salmaso V, Sturlese M, Cuzzolin A, Moro S. Combining self- and cross-docking as benchmark tools: the performance of DockBench in the D3R Grand Challenge 2. J Comput Aided Mol Des. 2018 Jan;32(1):251-264. doi: 10.1007/s10822-017-0051-4

Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, Pande VS. Accelerating molecular dynamic simulation on graphics processing units. J Comput Chem. 2009 Apr 30;30(6):864-72. doi: 10.1002/jcc.21209

Kästner J. Umbrella sampling. WIREs Comput Mol Sci, 2011, 1: 932-942. https://doi.org/10.1002/wcms.66

Qi R, Wei G, Ma B, Nussinov R. Replica Exchange Molecular Dynamics: A Practical Application Protocol with Solutions to Common Problems and a Peptide Aggregation and Self-Assembly Example. Methods Mol Biol. 2018;1777:101-119. doi: 10.1007/978-1-4939-7811-3_5.

Zhou R, Duan M. Metadynamics Simulations to Study the Structural Ensembles and Binding Processes of Intrinsically Disordered Proteins. Methods Mol Biol. 2022;2405:169-178. doi: 10.1007/978-1-0716-1855-4_9

Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004 Jun 22;120(24):11919-29. doi: 10.1063/1.1755656

Verli H, Calazans A, Brindeiro R, Tanuri A, Guimarães JA. Molecular dynamics analysis of HIV-1 matrix protein: clarifying differences between crystallographic and solution structures. J Mol Graph Model. 2007 Jul;26(1):62-8. doi: 10.1016/j.jmgm.2006.09.009

Monje-Galvan V, Voth GA. Binding mechanism of the matrix domain of HIV-1 gag on lipid membranes. Elife. 2020 Aug 18;9:e58621. doi: 10.7554/eLife.58621

Zou X, Morganella S, Glodzik D, Davies H, Li Y, Stratton MR, Nik-Zainal S. Short inverted repeats contribute to localized mutability in human somatic cells. Nucleic Acids Res. 2017 Nov 2;45(19):11213-11221. doi: 10.1093/nar/gkx731

Lovett ST. Encoded errors: mutations and rearrangements mediated by misalignment at repetitive DNA sequences. Mol Microbiol. 2004 Jun;52(5):1243-53. doi: 10.1111/j.1365-2958.2004.04076.x

Sun N, Yau SS. In-depth investigation of the point mutation pattern of HIV-1. Front Cell Infect Microbiol. 2022 Nov 15;12:1033481. doi: 10.3389/fcimb.2022.1033481

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res. 2000 Jan 1;28(1):235-42. doi: 10.1093/nar/28.1.235

Dolcetti R, Giagulli C, He W, Selleri M, Caccuri F, Eyzaguirre LM, Mazzuca P, Corbellini S, Campilongo F, Marsico S, Giombini E, Muraro E, Rozera G, De Paoli P, Carbone A, Capobianchi MR, Ippolito G, Fiorentini S, Blattner WA, Lu W, Gallo RC, Caruso A. Role of HIV-1 matrix protein p17 variants in lymphoma pathogenesis. Proc Natl Acad Sci U S A. 2015 Nov 17;112(46):14331-6. doi: 10.1073/pnas.1514748112

Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993 Dec 5;234(3):779-815. doi: 10.1006/jmbi.1993.1626

Laskowski RA, MacArthur, MW, Moss DS, Thornton JM. PROCHECK—A program to check the stereochemical quality of protein structures. J. App. Cryst. 1993, 26, 283–291.

Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics 2020, 36, 1765–1771.

Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Izadi S, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Mermelstein DJ, Merz KM, Miao Y, Monard G, Nguyen C, Nguyen H, Omelyan I, Onufriev A, Pan F, Qi R, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Smith J, Salomon-Ferrer R, Swails J, Walker RC, Wang J, Wei H, Wolf RM, Wu X, Xiao L, York DM and Kollman PA. Amber18; University of California, San Francisco, 2018.

Machado MR, Pantano S, Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations. J. Chem. Theory Comput. 2020, 16, 1367–1372.

Durrant JD, McCammon JA. HBonanza: a computer algorithm for molecular-dynamics-trajectory hydrogen-bond analysis. J Mol Graph Model. 2011 Nov;31:5-9. doi: 10.1016/j.jmgm.2011.07.008

Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys. 2004 Jun 22;120(24):11919-29. doi: 10.1063/1.1755656

Pierce LC, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. J Chem Theory Comput. 2012 Sep 11;8(9):2997-3002. doi: 10.1021/ct300284c

Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics. 2006 Mar 22;7:166. doi: 10.1186/1471-2105-7-166

Feichtinger H, Putkonen P, Parravicini C, Li SL, Kaaya EE, Böttiger D, Biberfeld G, Biberfeld P. Malignant lymphomas in cynomolgus monkeys infected with simian immunodeficiency virus. Am J Pathol. 1990 Dec;137(6):1311-5.

Rao Z, Belyaev AS, Fry E, Roy P, Jones IM, Stuart DI. Crystal structure of SIV matrix antigen and implications for virus assembly. Nature. 1995 Dec 14;378(6558):743-7. doi: 10.1038/378743a0

Galindo-Murillo R, Roe DR, Cheatham TE 3rd. Convergence and reproducibility in molecular dynamics simulations of the DNA duplex d(GCACGAACGAACGAACGC). Biochim Biophys Acta. 2015 May;1850(5):1041-1058. doi: 10.1016/j.bbagen.2014.09.007

De Francesco MA, Baronio M, Fiorentini S, Signorini C, Bonfanti C, Poiesi C, Popovic M, Grassi M, Garrafa E, Bozzo L, Lewis GK, Licenziati S, Gallo RC, Caruso A. HIV-1 matrix protein p17 increases the production of proinflammatory cytokines and counteracts IL-4 activity by binding to a cellular receptor. Proc Natl Acad Sci U S A. 2002 Jul 23;99(15):9972-7. doi: 10.1073/pnas.142274699

Caccuri F, Giagulli C, Bugatti A, Benetti A, Alessandri G, Ribatti D, Marsico S, Apostoli P, Slevin MA, Rusnati M, Guzman CA, Fiorentini S, Caruso A. HIV-1 matrix protein p17 promotes angiogenesis via chemokine receptors CXCR1 and CXCR2. Proc Natl Acad Sci U S A. 2012 Sep 4;109(36):14580-5. doi: 10.1073/pnas.1206605109

Dolcetti R, Giagulli C, He W, Selleri M, Caccuri F, Eyzaguirre LM, Mazzuca P, Corbellini S, Campilongo F, Marsico S, Giombini E, Muraro E, Rozera G, De Paoli P, Carbone A, Capobianchi MR, Ippolito G, Fiorentini S, Blattner WA, Lu W, Gallo RC, Caruso A. Role of HIV-1 matrix protein p17 variants in lymphoma pathogenesis. Proc Natl Acad Sci U S A. 2015 Nov 17;112(46):14331-6. doi: 10.1073/pnas.1514748112. Epub 2015 Nov 2. Erratum in: Proc Natl Acad Sci U S A. 2015 Dec 15;112(50):E7033. doi: 10.1073/pnas.1522501112

Fiorentini S, Giagulli C, Caccuri F, Magiera AK, Caruso A. HIV-1 matrix protein p17: a candidate antigen for therapeutic vaccines against AIDS. Pharmacol Ther. 2010 Dec;128(3):433-44. doi: 10.1016/j.pharmthera.2010.08.005

Iaria ML, Fiorentini S, Focà E, Zicari S, Giagulli C, Caccuri F, Francisci D, Di Perri G, Castelli F, Baldelli F, Caruso A. Synthetic HIV-1 matrix protein p17-based AT20-KLH therapeutic immunization in HIV-1-infected patients receiving antiretroviral treatment: A phase I safety and immunogenicity study. Vaccine. 2014 Feb 19;32(9):1072-8. doi: 10.1016/j.vaccine.2013.12.051

Focà E, Iaria ML, Caccuri F, Fiorentini S, Motta D, Giagulli C, Castelli F, Caruso A. Long-lasting humoral immune response induced in HIV-1-infected patients by a synthetic peptide (AT20) derived from the HIV-1 matrix protein p17 functional epitope. HIV Clin Trials. 2015 Aug;16(4):157-62. doi: 10.1179/1528433614Z.0000000018

Jourdan JP, Bureau R, Rochais C, Dallemagne P. Drug repositioning: a brief overview. J Pharm Pharmacol. 2020 Sep;72(9):1145-1151. doi: 10.1111/jphp.13273

Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinformatics. 2009 Jun 2;10:168. doi: 10.1186/1471-2105-10-168

Fossa P, Uggeri M, Orro A, Urbinati C, Rondina A, Milanesi M, Pedemonte N, Pesce E, Padoan R, Ford RC, Meng X, Rusnati M, D'Ursi P. Virtual Drug Repositioning as a Tool to Identify Natural Small Molecules That Synergize with Lumacaftor in F508del-CFTR Binding and Rescuing. Int J Mol Sci. 2022 Oct 14;23(20):12274. doi: 10.3390/ijms232012274

Johann Gasteiger, Mario Marsili. A new model for calculating atomic charges in molecules, Tetrahedron Letters, Volume 19, Issue 34, 1978, Pages 3181-3184, ISSN 0040-4039, https://doi.org/10.1016/S0040-4039(01)94977-9.

Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010 Jan 30;31(2):455-61. doi: 10.1002/jcc.21334

Pavan M, Menin S, Bassani D, Sturlese M, Moro S. Qualitative Estimation of Protein-Ligand Complex Stability through Thermal Titration Molecular Dynamics Simulations. J Chem Inf Model. 2022 Nov 28;62(22):5715-5728. doi: 10.1021/acs.jcim.2c00995