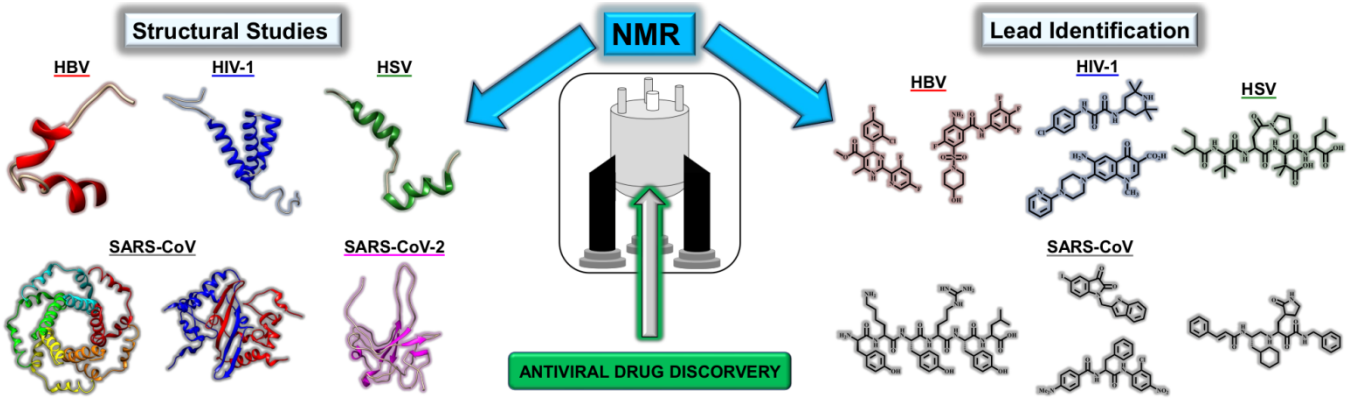
Send Orders for Reprints to reprints@benthamscience.ae

Current Medicinal Chemistry, 2019, *Volume, Page Enation*

The fight against human viruses: how NMR can help.

Marian Vincenzi[a] and Marilisa Leone[*a]

[a]*Institute of Biostructures and Bioimaging, National Research Council of Italy, Via Mezzocannone 16, 80134, Naples, Italy*



A variety of NMR-based tools can be employed to study viral proteins and their interactions to support the development of original antiviral agents along with potential diagnostic and therapeutic routes.

0929-8673/19 $58.00+.00 © 2019 Bentham Science Publishers

ARTICLE TYPE

The Fight against human viruses: How NMR can help.

Marian Vincenzi[a] and Marilisa Leone[*a]

[a]*Institute of Biostructures and Bioimaging, National Research Council of Italy, Via Mezzocannone 16, 80134, Naples, Italy*

ARTICLE HISTORY

Received:
Revised:
Accepted:

DOI:

Abstract:

Background: COVID-19 has brought the world to its knees and there is an urgent need for new strategies to identify molecules capable of fighting the pandemic. During the last few decades NMR (Nuclear Magnetic Resonance) spectroscopy has emerged as an intriguing structural biology instrument in the antiviral drug discovery field.

Objective: The review highlights how a variety of NMR-based tools can be employed to better understand viral machineries, develop anti-viral agents and set-up diagnostic and therapeutic routes.

Method: Works summarized herein were searched through PubMed database and the Web.

Results: The review focuses on a subset of human viruses that have been largely studied through NMR techniques. Indeed, NMR solid- or solution-state methodologies allow to gain structural information on viral proteins and viral genomes either in isolation or bound to diverse binding partners. NMR data can be employed to set up structure-based approaches to design efficient antiviral agents inhibiting crucial steps of viral life cycle. In addition, NMR-based metabolomics analyses of biofluids from virus-infected patients let identify metabolites biomarkers of the disease and follow changes in metabolic profiles associated with antiviral therapy thus paving the way for novel diagnostic and therapeutic approaches.

Conclusion: Considering the NMR-based work conducted on different viruses, we believe that in the close future much more NMR efforts will be devoted to discover novel anti SARS-COV-2 agents.

Keywords: NMR, viruses, viral infections, antiviral agents, drug discovery, structure-based design, metabolomics

1. INTRODUCTION

The route to the knowledge of human viruses started in 1901 when the first virus able to infect humans was identified [1, 2]. This was the case of YFV (or Yellow Fever Virus) coming from Africa and moving to America as a consequence of the slave trade [1]. The key events which raised the attention on viral pathogens were the "Spanish flue", that caused tens of millions deaths at the end of the second decade of twentieth century and the "Asian flu" in 1957 along with the "Hong Kong flu" in 1968, that brought other three millions deaths [3]. On the other hand, only the identification of HIV (Human Immunodeficiency Virus) in the early 1980s induced researchers to spend tremendous efforts to better understand viral pathogens [3].

Researches related to viruses led to the identification of different types of viral proteins linked to different functions [4]. Indeed viral proteins can be subdivided into four principal functional groups: structural, nonstructural, regulatory, and accessory proteins [4]. Structural proteins are those of the viral capsid; in fact, the genome of viruses is protected by a shell called capsid and formed by many protein subunits named protomers or capsomers [4]. Considering the genetic material, viral pathogens able to infect humans can be divided in DNA (DeoxyriboNucleic Acid), and RNA (RiboNucleic Acid) viruses [5, 6]. The first group includes "ds (double strand) DNA" and "ss (single strand) DNA" whereas, the second group includes "ds RNA", "ss RNA", and circular "ss RNA" [5, 6]. Single strand RNA can have (+) positive or (-) negative sense based on whether or not it can be straightly translated to generate viral proteins [6]. In addition, there is the retroviruses group, whose genetic material is made up of RNA(+) that is reversely-transcribed to DNA during the course of viral life cycle [5, 6].

The shell containing this genome (i.e., capsid) assumes a shape which depends on the type of packing in which the protomers of structural proteins are involved and can be helical, icosahedral or complex; in a few cases, the capsid is coated by a lipid membrane and composed by portions received from the host cell membrane during viral infection [4, 6]. The membrane-coated capsid (known as "viral envelope") can also contain glycoproteins [4].

A second group of proteins encoded by the viral genome is represented by the nonstructural proteins that are not comprised in the viral particles but work during virus

replication and assembly phases into the infected cells and can also play a role in gene transactivation [4].

Viral accessory and regulatory proteins play different roles and indirectly act in various processes like modulation of the transcription rate for the viral genes associated with structural proteins and the regulation of host cell functions [4]. Interestingly, viral proteins are often characterized by intrinsically disordered regions which make them able to properly answer to the high mutation rates of viral genomes and to adapt to the continuous alterations of the hostile environmental conditions [4].

For instance, hepatitis C and yellow fever viruses are examples of the "ss RNA(+)" group [6].

Instead, herpesviruses and adenoviruses belong both to the group of viruses with a "ds DNA" genome but, the first "viral family" possesses also an envelope whereas, the second one lacks it [6]. HIV and HBV (Hepatitis B Virus) are two examples of RT (Reverse Transcribing) viruses which are based on an RNA genome and a DNA genome, respectively [6].

Particular awareness and research interest in novel emerging viral pathogens came more recently from the epidemics associated with coronaviruses like MERS-CoV (Middle East Respiratory Syndrome - CoronaVirus), SARS-CoV (Severe Acute Respiratory Syndrome - CoronaVirus) and SARS-CoV-2.

SARS was responsible for more than 800 deaths and infected 8000 people globally in 2003 developing between 2002 and 2003 [1, 3, 7]. The first infected patient by MERS-CoV was reported in 2012 and MERS had a lethality rate higher (34%) with respect to SARS (10%) [1, 3, 7].

The name "MERS" derives from its origin set in the Middle East whereas, the "SARS" denomination is linked to the associated symptoms (i.e., fever with a temperature at least 38 °C, dry cough and shortness of breath) [7]. Although these coronaviruses transmit mainly from human-to-human, farm mammal or wild animal species were identified as possible reservoir hosts from which they started [3]. Indeed, dromedary camels were described as the major source for MERS-CoV infection in humans [8]. In the case of SARS-CoV epidemy, the mammalian reservoir host was the masked-palm or gem-faced civet cats that could be bought in Chinese wildlife markets [3].

COVID-19 (CoV Disease 2019) caused by SARS-CoV-2, firstly identified in China by the end of 2019 induced a pandemic which is still affecting the world [9-11]. Based to the WHO (World Health Organization) situation report of 19 July 2020, 14.043.176 cases of contagion and 596.583 deaths were globally identified [7].

MERS-CoV, SARS-CoV and SARS-CoV-2 viruses belong to the Coronaviridae family and the order of Nidovirales, whose members are characterized by the largest known genomes among "ss RNA" viruses encoding for 16 non-structural and 4 structural proteins [9, 11, 12]. The name of these viruses is related to the presence on the envelope of proteins that look like the spikes of a crown and thus, are defined spike proteins [9, 11, 12].

The newest COVID-19 pandemic with its awful impact on the every-day life worldwide tremendously raised the interest in the search of novel antiviral drugs and vaccines to block viral infections. NMR (Nuclear Magnetic Resonance) spectroscopy is a powerful technique to support this field. Indeed, NMR applications to structural studies of viral proteins and genome, NMR screening assays to develop potential antiviral agents, NMR-based metabolomics approaches on biofluids from infected patients will be described in the next subparagraphs. The review will be centered on a few of the better known viruses able to infect humans [13].

1.1. Human Viruses object of this study

This review will be centered on a subset of well-studied human viruses for which antiviral drugs have been approved: HIV, HCV (Hepatitis C Virus), HBV (Hepatitis B Virus), HSV (Herpes Simplex Virus) [13]. In addition studies on coronaviruses and related epidemics/pandemics along with routes to face them, will also be briefly summarized in the latest paragraphs of this review [14].

HIV belongs to the Orthoretrovirinae family and HCV to the Flaviviradae family, these viruses are based on a "ss RNA" genome and are responsible for AIDS (Acquired Immune Deficiency Syndrome) and severe liver diseases, respectively [13]. Instead, HBV, HSV belong to the group of viruses with "ds DNA" genome; HBV induces severe diseases including hepatitis and even liver cancer while, HSV generates oral and genital herpes infections accompanied by skin lesions and cold or blistering sores, ache during urination [13]. In addition, different coronaviruses (MERS-CoV, SARS-CoV and SARS-CoV-2) are associated with fever, cough, and/or shortness of breath. Interestingly, most of these viruses are defined as zoonotic because they crossed the specie barriers and jumped from animals to humans [9, 10]. The origin of these viruses makes them dangerous because human immune system is not ready to face a pathogen never met before [15].

1.2. Principal Strategies to discover antiviral agents

The first step of viral infections usually consists in binding of the virus to the host cells, followed by entry and replication [16]. However, several viruses able to infect humans (e.g., HIV-1 and HSV) can pass from a cell to another without diffusing in the extracellular environment and this kind of viral transmission through cells is responsible for their rapid dissemination and their ability to by-pass the immune system and enhance disease effects [16].

The mechanisms through which viruses entry the cells and disseminate have been widely investigated as potential route for the development of antiviral agents [17].

The need for novel antiviral agents led indeed to design a multiplicity of strategies able to block viral infection [18]. From a general point of view, drugs against viral infection can be divided on the basis of their level of action [18]. Indeed, there are compounds able to block the entrance of the virus into the host cells (e.g., peptides against glycoprotein b of herpes viruses), those acting on specific viral enzymes (e.g., didanosine against reverse transcriptase of RT viruses) and others inhibiting the production of viral proteins in the infected cells (e.g., some drugs used against HIV) [18]. For instance, acyclovir triphosphate is a molecule used as drug to inhibit HSV replication; it directly blocks viral DNA polymerase. In detail acyclovir specifically binds the thymidine kinase of HSV and is converted in the active triphosphate form; being a nucleoside analogue acyclovir triphosphate is incorporated in the viral DNA where it plays the function of a chain terminator as it lack the 3′-hydroxyl group, thus blocking the viral action [18]. Another example is given by zidovudine (AZT) which is an inhibitor of reverse transcriptase that HIV utilizes to convert its RNA into DNA and interferes with HIV replication [18]. Peptides have also recently been employed to hamper HIV-1 virus entry in the host cell (e.g., LRSRTKIIRIRH, also named G1, and MPRRRRIRRRQK, also named G2). G1 and G2 work by binding to Heparan Sulfates (HS), that are present on the surface of many cells, avoiding their interaction with viral glycoproteins and consequently depriving the virus from important attachment points for the entry into host cells [18].

From a general point of view, a protocol to produce efficient antiviral agents requires at least three sequential steps: "lead generation", "lead optimization" and "lead development" [19]. The "lead generation" can be achieved starting from the structure of the substrate of a particular target viral enzyme or from results of the screening of hundreds of thousands of compounds against a viral protein [19]. The majority of drugs on the market is directed against viral enzymes or transmembrane receptors [19]. In the search of novel lead against viral infection it is worth mentioning saquinavir that was obtained through a substrate relying approach; this compound represents the first marketed inhibitor of HIV protease and was designed based on the structural comparison between viral protease and the mammalian aspartyl protease renin, for which previous inhibitors projects had been conducted [19]. Instead, a screening approach led to the identification of a series of 4-substituted 2,4-dioxobutanoic acids with inhibitory activity towards the replication of influenza A and B viruses in cell culture working as blocking agents of the viral polymerase [19]. Molecule inhibitors identified in the first step usually have not the proper features and desired potency and require optimization that at this stage refers more to the *in vitro* activity [19]. Therefore, "lead optimization" is achievable through different approaches including biostructural methods (employing high resolution structural data that are instead more difficult to use at the lead generation phase), and high-throughput chemistry [19]. The resulting optimized molecules can be employed for the final clinical development phase only after a further optimization step (described as "lead development") focused on the pharmacokinetic properties of the molecules. For instance, a valyl ester derivative of acyclovir showed improved bioavailability and clinical profile [19].

Many antiviral agents were identified by exploiting structural details of viral proteases, which represent key enzymes involved in the modulation of different events, such as replication rate of infected organism and interference with normal immunological host defense [20, 21]. More, in detail proteases from different viruses (e.g., HIV, HBV, HCV and coronaviruses) are key elements in the viral replication and thus have gathered attention as potential target in the development of antiviral agents [20, 21]. Interestingly, viral proteases may be distinguished from the ones of host cells and the different structural details may lead to inhibitors selective for viral proteases. This is an important issue as blocking mammals host protease may cause problems to many regular physiological functions [21]. The host cell contains aspartic proteases (e.g., pepsin, chymosin and memapsin) with a pepsin-like fold consisting in a bilobal shape made up of two similar β-barrel domains which contribute individually with one aspartic acid residue to the catalytic site [20, 21]. The retroviral versions of these aspartic proteases (also known as retro-pepsins such as those associated with HIV and RSV) are characterized by an active homodimer formed by two identical β-barrel domains and by a considerable shortness of the sequence (approximately 100 residues instead of more than 300 in pepsin-like proteases) [20, 21]. Instead serine proteases present generally two main types of folds: trypsin-like (also named chymotrypsin-like) fold and subtilisin-like fold with the first one being the most common among eukaryotic, prokaryotic and viral serine proteases. Serine proteases of HCV adopt for example a trypsin-like fold which is characterized by two β-barrels and a catalytic triad formed at the interface between the two domains [20, 21]. Instead, herpes viruses and HCMV (Human Cytomegalovirus) serine proteases represent another small family of proteases which possess a fold made up of seven helices which surround a seven-stranded, mostly antiparallel, β-barrel [20, 21]. On the basis of the structural information related to viral proteases, different inhibitors were generated, including the non-peptide compound UCSF8 (used against HIV aspartic protease) [21] and the peptide CMK (Cbz–Val–Asn–Ser–Thr–Leu–Gln–CMK, where Cbz=benzyloxycarbonyl and CMK=ChloroMethyl Ketone, used against SARS-CoV main protease) [22]. Similarly structural biology tools were employed to generate different inhibitors of HCV NS3 (NonStructural Protein 3) protease, such as WO-9950230 and US6187905 [23].

Interestingly, the life cycles of various DNA and RNA viruses significantly depend on the channel-forming activity of viroporins, which form a family of small hydrophobic integral membrane proteins [24]. More in details, viroporins

are able to assemble forming ion channels and pores inside the host cell membrane making it more permeable thus facilitating the exit of virions [24]. Therefore, the development of molecules able to act on these proteins is considered a valuable route to get antiviral agents [24]. Amantadine and rimantadine are two examples of licensed antivirals that are used also to alleviate symptoms of Influenza A by interacting with the viroporin M2 (Matrix 2 protein) channel of Influenza A virus [24]. However, extensive use of viroporins as target for antiviral agents still requires a better understanding on the effective role of viroporin channel activity within virus life cycles [24].

An interesting route to get molecules able to block viral infection is based on targeting the viral capsid [25, 26]. The capsid is formed by the CA (capsid) protein that spontaneously assembles into an oligomeric structure and plays the fundamental function of safe carrier of the viral genome to the host cell [26]. In the capsid CA monomers or very small multimers pack each other precisely under the effect of spontaneous thermodynamics interactions [26].

Drug-discovery approaches targeting viral capsids are based on the idea that ad hoc designed molecules could be able to insert in the architecture of capsid and hamper its formation [26]. The interest on the capsid as drug target rose also from the need to overcome the issue of the continuous mutations of viruses which can make them resistant to the canonical drugs and can be achieved by molecules affecting the capsid oligomeric assembly as a whole when interfering with certain interfaces in the well-organized capsid lattice [26].

Capsid based drug discovery strategies allowed the identification of potential antiviral agents belonging to different class of molecules, such as a) synthetized organic compounds (e.g., BAY41-4109 and GLS4 (ethyl 4-[2-bromo-4-fluorophenyl]-6-[morpholinomethyl]-2-[2-thiazolyl]-1,4-dihydro-pyrimidine-5-carboxylate) against the HBV capsid; GS-CA1 against HIV capsid; pillarene and D-Pro-C4A against HPV), b) natural products and their derivates (e.g., isothiafludine, against HBV representing a derivate of the leucamide A cyclic peptide extracted from the Australian marine sponge *Leucetta microraphis*) and c) peptides (e.g., pep14 (DLDQFPLGRKFLLQ) used against HPV) [26].

Interestingly, a recent review reported on different antiviral agents extracted from marine sources [27]. For instance, reverse transcriptase and protease of HIV-1 are inhibited by 8,8′-bieckol, that is a tannin derivate coming from the brown algae *Ecklonia cava* [27]. Anti HIV-1 activity was associated also with sulfated derivates of chitin and chitosan; chitin can be found in crustaceans, fungi, invertebrates, and insects whereas, chitosan derives from chitin deacetylation [27]. In addition, CaSP (Calcium SPirulan) is a sulfated polysaccharide which can be extracted from the marine blue-green alga *Arthrospira platensis* and characterized by a potent antiviral activity against HIV-1 [27].

Replication, transcription, and translation of viral genome strongly depend on the activity of helicases, enzymes that separate double stranded nucleic acids into single filaments strands exploiting the energy coming from ATP hydrolysis [28]. During viral genome replication, helicases function can be employed at different stages. The separation of the double strands is required in virus with “ds RNA” and “ds DNA” genome to allow copying single filaments whereas, in viruses with single stranded genomes, helicases need to divide double strands formed after replication. Helicase activity is involved in the transcription of viral mRNAs, translation, disruption of RNA-protein complexes, and assembly of nucleic acids into virions [28]. Therefore, inhibition of helicases was described as potential route to identify novel antiviral agents [28]. The basic idea behind inhibitors described for helicases consists in blocking or somehow slowing down the unwinding of double stranded nucleic acids by different mechanisms, such as avoiding ATP hydrolysis, inhibiting nucleic acid binding and nucleic acid release, hampering the binding of a needed cofactor to the helicase [28].

Targeting viral RNA by small molecules represents another prolific route for the development of antiviral agents able to block viral replication at different levels of the process. Target RNAs for such drug discovery approaches have been identified in several viruses such as HIV, HCV, SARS-CoV and influenza A virus. For example, TAR (TransActivation Response) element of ss RNA(+) genome of HIV includes the binding site for Tat protein, the TAR/Tat complex induces transcription elongation to produce full length transcript using also host cell factors [29]. Instead, the IRES (Internal Ribosome Entry Site) region of HCV RNA is important for translation of viral genome by engaging ribosomes at the beginning of viral start codon without requiring most host cells initiation factors [29]. Another level of inhibition was reached starting from the frameshifting PK (PseudoKnot) RNA from SARS-CoV [29].

This region modulates the programmed ribosomal frameshifting, a mechanism in which overlapping reading frames can be translated, thus maximizing the coding content of genome [29]. For example, translation of viral proteins necessary for replication in SARS-CoV such as the RNA-dependent RNA polymerase is started by a specific frameshift which is activated by a three stemmed RNA PK [29].

Molecules able to bind these specific regions of RNA from different viruses have been identified and behave as promising antiviral agents thus highlighting the potentiality of viral RNA regions as drug target alternative to the canonical viral proteins [29].

A lot of interest has always been devoted also to the discovery of vaccines to fight viral infections. It is possible to generate a vaccine by using the virus itself in a weakened or inactivated form to make the adaptive immune system able to recognize the viral pathogen [30]. To weaken a virus it can be continuously passed through host cells (animals or

humans) to induce a series of mutations sufficient to reduce the ability of virus to cause the pathological conditions [30].

For example, it could in theory be possible to alter the genetic code of a virus so that viral proteins are generated less efficiently and use this altered form to make a vaccine.

Nevertheless, to generate a vaccine an inactivated virus can be employed and the inactivation can be achieved through use of chemicals (e.g., formaldehyde, β-propiolactone and formalin), or heat to make viruses unable to start an infection [30, 31]. Other types of vaccines not relying on the virus itself such as nucleic acid vaccines, viral vector vaccines, protein-based vaccines, can be also produced [30].

The major difficulty in the process to get novel vaccines rises from the highly heterogeneity which characterizes the populations of virus strains circulating in animal reservoirs [31]. Indeed, it was suggested that an unique vaccine may be designed for different coronaviruses only if a proper level of information on the antigenicity and neutralizing antibody footprint of each coronavirus is reached [31].

The route to approve a novel antiviral agent or to get a vaccine may be too long to be effective during a viral pandemic [32]. A valid alternative may be to test molecules approved for other diseases and validate their use in an antiviral therapy (i.e. "drug repurposing") [32]. This approach presents certain advantages, including the reduced cost connected to drug development and the faster entrance of the drug in the clinical trials [32]. In addition, the so-called "repurposed drug" can be combined with other monotherapies and thus increase the chance to overcome viral resistance issues [32]. On the other hand, the elaboration of therapies based on the combined use of different drugs can be arduous especially when the repurposed drugs shows polypharmacology i.e., ability to act on multiple targets, thus eventually producing elevated toxicity [32].

1.3. Potential applications of NMR Spectroscopy in the antiviral drug discovery field

NMR spectroscopy represents one the most powerful techniques in structural biology to investigate the structure and interactions of biological macromolecules [33]. NMR is also widely employed in the drug discovery field not only as a screening technique but, as a versatile tool to be exploited at different stages of the drug design and development process [34, 35].

Different NMR strategies can be implemented in fragment based lead discovery for both hit identification and hit validation. Two main strategies can be employed to discover ligands of a protein target by screening libraries of small molecules (or even peptides): those based on observation of a protein target (that could be a chosen viral protein) and those relying on detection of the ligand [35]. Methods based on protein include the well-known CSP (Chemical Shift Perturbation) studies [35]. The most common CSP strategy requires acquisition of 2D ([^{1}H, ^{15}N] or [^{1}H, ^{13}C]) heteronuclear correlation spectra of an uniformly ^{15}N and/or ^{13}C labeled protein in absence and in presence of a compound (or a mixture of compounds). The possible ligands are identified by comparing the two spectra (with and without compound/compounds) and looking for chemical shifts variations or intensity changes [34, 35]. This strategy is very powerful as, if the resonances of the target protein are known, it allows to identify the binding pockets on the protein surface [35, 36]. Methods based on observation of protein resonances are very informative but more expensive as require a relative large amount of labeled protein with respect to techniques based on ligand observation [35]. In the case of ligands positioned in a protein binding pocket in proximity of the methyl groups of valine, leucine and isoleucines residues, the sensitivity can be considerably increased by labeling only the protein methyl groups with ^{13}C isotope and performing CSP analysis by acquiring 2D [^{1}H, ^{13}C] heteronuclear correlation spectra [35, 36]. The CSP approach can be exploited to study weak and strong interactors. By running NMR spectra at increasing concentrations of ligand and monitoring consequent gradual changes in chemical shifts or intensity it is also possible to get estimates of dissociation constants (K_D) of the protein-ligand complex [36, 37].

There are many ligand-based screening techniques [38, 39]. For instance, the STD (Saturation Transfer Difference) NMR experiment is commonly used to identify molecules binding to target proteins as well as to achieve precise group epitope mapping and it also allows, when performing titration experiments, to get estimates of dissociation constants [40-42]. The STD relies on selective saturation of protein and transfer of magnetization due to spin diffusion from the protein to the bound ligand; briefly two spectra are acquired one with selective irradiation of protein in a region where signals of compounds are absent (often at negative ppm) known as "on resonance spectrum" and another one without protein irradiation ("off resonance spectrum"). The bound ligand will undergo a decrease of intensity upon passage of saturation from the protein and can be detected in the STD spectrum (obtained by subtracting from the on resonance spectrum the off resonance one) [40, 42].

A valuable application of STD in the antiviral drug discovery field is reported in a work conducted on HRV2 (Human Rhinovirus Serotype 2) describing the identification of an antiviral agent along with the characterization of the ligand-virus interaction [43].

WaterLOGSY (Water-Ligand Observed via Gradient SpectroscopY) is another NMR experiment relying on detection of ligand resonances that can be implemented to discover novel hits interacting with a target protein [35, 39, 44]. This experiment is similar to the STD but based on the transfer of magnetization from the bulk water in the binding site to the protein and next to the bound ligand, through Nuclear Overhauser (NOE) effect and chemical exchange [45].

Ligands can be also identified by a simple screening through 2D [^{1}H, ^{1}H] NOESY experiments that are recorded for a small molecule alone and in presence of a small amount of protein. Small molecules have short correlation times and are characterized by small positive NOEs (diagonal and cross-peaks are in anti-phase) while, when a small molecule binds to a protein target, it assumes a larger correlation time and the sign of NOE becomes negative (diagonal and cross-peaks are in phase) [35, 38].

The above mentioned NMR screening methodologies based on observation of ligand resonances work for weak interactors with dissociation constants in the micromolar – millimolar range [35].

The main concept behind FBS (Fragment-based Screening)-NMR is the identification of a "fragment" interacting with a specific target, among the members of libraries of compounds that are small in size and thus can cover a large chemical space [46]. Therefore, FBS is characterized by higher hit rates and binding efficiencies with respect to HTS (High Throughput Screening) [46]. On the other hand, the binding of a small fragment to a protein target is generally weak and a real lead can be generated only after chemically functionalize and optimize the starting compound generating a more complex chemical entity [46]. Interestingly, the SAR (Structure-Activity Relationship) by ILOEs (Inter ligand NOEs) represents an NMR strategy that through acquisition of simple 2D [^{1}H, ^{1}H] NOESY experiments allows to recognize two small molecules that bind a protein target simultaneously into two adjacent close binding pockets. NOE build up experiments allow as well to identify the mutual orientation of the two fragments next, molecular modeling can be implemented to design a proper linker to join the two compounds together. The compound obtained by linking two fragments together, could result a much stronger binder respect to the isolated "pieces" with a dissociation constant value even equal to the products of the K_D values of the single compounds [47]. Of course SAR by NMR based on acquisition of CSP experiments can be as well implemented to find fragments binding to close sites of a target protein when the resonance assignments of the latter are known [35].

Based on the concept that pathological conditions can be associated with alterations of metabolic pathways, metabolomics studies may provide an alternative original way to find novel antiviral agents [48].

NMR represents together with mass spectrometry and chromatographic methods one of the principal analytical techniques in metabolomics studies [49]. Metabolites analysis of biological fluids and tissues of patient affected by certain diseases and/or viral infections can be performed by NMR spectroscopy upon careful selection of sample preparation protocols [50].

It is also important to establish the enzymes involved in alterations of metabolic pathways as they can be considered novel targets to build original therapeutic routes [48]. For example, the infections induced by IVA (influenza A virus) and HSV-1 (Herpes Simplex Virus Type-1) were associated with an upregulation of acetylneuraminic acid and alterations of 2′-deoxynucleotides (dTTP (deoxyThymidine TriPhosphate) and dTMP (deoxyThymidine MonoPhosphate)) in human fibroblasts, respectively [48]. This is coherent with the activity of antiviral agents found for IVA (i.e., oseltamivir) and HSV-1 (i.e., acyclovir) [48]. Indeed, oseltamivir acts as inhibitor of neuraminidase (enzyme involved in the modulation of acetylneuraminic acid levels) and acyclovir inhibits thymidine kinase (enzyme regulator of 2′-deoxynucleotides levels) [48].

A molecule able to neutralize the alterations which a viral pathogen can induce in the levels of specific metabolites could work as antiviral agent [48]. Nevertheless, relating information from several metabolomics studies obtained with diverse techniques (like NMR and mass spectrometry) or correlating metabolomics data with those from genomics and transcriptomics may bring to an improved knowledge of metabolic pathways [51, 52].

In metabolomics studies measurements of metabolites concentrations are important but they do not give information about variations in influx and efflux from which the perturbation of metabolite levels depend [53]. Metabolites are generated and consumed by a series of related enzymatic reactions that transform received nutrients into energy and biomass [54]. When the steady state is reached, a balance between the quantitative inflows and effluxes deriving by each metabolite must be achieved [54].

Thus it is important when conducting metabolomics studies also analyzing pathway activity, that can be quantified considering metabolic flux i.e., material flow per unit time [54]. Interestingly, the influx and efflux pathways are responsible for the perturbation of metabolites levels [53].

Metabolomics studies along with the use of isotope tracer are important in the identification of novel drug targets [53].

The use of isotopic tracers (like ^{13}C-labeled glucose) to feed cultured human fibroblasts revealed through metabolomics analyses that fibroblasts infected by HCMV virus are characterized by significant amount of citrate molecules provided with two labeled ^{13}C-carbon atoms whereas, those infected by HSV-1 by significant amount of citrate labeled with three ^{13}C carbons, indicating that the two viruses induce two different TCA (tricarboxylic acid) cycle influx pathways and suggesting that different groups of enzymes could be targeted for the development of selective anti-HCMV and anti-HSV-1 agents [53]. The above mentioned study was largely conducted by mass-spectrometry technique however, employment of NMR technique in stable isotope-enriched tracer studies to investigate biochemical pathways and networks have been reported as well [55].

NMR spectroscopy can be employed also to get 3D structures of proteins involved in viral infection as isolated entities or in complex with other proteins or small molecules including inhibitors. NMR spectroscopy works rather well

with small proteins (MW below 20 kDa) and is more advantageous with respect to X-ray studies when dealing with dynamic systems and particularly flexible regions such as IDR (Intrinsically disordered region) or when trying to get information on small peptide ligands. These structure information are important for structure-based design of antiviral agents [33]. An example is provided by the E6 protein of HPV and its interaction with E6AP (E6-Associated Protein) [56]. E6 protein acts mainly by degrading the tumor-suppressor protein p53 to avoid growth arrest and/or apoptosis. Binding of E6 to a cellular E3 ubiquitin ligase (named E6AP) is important to form a complex able to engage p53 and induce its ubiquitination and subsequent degradation [56]. Indeed, NMR studies of the interaction between E6 and short peptides reproducing its consensus binding motif on E6AP and other interaction partners allowed to build a pharmacophore model that was next implemented in a virtual screening approach aiming at identifying ligands of the E6 protein and inhibitors of the E6-E6AP complex [56].

Finally NMR spectroscopy is of course a powerful analytical technique that can be employed to verify identity and purity of synthesized compounds and support medicinal chemistry efforts to obtained the desired products.

2. HEPATITIS C VIRUS (HCV)

2.1. Mechanism of action

HCV is an enveloped virus discovered in 1989, characterized by "ss RNA(+)" genome and belonging to the *Flaviviridae* family (the same family of yellow fever virus and dengue fever virus) [57-63]. This virus is the main responsible for chronic liver diseases in the world [58]. Interestingly, there are different genotypes for HCV and the most frequent infections are linked to the genotype 1 (GT1) [59, 60]. The diffusion of this virus depends on the direct contact with infected blood and thus can occur by blood transfusions, health-care-related parenteral administrations, or drug injection [59].

The starting point of the HCV infection is the interaction between viral single particles (also named virions) and the cell entry factors anchored on the host cell membrane [61].

Molecules exposed on the host surface, that are needed for virus entry, can be represented by either receptors or co-receptors. Many viruses can use just a single molecule as a receptor but, there are also viruses for which a co-receptor positioned close to the receptors is necessary for their entry in the host cells. Carbohydrates and lipids have been shown to be important for virus-host cell interaction thus receptors are not exclusively membrane proteins [62].

For HCV, several molecules have been reported to be candidates for receptor or co-receptor for viral entry [62]. Indeed the identified entry factors are more than 30 and the best known are CD81, SR-BI (Scavenger Receptor class B type I), CLDN1 (CLauDiN-1), OCcLudiN (OCLN) and EGFR (Epidermal Growth Factor Receptor) [62].

In details, HCV virions are captured by glycosaminoglycans prior being relocated to a cell surface receptor and/or co-receptor and next internalized into cells by endocytosis. After the capsid shells are disassembled (i.e., uncoating process), the viral RNA is translated first into a precursor polyprotein that is next cleaved by cellular and viral proteases to generate each viral protein [62]. Once the virus complete its entry in the cytoplasm, the NS (Non Structural) proteins catch the cellular membranes from the ER (endoplasmic reticulum) to generate the so-called membranous web where spatially coordinated replication complexes are formed by viral proteins and genomes [60, 62].

2.2. NMR investigations

Many NMR studies related to HCV have been conducted, examples are provided by structural investigations to identify features important for the interaction with membrane of the non-structural protein NS4B, that is crucial for formation of the replication complex [63]; solution structures of IRES RNA domains [64-67]; interaction studies between viral proteins and RNA [68]; design and evaluation of different HCV proteins inhibitors [69-71].

2.2.1. Structural and Interaction studies

The HCV genome encodes just for a single viroporin p7 that is made up of 63 residues and forms oligomers in the host membrane thus creating ion channels that exhibit cation selectivity [72]. NMR spectroscopy allowed understanding the structural details behind channel formation by multiple monomers of the p7 membrane protein from HCV [72]. The attention on this protein is due to its role in the assembly and release of infectious virions and thus to its potential as target for antiviral drug development [72]. Several inhibitors of the p7 channel fuction have been reported in literature such as amantadine and rimantadine that also target the influenza M2 channel, hexamethylene amiloride (HMA), and iminosugar derivatives with long-alkyl-chain [72, 73].

The NMR structure of p7 shows an uncanonical hexameric assembly, with single p7 monomers (a), not only contacting other immediately close unities, but also going further and assembling with the a+2 and a+3 monomers, generating a complex funnel-like helical structural topology (a "flower-like" architecture) [73]. This structure was obtained by cutting edge NMR techniques in presence of DPC (DodecylPhosphoColine) and by using a precise isotope labeling scheme. As first step, NOE-based experiments were employed to determine the secondary structure elements of the single monomers [73]. Then, a mixed sample was prepared in which 50% monomers were [^{2}H, ^{15}N]-labeled and the other half ^{13}C-labeled [73].

This labeling scheme allowed to reveal NOE contacts between the ^{15}N -attached protons of one monomer and the ^{13}C-attached protons of the neighboring monomer and together with orientational restraints from RDC (Residual Dipolar Couplings) led to a better knowledge on the assembling of the central cavity of the hexamer [73].

NMR studies showed that adamantane interacts with six equivalent hydrophobic sites (as the p7 hexamer presents a 6-fold symmetry) close to the bend of the middle pore-forming helices, made up of portions from diverse helical regions and subunits [72, 73].

NMR titration along with NOE experiments were implemented to study the interaction between the p7 channel and its inhibitors. P7 was dissolved in a lipid environment (DPC DodecylPhosphoCholine) and high quality [^{1}H, ^{13}C] HSQC (Heteronuclear Single Quantum Coherence) spectra of methyl groups and [^{1}H, ^{15}N] TROSY (Transverse Relaxation Optimized)-HSQC spectra were recorded at increasing HMA and rimantadine concentrations [72]. These studies revealed the importance of hydrophobic interactions in binding of p7 to its inhibitors [72]. Indeed, reduction of hydrophobicity by substitution of crucial leucine residues in p7 binding site with alanines was associated with a defeat in rimantadine sensitivity [72].

NMR studies were also performed in trifluoroethanol/water mixtures and ^{1}H and ^{13}C NMR spectra were recorded to determine the secondary structure elements characterizing the monomeric form of p7 protein [74]. The structure of p7 was as well determined by NMR in presence of micelles formed by DHPC (1,2-dihexanoyl-*sn*-glycero-3-phosphocholine), a short chain phospholipid which simulates a membrane-like environment [75, 76]. NMR experiments to study structure, dynamics and overall structural topology of p7 comprised hydrogen/deuterium exchange, paramagnetic relaxation enhancement and bicelle "q-titration" (this kind of titration monitors change in intensity of NMR signals by varying the ratio of short (DHPC) and long (DMPC, 1,2-dimyristoyl-sn-glycero-3-phosphocholine) chain lipids and gives information on local motions of residues in the membrane) [76].

Several structural studies by NMR focused on HCV RNA [68, 77]. Translation of the HCV RNA starts from a largely structured internal ribosomal entry site (IRES) in the 5′ untranslated portion (5′ UTR) of the RNA genome [77]. The HCV IRES is made up of four domains I, II, III, and IV; the 40S ribosomal subunit recognizes domains II - IV whereas, eukaryotic initiation factor 3 (eIF3) interacts with domain III to start translation [68].

An interesting work reported the NMR structure of the complex between a domain IIa construct of IRES and Isis-11, a potent small-molecule inhibitor of HCV replication [77]. These structural studies allowed to better comprehend the kind of inhibitors that could be used to alter the RNA structure [77]. It was in fact demonstrated that a compound inhibitor, even if interacting with RNA with micromolar affinity, is able to produce a large structural rearrangement in the IRES domain IIa RNA [77]. NMR studies let speculate that alterations in IRES structure may be fundamental to achieve efficient inhibition of HCV replication [77].

Another NMR study reported on the interaction of IRES domain IV with the peptides LKKLLKLLKKLLKLKG (dissociation constant K_D = ~ 67 nM) and a truncated form (LKKLLKLLKKLLK) [68, 78]. CSP studies were carried out through acquisition of [^{1}H, ^{13}C] HSQC spectra of IRES domain IV in presence and absence of peptides and showed a significant decrease in binding affinity by truncation of the last three C-terminal residues of the LKKLLKLLKKLLKLKG sequence, thus suggesting the critical role of these residues for the RNA-peptide interaction [68]. Nevertheless, CSP experiments demonstrated that the truncated peptide (LKKLLKLLKKLLK) by binding to the upper stem region of IRES domain IV could consequently block the initiation of translation by avoiding recognition of the AUG start codon [68].

Another viral target widely studied by NMR is represented by the group of NS (NonStructural) proteins (i.e., the NS2, NS3, NS4A, NS4B, NS5A, NS5B) that play a role in RNA replication and virion formation [79, 80].
Proteins from NS3 to NS5B interact with DMV (Double-Membrane Vesicles) made up by the host endoplasmic reticulum (ER), and form the viral replication apparatus. This assembly, through the NS5B RNA-dependent RNA polymerase activity allows viral genome replication [79, 80]. NS3 possesses an N-terminal domain including a serine protease which represents a member of chymotrypsin family and plays a fundamental role in the maturation of viral polyproteins [79, 80]. Therefore, this domain was identified as potential drug target for the development of antiviral agents [80]. Different NMR experiments were carried out to get resonance assignments, distance restraints and angular restraints required to calculate the structure of the complex formed by NS3 protein serine protease domain and the covalently bound reversible inhibitor α-ketoacid Boc-Glu-Leu-(γdi-fluoro)Abu (where Abu= 2-Aminobutyric acid) (Fig. **1**) [80]. The study showed that the residue in position P2 (when referring to proteolytic cleavage of peptide substrates the P1 and P1' are the sites in between which cleavage occurs while the P2 represents the N-terminal position with respect to P1 site) of the inhibitor (i.e., the Leu) play a crucial role in stabilizing the active site His-Asp hydrogen bond by avoiding solvent exposure. This effect is important for enzyme activation [80].

Nevertheless, a multidisciplinary approach including NMR experiments along with X-ray crystallography and kinetic analysis helped identifying structural features responsible for the interaction of the protease NS3 with the inhibitor BI201335, that entered Phase II b trials [81]. BI201335 is made up of a C-terminal carboxylic acid that interacts non covalently with the active site and a bromo-quinoline moiety on a proline residue that confers great potency [81]. NMR was particularly useful to study the

protonation state and hydrogen bonding capacity of the catalytic histidine His57 in its bound form [81].

NMR represents a useful technique to get structural information on disordered domains. This is the case of domain 2 of NS5A (also named NS5A-D2), a protein module characterized by a proline-rich sequence which comprises a short structural motif known as PW-turn [79]. NS5A-D2 binds and also represents a substrate of a crucial host factor for HCV replication, the human peptidyl-prolyl cis/trans isomerase CypA (Cyclophilin A). The principal CypA interaction site can be located inside a small NS5A-D2 region, made up of roughly 20 residues, in which a few highly conserved amino acids among the different HCV genotypes, are contained including 3 prolines. The PW-turn structural motif (i.e., 310-PAWA-313 in JFH1 HCV strain) is also positioned inside the same region and is important for HCV replication. The NS5A-D2/CypA complex can be considered as a valuable drug target to design inhibitors of HCV replication. NMR highlighted the presence of two states for the PW-turn motif : a folded and a disorder state that are in equilibrium with each other and revealed an allosteric modulation linked to prolines cis-trans isomerization [79]. In major details NMR data pointed out that the PW-turn motif is linked from a structural point of view to a wider portion of NS5-D2 including residues 304-323, thus, this larger peptide can adopt a structured motif in equilibrium with a disordered conformation. Alterations within this portion influence the equilibrium between the folded and unfolded forms. Different mutations (A311G, D316E, D316E-Y317N, P306A, P310A, P315A, P319A and P320A) along with the cis-state of P306, P310, P315, P319 and P320, represent allosteric alterations [79]. In addition, viral RNA replication efficiency seemed to be modulated by this “structured/unstructured” fraction. These NMR studies gave insights on potential novel strategies to regulate the efficiency of viral RNA replication by modulating the equilibrium between folded and disorder states of the PW-turn motif [79].

NMR was employed to investigate as well structural determinants connected to the interactions of viral proteins with host receptors/co-receptors that are necessary for virus entry in the host cell. In this context, NMR revealed powerful for studying under physiological solution conditions, the Large Extracellular Loop (LEL) of CD81. Tetraspanin CD81 represents an entry factor which is fundamental for the HCV infection of hepatocytes [82]. The protein (CD81-LEL) was expressed in mammalian cells and resonance assignments were determined [82]. NMR interaction studies characterized the binding between a dynamic loop located on the helical bundle of CD81 and the E2 glycoprotein of HCV and identified the position of the binding site on the surface of the receptor along with a crucial functional loop conformation [82]. The same studies also revealed the presence of a second region inside the exposed CD81 extracellular loop, which is close to the E2 glycoprotein binding site, and responsible for binding to the membrane [82]. The membrane orientation of CD81 may be important to well dispose the exposed E2 binding sites of CD81 for the interaction with the HCV E2 protein [82]. Having in hand resonance assignments for CD81-LEL, CSP NMR studies with [^{1}H, ^{15}N] HSQC experiments help also identifying the binding sites for potential inhibitors of the CD81/E2 interaction paving the way for the design of novel therapeutic compounds against HIV infection targeting CD81-LEL [82].

The two envelope proteins E1 and E2 of HCV are essential for receptor binding and fusion with the host membrane. These proteins are glycosylated in their N-terminal regions while the C-terminal sides contain the transmembrane domain responsible for anchorage within the viral membrane; the transmembrane regions in E1 and E2 interact forming an heterodimer through non covalent interactions. Noteworthy, NMR was effective also to determine the structural features associated with the 314-342 transmembrane region of the HCV E1 glycoprotein [83]. NMR structural studies were conducted in a mixture composed of water and a fluorinated alcohol (i.e., HFIP (hexafluoroisopropanol)) that was employed to somehow mimic an environment characterized by low polarity like that of the inner membrane. Also in this case NMR provided structural details that could be exploited to generate antiviral agents aiming possibly to interfere with the complex mechanism of membrane fusion and virus entry [83].

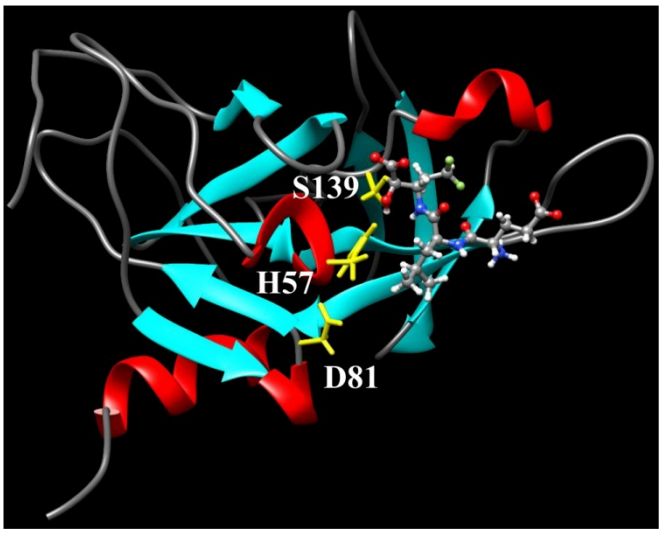


Fig. (**1**). NMR solution structure of the HCV NS3 protease in complex with a peptide inhibitor α-ketoacid Boc-Glu-Leu-(γdi-fluoro)Abu (structures n.1 and 2 of the NMR ensemble) (pdb entry code 1DXW). The residues of the catalytic triad His57, Asp81 and Ser139 are shown in yellow whereas the peptide inhibitor is report in a ball and stick representation [80].

2.2.2. Metabolomics studies

Improved diagnosis and prognosis of human liver diseases like those generated by HCV infection can be surely

gained by NMR metabolomics analyses [84]. The method presents some advantageous features and is first of all simple, fast, non-invasive, inexpensive with rather good sensitivity and specificity [52].

The key element in a metabolomics study by NMR is represented by the identification and validation of a biomarker that is a major point when dealing with liver diseases as it can support the diagnosis and help generating personalized therapeutic approaches [52]. A right biomarker needs to be discriminant, accurate, robust and universal. Metabolomics analyses in biological fluids through high-throughput approaches permit to gain information on many different metabolites at the same time and rarely studies, focused on a single metabolite, result really useful. It is also important to establish a link between biomarker levels and clinical outcomes. In addition it is recommended not to perform a single measurement at a specific time but, to study variations in metabolite levels during the course of a follow-up and at various phases of a treatment. Another fundamental element in the metabolomics is the choice of the biological fluid [52]. For instance, serum and plasma were described as more suitable for hepatological diseases [52]. In addition, urine is a biofluid which is abundant, readily available, easily stored and collectable by simple non-invasive techniques, thus suggesting the utility of its use in metabolomics [51, 52]. Another kind of sample that can be employed for metabolomics studies consists of intact tissues (like liver biopsies) that can be studied by high-resolution magic-angle spinning (HR-MAS) NMR spectroscopy [52].

Once the data are acquired, the next step consists in their statistical interpretation that is fundamental to get reliable results and is generally achieved by multivariate analysis methods (supervised and unsupervised analysis) [52]. Robust statistical models for precise interpretation of changes in metabolomics fingerprints are crucial to deliver important insights for patient management [52].

A few NMR metabolomics studies related to HCV infection have been reported in literature [84-88]. An interesting work was conducted starting from serum samples of a group of HCV-positive patients and based on ^{1}H NMR analysis of data obtained before and after viral eradication upon treatment with a DAA (Direct-Acting Antiviral) [84].

NMR spectra of 294 serum samples were recorded; samples were collected from 67 HCV positive patients at three different time lapses: during the infection (i.e., baseline point), 12 weeks and 24 weeks following a relevant virological response [84]. The same NMR metabolomics-based study also analyzed comparatively the serum profile of naïve HCV patients, naïve HBV patients (50 infected patients were analyzed) and healthy controls (43 samples). Results pointed out that effectively treatment with DAA caused differences in the metabolomics profile of HCV patients. Thus following therapy and virus eradication a metabolic switch took place; variations in metabolite levels of patient serum could be detected during DAA therapy and indicated changes in many pathways including but not limited to nitrogen metabolism, biosynthesis of several amino acids and aminoacyl-tRNA [84].

HCV is linked with dysmetabolic syndrome and a relevant high cardiovascular risk with respect to HBV. Interestingly HCV baseline samples were characterized by high levels of 2-oxoglutarate and 3-hydroxybutyrate respect to HBV samples. This result is interesting considering that 2-oxoglutarate and 3-hydroxybutyrate are known as potential biomarkers of cardiovascular risk.

Levels of several metabolites were found significantly different in HCV, HBV infected patients and healthy control groups. This study through a remarkable portrait of metabolic fingerprints stressed out how HCV and HBV act differently on cellular metabolism. Results gained through comparison with the healthy group samples showed how this metabolomics approach could be applied to establish fast, sensitive and rather inexpensive diagnostic tests and be implemented also to establish a prognosis [84].

Metabolomics studies based on ^{1}H NMR revealed to be rather valid also to identify with high sensitivity and specificity HCV-positive patients simply starting from urine samples [86].

Similar metabolomics-based work identified choline, acetoacetate and low-density lipoproteins as the most informative biomarkers to establish the presence of cirrhosis in HCV patients [88].

2.2.3. NMR drug discovery

In the paragraph 1.3. different NMR screening methods, based on either protein target or ligand observation, were described. Therefore, this subparagraph will be dedicated to the successful applications of these strategies in the development of antiviral molecules against HCV.

An interesting study reported how NMR can support other techniques (including X-ray crystallography and SPR) in the search of lead compounds targeting HCV helicase [89]. STD was employed to conduct an NMR screening of a fragment library against NS3 helicase of HCV [89]. First, 1D ^{1}H NMR spectra were recorded to evaluate the solubility grade and to identify diagnostic resonances for each fragment [89]. Then, the fragments able to bind the helicase were established by analyzing 1D STD NMR spectra; the screening led to indole based compounds as helicase ligands [89]. In addition, the interaction of the identified fragments was further verified by CSP studies through methyl shift NMR assays looking at variations in the methyl region of helicase 1D [^{1}H] spectra, that were recorded in presence and absence of a compound, or through [^{1}H, ^{15}N] TROSY spectra acquired for ^{15}N labeled helicase alone or in presence of a ligand [89]. CSP experiments were also set up through acquisition of ^{19}F NMR spectra for compound provided with a fluorine atom [89]. 1D ^{1}H methyl shift NMR assays recorded at increasing compound concentrations allowed to get also K_D estimates. Instead, NMR competition-type

interaction assays were employed to get insights in the binding pocket on the helicase surface [89].

A nice multidisciplinary approach relying on NMR, X-ray crystallography, molecular modelling and medicinal chemistry was employed to design inhibitors of the NS3 HCV protease starting from the peptide -DDIVPC- [90]. This strategy consisted in a ligand-based approach and paved particular attention to the bioactive conformation at all stages of the drug discovery campaign. Several ^{1}H and ^{13}C NMR spectra resulted particular useful to gain crucial information for compounds optimization, for example ^{13}C T1 and transferred ^{13}C T1 relaxation experiments [70] revealed important differences between the free and bound states of ligands [90]. It's worth noting that the relative variations in ^{13}C T1 relaxation gives a qualitative information on the site-specific changes in the immobilization of a ligand following interaction with the protease [70]. In the end, this study contributed marketed and advanced drug candidates - like Faldaprevir and Ciluprevir (BILN 2061) - with improved potency against the NS3 protease of HCV with respect to the starting peptide (Fig. **2**) [90]. Enhancement in protein-ligand binding comes generally and simultaneously by both stabilization of the bioactive conformation, and establishment of better interactions. In particular, this work showed through a detailed SAR and a group of supportive NMR experiments that improved affinity and potency can be accompanied by an increase in the bioactive conformation in the free unbound state, through for example shortening and rigidification of a peptide scaffold [90].

Similarly, NMR techniques were also applied to optimize compounds against HCV NS5B polymerase [91]. A combination of ROESY (Rotating Frame Overhauser Enhancement Spectroscopy) spectra (employed to analyze the compounds conformational preferences in the free state) and transferred NOESY spectra (to study the structural features of bound compounds states) resulted essential to suggest the best conformational restrictions and scaffolds to be inserted in the starting compounds to stabilize the conformations and improve affinity and potency [91]. This study led to the phase III clinical candidate deleobuvir (also named BI207127) [91].

An NMR-based screening of a fragment library brought to the discovery of 16 small-molecule ligands interacting with the substrate binding site of HCV NS4A-bound NS3 protease. These starting ligands represented weak hits provided with dissociation constant values ranging from ≈100 µM to 10 mM [71]. The screening was achieved using ^{15}N labeled protein and [^{1}H, ^{15}N] HSQC spectra. Analogues of these hits were also tested and optimized compound ligand were identified by NMR and biochemical assays [71]. Next, NMR CSP studies gave detailed insights in the locations and orientations of the scaffolds into the active site and pointed out a couple of molecules interacting in the adjacent S1-S3 and S2' substrate pocket [71]. Thus, a bidentate compound was generated by joining the two ligands together through an *ad hoc* designed linker [71]. Interestingly, the final compound was characterized by a significantly higher potency, if compared to the two separated molecules [71].

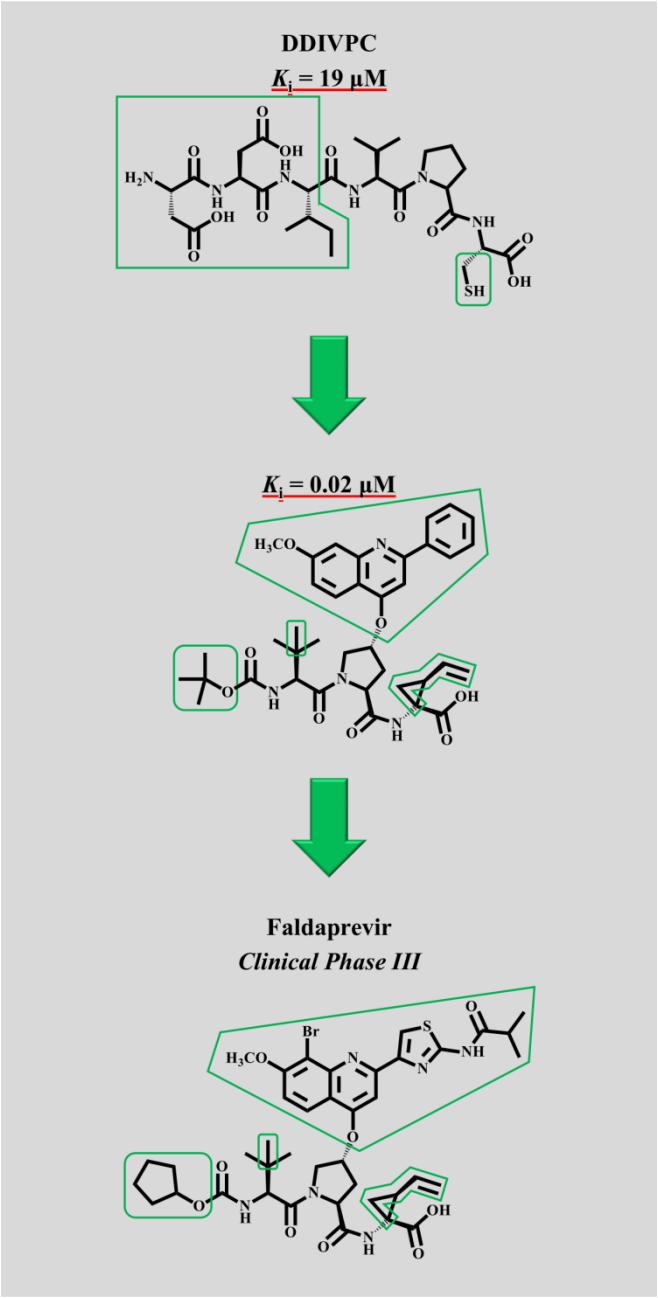


Fig. (**2**). Design of inhibitors of HCV NS3 protease starting from the DDIVPC peptide [90]. The conserved peptide scaffold is represented by the groups of atoms not included in the green boxes.

Other studies conducted on HCV NS3 protease stressed out how transferred NOESY, line broadening experiments and ^{19}F NMR combined together provide a detailed picture of the conformational and interaction properties of peptide inhibitors. Again while for a short flexible peptide 2D [^{1}H, ^{1}H] ROESY can be employed to get structural features of the free state, the conformational properties of the bound peptide

can be investigate by transferred NOESY [33] and line-broadening experiments [92]. These experiments are important to reveal the structure details characterizing the bound state that are absent in the free state and thus gain insights on how to optimize the free state conformation, as explained before [92]. Using a ^{19}F labeled inhibitor ^{19}F NMR was implemented to verify that the peptide bind the protein forming one single complex (using a 1 : 1 protein : inhibitor ratio), this information is not accessible through transferred NOESY as the latter is recorded using a large excess of inhibitor with respect to the protein and thus in presence of multiple conformations, it provides only an average picture [92].

The reported examples highlight the prominent multifaceted information that can be accessed through a vast array of NMR experiments to support antiviral hit discovery and optimization.

2.2.4. Other studies

NMR can be employed also to determine the formation of aggregates in aqueous solutions [69].

Line broadening can affect the spectra of aggregated species and, as explained in paragraph 1.3, the sign of NOE can be supportive too. More in detail, sharp resonances are associated with fast-tumbling small molecules whereas, broad resonances are associated with slow-tumbling aggregated compounds [69]. As concerning NOESY spectra, small molecules present small positive (or absent) NOEs and aggregated compounds negative NOEs [38, 69]. Consequently by acquiring 1D [^1H] spectra of a small organic compound at increasing concentrations and looking at changes in chemical shifts, line width (and consequently intensity), peaks shape, number of peaks, it is possible to find out molecules with a propensity to aggregate [69]. Changes in the appearance of a 2D [^1H, ^1H] NOESY spectrum as function of compound concentration are similarly suggestive of aggregation [69]. This information is important if these compounds are implemented in libraries for drug discovery studies or if we are dealing with small molecule drugs. In fact, such an NMR approach was tested on 11 HCV NS5B Polymerase Inhibitors finding a strong correlation between aggregation and promiscuity (i.e., "off target effect") [69].

Atropisomerism represents an axial dynamic chirality type [93, 94]. Among US FDA (Food and Drug Administration)-approved drugs there are many experimental compounds provided with stable atropisomers, usually characterized by large differences in biological activities [93]. NMR was described also as a tool to evaluate the existence of atropisomer chirality. This is the case of an anthranilic acid series of inhibitors targeting the allosteric thumb pocket 2 of the HCV NS5B polymerase [95]. Simple acquisition of ^1H NMR spectra at different temperatures was not enough to indicate the presence of mirror image enantiomers [95]. Therefore, a second chiral center was added in the starting molecule (Fig. **3**) that coupled to the atropisomerism led to diastereomers clearly visible in 1D [^1H] NMR spectra as characterized by two distinct sets of signals [95]. 2D [^1H, ^1H] ROESY spectra resulted useful to understand the dynamics (fast or slow motions) of bond rotations inducing the different isomers [95].

Interestingly, the compound reported in Fig. **3** inhibits both HCV polymerase and HIV matrix but, X-ray crystallography shows it binds the two proteins into two atropisomeric forms [95]. Indeed, modulation of the rotational equilibrium involving atropisomers gathers a lot of attention in drug discovery as potential way to get selectivity [94].

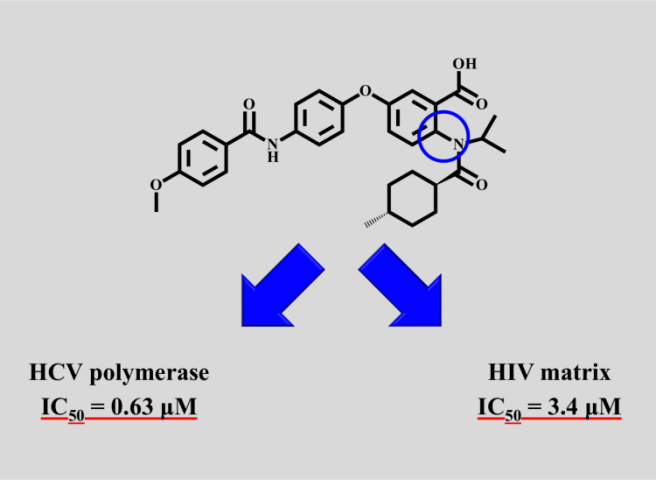


Fig. (**3**). Chemical structure of a compound inhibitor of HCV and HIV. The bond whose rotation leads to atropisomerism is highlighted by the blue circle [95].

1D ^1H and ^{13}C NMR spectra can be useful in medicinal chemistry also to evaluate the purity grade of a compound, to check the "status" of a reaction and to prove that the right product has been obtained. In this context NMR as analytical technique was employed in several medicinal chemistry works focus on inhibitors of HCV NS4B protein [96], of HCV NS5B polymerase [97-99], NS3/4A proteases [100, 101], inhibitors of hepatitis C virus E2 binding to CD81 [102].

3. HEPATITIS B VIRUS (HBV)

3.1. Mechanism of action

HBV represents the best described member of *Hepadnaviridae* family and is characterized by a small and partially double stranded DNA genome [103-106]. The infection by this virus causes acute and chronic hepatitis in humans and its chronic form is associated with a higher risk to develop liver cirrhosis and HCC (HepatoCellular Carcinoma) [103-108]. This virus gathers a lot of attention due to its infectiveness that is indeed 200 times higher than HIV and with a larger mortality incidence [107].

The particle responsible for the infection (known as Dane particle) possesses spherical-like shape of approximately 42 nm; the virion has an envelope composed by a lipid bilayer coming from the membrane of the host cells [107]. The lipids are coupled to three types of HBsAg (Hepatitis B surface Antigen) surface proteins (i.e., S (Small)-HBsAg, M (Medium)-HBsAg and L (Large)-HBsAg) [107].

The viral nucleocapsid is located inside the envelope and is made up by several copies of a Core protein (Cp) known as HBcAg (Hepatitis B core antigen) [107]. In addition, the "ds DNA" genome, which resides in the capsid, is covalently linked to a polymerase provided with DNA-dependent DNA polymerase and reverse transcriptase activities [107]. Interestingly, the serum of patients infected by HBV is characterized also by the presence of spherical or filamentous non-infective particles which serve to fool human's immune system [107].

HBV life cycle includes different stages, such as receptor binding and entry in the host cells, transport of nucleocapsids to nuclear pores so that HBV genomes can be released into the nuclei, creation of ccc (covalently closed circular)-DNA, transcription, pgRNA (pregenomic RNA) encapsidation, reverse transcription, DNA production, envelopment, and release of virions and subviral particles [104, 105]. Necessary for virus entry in hepatocytes is the low-affinity interaction between HSPG (Heparan Sulfate ProteoGlycans), glycoproteins with one or more covalently attached HS (Heparan Sulfate) chains and located on hepatocytes surface, and AGL (AntiGenic Loop) regions belonging to HBV envelope proteins [104, 109, 110]. The next step is the high-affinity interaction between the PreS1 domain of L-HBsAg surface protein and its liver-specific receptor NTCP (Sodium Taurocholate Cotransporting Polypeptide) [104, 107]. After this step, the internalization of the virion, followed by de-envelopment, occurs and thus leads to the release of the bare nucleocapsids into the cytoplasm of the host cells [104]. The following phases are the opening of nucleocapsid at the nuclear pore complex and the consequently entering of viral genome rcDNA (relaxed circular DNA) into the nucleus of the host cell [104]. In the final step, the rcDNA is transformed into cccDNA, the template used for the transcription of different viral mRNA species [104].

Interestingly, the chronic HBV infection depends on the viral replication which in turn is linked to two fundamental features of HBV life cycle [105]. The first one is formation of the small cccDNA in the nucleus of infected hepatocytes whereas, the other element is the expression of the regulatory HBx (Hepatitis B virus X protein) protein. HBx is a crucial protein that re-directs the host machinery to induce cccDNA transcription [105]. Interestingly, a therapy based on inhibition of HBV DNA polymerase is able only to decrease the levels of other HBV replicative intermediates but not those of cccDNA [105]. Therefore, following polymerase inhibitors treatments, viral replication is still able to restart with the transcription of cccDNA [105]. In addition, HBx favors the progression of hepatocellular carcinogenesis by blocking several host signaling pathways, thus suggesting its role as promising drug target for the development of therapeutic routes for HBV disease [105].

One of the strategies proposed against HBV is based on the inhibition of the virus entry step and different inhibitors (e.g., Cyclosporin A, Ritonavir and Ezetimibe) were described for their ability to avoid the interaction of HBV with NTCP, and thus hamper viral internalization [109].

An alternative way to block viral action consists in the design of molecules able to lower the cccDNA levels and different methods were developed [109]. For instance, different DSS (DiSubstituted Sulfonamides) molecules, (e.g., CCC-0975 and CCC-0346) were described as able to hamper the conversion of rcDNA into cccDNA and also to avoid de novo formation of cccDNA [109]. Interestingly, many small molecules for therapies against HBV infection are analogues of nucleosides [108]. Indeed, Adefovir dipivoxil (known as Hepsera) and Tenofovir (known as Viread) are two adenosine analogues whereas, Lamivudine (3TC) (known as Zeffix), Entercavir (known as Baraclude) and Telbivudine (known as Tyzeka) are analogues of cytidine, guanine and thymidine, respectively [108].

It was also proposed the use of endonucleases, such as CRISPR/Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats/ CRISPR associated protein 9), to cleave and eradicate cccDNA from infected host cells [109].

The viral infection can be blocked also by siRNAs (small interfering RNA) which are short sequences of nucleotides able to promote the gene silencing at the post-transcriptional level and thus favor a decrease in the expression of the viral genes [109]. Indeed, valuable examples of this approach can be found in ARC-520 (phase II/III clinical trials) and TKM-HBV/ARB-001467 (phase II clinical trials) [109].

The nucleocapsid of HBV was described as fundamental for different processes (e.g., packaging of viral genome, reverse transcription, intracellular trafficking and maintenance of chronic infection) and different molecules able to damage the nucleocapsid (e.g., morphothiadine mesilate GLS-4 in phase II clinical trials in China and the sulfamoylbenzamide NVR 3-778 that has completed phase 1) were identified as promising antiviral agents [109].

3.2 NMR investigations

Similarly to HCV, several NMR studies were reported to investigate HBV and its mechanism of action. Indeed, NMR delivered crucial structural information on proteins involved in viral infection like the surface antigens preS1 [111, 112] and preS2 [113], HBx [114] and XIP (HBx-Interacting Protein) [115]), viral DNA [116] and RNA regions [117].

In addition, a few works described the use of NMR to help medicinal chemistry efforts in the synthesis of inhibitors of HBV infection [118-120] and on NMR-based metabolomics strategy [84, 121, 122] to implement novel diagnostic tools.

3.2.1. Structural and Interaction studies

Pre-S proteins play an important function linked to virus assembly and entrance into the host cell [123]. NMR provided much structural information on pre-S1; this protein recognizes specific receptors that are involved in the engagement of HBV to human hepatoma HepG2 cells and human liver plasma membranes [111]. The conformational analysis of a synthetic peptide comprising the fragment A20-P47 of the N-terminal region of preS1 protein, that contains the receptor binding sites, was conducted through NMR experiments [111]. More in detail, combined analysis of different spectra (DQF-COSY (Double Quantum Filtered COrrelation SpectroscopY), TOCSY (TOtal Correlation SpectroscopY), E.COSY (Exclusive Correlation Spectroscopy) and NOESY) led to achieve resonance assignments for this peptide and to obtain 157 distance and 55 dihedral angle constraints [111]. These restraints were exploited in NMR structural calculations [111]. NMR data allowed the identification, inside the receptor binding region (i.e., P21-P47), of four reverse turns encompassing peptide segments L22-F25, N37-N40, P41-D44 and W43-N46 [111]. Immunological studies showed as well the presence of three main antibody binding sites inside the same region: 16-27, 32-53, 41-53. Consequently, NMR studies pointed out that turn conformation may be relevant for antigen-antibody recognition and for receptor binding [111].

A more recent work was conducted on the entire pre-S1 protein (119 residues-long), through heteronuclear NMR techniques, and revealed further structural insights [112]. Pre-S1 is largely unstructured with no tertiary structure due to the poor chemical-shift dispersion in both dimensions in the 2D [^{1}H, ^{15}N] HSQC spectrum recorded under non-denaturing experimental conditions [112]. The absence of tertiary structure was also confirmed by hydrogen/deuterium exchange experiments, by observing the disappearance of all backbone amide protons within 10 min after addition of $^{2}H_2O$ [112]. Interestingly, the joint analysis of different NMR spectra allowed achieving resonance assignments and provided a detailed structural portrait. The occurrence of d_{NN}, $d_{\alpha N}(i + 2)$, $d_{\alpha N}(i +3)$ NOE contacts along with specific $^{3}J_{HNH\alpha}$ coupling constants values and the Hα CSI (Chemical Shift Index) profile suggested the presence of transient local structural elements in the N-terminal region of pre-S1 ranging from M12 to D50 [112]. NMR data indicated two helical turn motifs encompassing fragments P32-A36 and P41-F45 and amide protons temperature coefficients indicated as well a set of hydrogen bonds in the region N40-K49 [112]. Interestingly, P32-A36 and P41-F45 regions are included in the P21-P47 segment identified as the putative HBD (Hepatocyte-Binding Domain), thus suggesting for the entire pre-S1 protein the simultaneously presence of a mainly large natively unstructured state and N-terminal regions with certain defined structures and involved in the binding to hepatocytes [112].

In another work the attention was focused on the structure of a region of pre-S2 (N1-G23 and named "adrl23") which is recognized by major B cells and is involved in the neutralization of HBV infection [113]. More in detail, a peptide with all residues of adrl23 (NSTTFHQALLDPRVRGLYFPAGG) was synthesized and different NMR experiments were conducted to get unambiguous resonance assignments and distance and angle constraints needed to calculate the NMR solution structure [113]. These studies revealed for adrl23 a fold characterized by a "L" shaped helix-turn-helix topology in which an α-helix encompasses the S2-Q7 region at the N-terminus, a 3_{10} helix is formed by the V14-Y18 segment at the C-terminus and two β-turns are including A8-D10 and D11-V13 fragments between the two helical structures (Fig. **4**) [113]. Interestingly, the site recognized by major B cell is partially included in the two β-turns [113]. In addition, the structure seems to be stabilized by a set of hydrophobic interactions which involve residues F5, L9, L10, V14, and Y18 located on one side of the protein and by F19 that is inserted in the concave side of the protein (Fig. **4**) [113]. The shape of the protein is stabilized by a H-bond network which involves T3, H6, and R15 residues; in addition the N-terminal end is made immobile by the H-bond formed by the backbone amide proton of N1 and the hydroxyl oxygen of T4 (Fig. **4**) [113]. The same work showed also that the side chains of residues D11, R13, V14, L17 and Y18, which play a fundamental role in the interaction between adr123 and a monoclonal antibody H8 mAb, are solvent exposed [113].

These structure features obtained by NMR could be exploited for the development of potential antiviral agents targeting pre-S2 protein of HBV [113].

The preS and S regions of the HBV large surface antigen (L-HBsAg) play an essential role in the assembly of the infectious virion, by binding in a synergic manner the viral core antigen (HBcAg) [118]. Consequently, peptides encompassing the preS and S sequences of L-HBsAg are considered possible inhibitors of viral assembly [118]. Two peptides including the sequences of these regions (RQPTPLSPPLRTTHPQAMHWNSTTF for preS peptide–made up of 17 residues from the C-terminal preS1 site plus 8 residues from the N-terminal preS2 site) and PISNHSPTSCPPTCPGYRWMCLRRF for S peptide) were synthesized and their structural features were investigated by NMR spectroscopy. Complete resonance assignments were achieved through analysis of homo-nuclear and hetero-nuclear NMR experiments and the resulting chemical shifts for Cα, Cβ, C′ and Hα atoms were analyzed by the CSI method [33, 118]. In detail, CSI data showed the tendency of the entire preS peptide and of the N-terminal portion of the S peptide to assume a β-strand like structure whereas, the C-terminal region of the S peptide showed a propensity towards an α-helical structuration [118]. The solution structures of the two peptides were calculated from NMR restraints and showed that peptide preS contains several β-turns structured regions at difference from the peptide S that

has instead a single structured segment; in both peptides bends flank the turn structures [118]. STD NMR provided epitope mapping and revealed the residues of preS (i.e., R1, H14, H19, W20, F25) and S peptides (I2, Y17, W19, M20, L22, R24 and F25) important for the interaction with HBcAg [118]. Nevertheless, docking analyses conducted with the two peptides and HBcAg indicated that they associate with the immunodominant region of HBcAg positioned in the viral capsid spikes [118]. These structure details are important to develop molecules able to disrupt this interaction and thus behave as a promising antiviral agent against HBV [118].

NMR and particularly solid-state techniques can be supportive to other structural biology tools to study capsid assembly [124]. The power of NMR is linked to the sensitivity of chemical shifts (mainly ^{13}C chemical shifts) to backbone dihedral angles, and to the possibility to identify conformational changes through chemical-shift perturbations, generally following addition of a ligand or upon formation of a complex. When a single system is characterized by several conformations (or more general crystallographically non-equivalent forms), this results in the observation of peak splitting in the spectra [124].

The HBV particle is made up by an external envelope and an inner nucleocapsid that is in turn constituted by a self-assembling core protein (Cp) with a molecular weight of 20 kDa [124]. The first 140 residues in Cp are sufficient to direct capsid assembly. The icosahedral nucleocapsid of HBV is pull together from 240 core protein molecules, that are indistinguishable from a chemical point of view, and includes four sets of symmetrically different subunits [124]. This asymmetry is reflected in solid-state NMR spectra of the capsids, in which a certain peak splitting is observed. First, cryoEM (cryo-Electron Microscopy) and crystallography data showed the structure of Cp149, revealing a mostly α-helical protein which dimerizes generating a four-helix bundle [124-126]. When the capsid assembles, the capsid coat is made through lateral dimer-dimer interactions whereas, the helix bundles look as perpendicular dimeric spikes punctuating it [124]. In detail solid state NMR spectra showed resonance splitting only for a subset of residues [124]. Almost all these residues were located close to the fivefold and quasi-sixfold interfaces formed by the subunits of capsid whereas, residues within the spikes of the capsids were not characterized by this splitting [124]. In addition these residues were distributed almost equally between the flexible loops and towards the start or end of the base plate helices [124].

In an interesting work, a strategy based on cell-free protein synthesis and solid-state NMR was set up to produce and investigate the assembly of higher-order complexes straight after exiting from the ribosome [25]. The approach was tested by implementing as model-system the entire HBV Cp183 capsid protein and revealed that this protein led to icosahedral capsids which are identical to those obtained by bacterial expression [25]. In addition, spectra were characterized by signal-to-noise ratios which allowed them to provide information on conformational changes of capsids [25]. If compound modulators of assembly are added to the cell-free reaction, they produce outcomes identical to those that can be retrieved by adding the compounds to pre-made Cp183 capsids. In summary, the strategy can be employed to better assess capsid assembly regulation right after synthesis by ribosome, and provide a clever route to analyze changes induced on capsids structures by natural or synthetic molecules, and eventually enzymes that accomplish post-translational modifications [25].

As described above, HBx is a protein which is expressed in liver cells that have been infected by HBV and holds great interest as it favors the activation of cccDNA transcription although it has a variety of functions [114]. The activities played by this protein are related to its ability to interact with different partners. To clarify the structural determinants of the multifaceted interaction properties, the structure of a truncated version of HBx (named Tr-HBx) was studied by different techniques, including NMR spectroscopy [114]. The 2D [^{1}H, ^{15}N] HSQC spectrum of Tr-HBx was characterized by a low chemical shift dispersion in the ^{1}H dimension which indicated an intrinsically unstructured/unfolded state for the protein [114]. More in detail, the number of cross peaks that appeared in the spectrum resulted lower than expected for this protein, without considering prolines and the last N-terminal residue [114]. These data indicated the coexistence of a N-terminal disordered region along with a more structurally organized C-terminal protein portion [114].

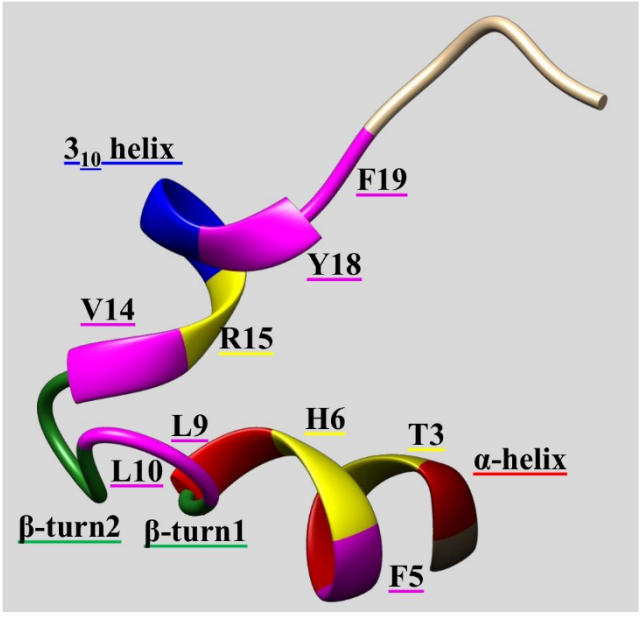


Fig. (**4**). NMR solution structure of adrl23 (first conformer, PDB code: 1WZ4 [113]). The α-helix (S2-Q7) is reported in red, the two β-turns (β-turn1 (A8-D11) and β-turn2 (D11-V13)) in green and the 3_{10} helix (V14-Y18) in blue. The residues involved in hydrophobic interactions (i.e., F5, L9, L10, V14, Y18, and F19) are colored in magenta whereas, the residues

participating in the H-bond network (T3, H6, and R15) are reported in yellow.

Interestingly, there exists a protein which can block the transactivation functions of HBx [115]. This protein, named XIP, was studied by NMR under physiological conditions and in an hydrophobic environment (reproduced by using SDS (Sodium Dodecyl Sulfate)) to get information on its structural behavior and on the structural elements important for binding to HBx [115]. The 1D ^{1}H spectrum of XIP in presence of SDS was characterized by a higher chemical shift dispersion of amide protons, if compared to the one obtained in absence of SDS, thus pointing out that a structural transition occurs upon addition of the ionic detergent to XIP [115]. The process of resonance assignments for XIP in SDS was achieved mainly by standard 3D NMR experiments [115]. Then, a CSI analysis was conducted on the basis of Cα, Cβ, CO and Hα resonances and showed the presence of two rather long α-helices (regions from A42 to K54 and from I82 to S91) and two short helices (regions from H8 to T12 and from G32 to S38) [115]. These results were employed to design different XIP mutants which were studied by *in vitro* binding assays [115]. GST (Glutathione S-transferases)-tagged XIP constructs were implemented for this study; control experiments were of course conducted and demonstrated that GST alone was unable to bind HBx. Interaction assays demonstrated that the entire XIP sequence was necessary for binding to HBx as the interaction was abolished in any truncation mutant. These studies highlighted the important role of the whole helical fold in XIP for its association to HBx [115].

Another interesting work, was focused on the HBV genome which, as described above, is characterized by "ds DNA" and represents the smallest known genome of a DNA virus (3.2 kilobases) [116]. 1D [^{1}H] and 2D [^{1}H, ^{1}H] NMR spectra were employed to study the duplex ([d(GGCAGAGGTGAAA)-d-(TTTCACCTCTGCC)]). This duplex contains the 11-base-pair direct repeat sequence of viral genome and an additional base pair from the HBV genome on each end to decrease end effects on the sequence [116]. Interestingly, NMR data pointed out a B-DNA structure for the studied sequence; in addition sugar pucker investigation and imino exchange rate studies evidenced a particular dynamics in the region centered around the G8-C19 and G7-C20 base-pairs in which the sugars are characterized by a S conformation for only 60-65% of the time [116].

A crucial step for virus replication is represented by reverse transcription of HBV pregenomic RNA. At the initial stage the HBV reverse transcriptase interacts with the highly conserved encapsidation signal, epsilon (ε), positioned close to the 5' terminus of the pregenome. NMR techniques were employed to study the structure of this ε encapsidation signal [117]. For this RNA region it was predicted a structure composed by a bulged stem-loop with the apical stem capped by a hexa-loop. In this study, isotope labeling was exploited to study the structural features of the ε signal in great details. The apical loop was found to be capped by a UGU tri-loop with a CG closing base, instead of the predicted hexa-loop [117]. In addition, an unpaired C near to the CG closure of the loop and an unpaired U, that is partially intercalated into the stem were detected within the apical stem loop [117]. Interestingly, the loop region was associated to a considerable number of NOE contacts and narrow resonances which suggested the absence of conformational exchange in the millisecond or slower time scale and thus the presence of a well-defined stable conformation [117]. This evidence seems to point out that the loop region with its stable fold could play a crucial role in the interaction with the HBV reverse transcriptase [117]. Noteworthy, analysis of a large number of HBV sequences, so to cover as much as possible all genotypes, highlighted that the key residues for the above described fold of ε were mainly conserved with only rare non-disruptive mutations [117]. Therefore, based on these NMR structural features it could be envisioned to design compounds able to target the ε encapsidation signal and work as inhibitors of viral replication [117].

3.2.2. Metabolomics studies

Different studies demonstrated the validity of NMR as a technique to identify biomarkers specific for diseases associated with HBV infection and thus, to define novel less-invasive diagnostic tools [84, 121, 122, 127-130].

As already reported in paragraph 2.2.2. NMR-based metabolomics is useful in revealing differences in the type and levels of metabolites in HBV infected patients compared to those affected by HCV infection [84].

Nevertheless, NMR-based serum metabolomics can be used to study the progression of LC (Liver Cirrhosis) in HBV infected patients [127, 128]. LC is a severe liver pathology, often induced by HBV, that can degenerate into hepatocellular carcinoma. Metabolic changes in the serum of healthy individuals, patients infected with HBV and those with HBV-LC were analyzed by a solid NMR-based approach. 1D [^{1}H] NMR spectra were acquired to achieve the metabolomics analysis and pulse sequence to reduce as much as possible the line-broadening effect due to the presence in the samples of macromolecules such as proteins and lipids were set-up; ambiguities in resonance assignments were solved by recording 2D [^{1}H, ^{1}H] COSY and 2D [^{1}H, ^{1}H] TOCSY [127]. Complex statistical analysis was exploited to get a general picture of the metabolic profiles of patients infected with HBV, HBV-LC and control subjects without pathological conditions and, to conduct comparative analysis [127]. Interestingly, serum histidine resulted a possible biomarker to distinguish HBV patients from healthy ones whereas, a set of biomarkers (i.e., acetate, formate, pyruvate and glutamine) in serum was associated to the transition from a simple HBV infection to a HBV infection

with liver cirrhosis. Furthermore, this work led to the identification of an extra set of biomarkers (i.e., phenylalanine, unsaturated lipid, n-acetylglycoprotein and acetone) which could be used to follow the progression of cirrhosis [127].

To get further insights into the biomarkers of liver diseases, a similar protocol was adopted to analyze the saliva metabolic profile of patients affected by HBV chronic infection [121]. 1D [^{1}H] NMR spectra of 16 saliva samples from healthy patients and 20 saliva samples from patients affected by chronic hepatitis B infection were recorded and analyzed. Results indicated increased levels of propionic acid, putrescine, acetic acid, succinic acid and tyrosine and lower levels for L-lactic acid, butyric acid, pyruvate, 4-pyridoxic acid and 4-hydroxybenzoic acid under HBV induced pathological chronic conditions [121]. These metabolites were linked to different metabolic pathways (i.e., phenylalanine metabolism, glycolysis or gluconeogenesis, pyruvate metabolism, propanoate metabolism, butanoate metabolism, taurine and hypotaurine metabolism, TCA cycle, tyrosine metabolism, and alanine, aspartate, and glutamate metabolism) which could be thus employed to discriminate the infected patients from the healthy ones [121]. For instance, this study showed that the HBV chronic condition led to higher levels of succinate, a fundamental substrate of TCA cycle, and lower levels of pyruvate and lactate, two products of glycolysis [121]. Another valuable example is represented by 4-pyridoxic acid, the catabolic product of vitamin B6 whose presence is due to the combined action of aldehyde oxidase (endogenous enzyme) and pyridoxal 4-dehydrogenase (microbial enzyme) [121]. Decreased levels of 4-pyridoxic acid could be linked to weakened metabolism of vitamin B6, variations in the levels of oral microbiota or lower amounts of vitamin B6 in patients affected by the chronic disease. Vitamin B6 possesses antioxidant and anti-inflammatory properties therefore, it was speculated that the host could employ vitamins to lower hepatic inflammatory stress [121].

This interesting study pointed out that diagnosis based on the saliva metabolomics is useful not only to distinguish between patients affected by chronic hepatitis B and healthy people but also to better understand mechanisms at the base of the disease.

NMR-based metabolomics was extensively used as method to determine biomarkers for HCC (HepatoCellular Carcinoma) whose major risk factors are HBV and HCV viruses [122]. In this context metabolic profiles were analyzed in HepG2.2.15 cells -arising from HepG2 (a human liver cancer cell line)- with stable expression and replication of HBV, that were employed as model-system [122]. Results of such metabolomics studies revealed that HBV infection could lead to HCC through upregulation of specific pathways: GFAT1(Glutamine-Fructose-6-phosphate AmidoTransferase 1)-activated hexosamine biosynthesis and CHKA (CHoline Kinase Alpha)-activated phosphatidylcholine biosynthesis [131].

HBx is an oncoprotein, that is related to HBV-linked HCC, and as briefly mentioned in previous paragraphs plays a function in HBV replication as well as DNA repair, transcriptional regulation, progression of cell-cycle [122]. An NMR-based metabolomic approach was employed to analyze the relationship between HBx and HCC [132]. HBx was found to interfere with the metabolism of glucose, lipids, amino acids, and largely nucleic acids. Based on these data a mechanism by which HBx could induce HCC has been proposed: the protein first produces DNA damage and next, by interfering with nucleic acid metabolism, inhibits DNA repair thus leading to cancer [132].

Additional work was focused on the NMR-based metabolomics analysis of urine samples coming from 43 Bangladeshi patients with HCC from HBV infection (group named HCC-HBV), 50 patients with a hepatitis B related cirrhosis (group named CIR), 48 patients affected by non-cirrhotic chronic hepatitis B related liver disease (group named CHB) and 8 people with no history of liver disease from the same Bangladeshi population (group named CTR) [129]. This work led to the identification of different metabolites (i.e., acetate, creatine, creatinine, dimethyamine (DMA), formate, glycine, hippurate, and trimethylamine-N-oxide (TMAO)) whose levels changed on the basis of the considered group of patients [129]. Indeed, significantly higher levels of carnitine and lower levels of creatinine, hippurate, and TMAO were found in HCC patients in comparison with the other groups [129]. A correlation between HBeAg, (Hepatitis B e-Antigen), a protein whose levels affect the disease progression, and metabolites levels was found [129]. Indeed, CHB patients with HBeAg, showed significant lower levels of creatinine if compared with CHB patients without HBeAg, whereas CIR members with HBeAg were characterized by higher levels of DMA, if compared to the CIR members without HBeAg [129].

BAY41-4109 (methyl(R)-4-(2-chloro-4-fluorophenyl)-2-(3,5-difluoro-2-pyridinyl)-6-methyl-1,4-dihydropyrimidine-5-carboxylate) is a member of HAP (2-Heteroaryl Pyrimidines) with a potent antiviral activity against HBV (Fig. **5**) [130]. Interestingly, anti-HBV agents are characterized by hepatotoxicity [130]. Indeed, high resolution ^{1}H NMR-metabolomics studies were conducted with an array of urine, serum and liver tissue aqueous and lipid extracts coming from BAY41-4109 treated rats and control groups, to get information on mechanism of compound hepatotoxicity [130]. For instance, urine and serum samples of rats treated with BAY41-4109 were characterized by an increased level of 3-HB (3-hydroxybutyric acid), which is one of the ketones produced by fatty acids metabolism in liver mitochondria [130]. Moreover, aqueous extracts of liver tissue from animals treated with the antiviral agent, were characterized by higher levels of lactate, 3-amino-isovalerate, pyruvate, choline, TMAO (trimethylamine-*N*-oxide) and lower levels of taurine, hippurate and D-glucose whereas, chloroform/methanol extracts of liver tissue from treated

animals revealed increased amounts of different lipid signals (e.g., triglyceride terminal methyl, methylene groups, and CH_2CO, $N^+(CH_3)_3$, CH_2OPO_2, CH_2OCOR) [130]. Therefore, it was suggested that the hepatotoxicity of BAY41-4109 could be largely related to its ability to alter fatty acid metabolism and mitochondrial functions [130].

3.2.3. NMR as an analytical technique to characterize synthetic and natural anti-HBV agents

HAPs (HeteroAryldihydroPyrimidines) (e.g., BAY41-4109, GLS-4 and HAP-12) are known as compounds able to block replication of HBV both *in vitro* and *in vivo* (Fig. **5**) [133]. HAPs interfere with capsid assembly; indeed at the highest concentrations HAPs induce formation of abnormal non-capsid polymers whereas, at stoichiometric levels they also allosterically lead to assembled-active forms and favor the rate of capsid assembly [133]. Thus, based on compounds concentrations, a competition occurs between active functional assembled states and misassembled ones [133].

The HBV nucleocapsid is made up by homo-dimers formed by the core protein with a mainly α-helical fold and with the dimeric interface including a four helix bundle, and crosslinked by a disulfide bridge involving a Cysteine positioned at the center of the bundle.

The binding of HAPs to HBV depends mainly on non-polar contacts and they seem to interfere with interactions between capsid proteins, causing quaternary changes and presumably misassembly [133]. Crystallographic studies of HBV capsid bound to HAP1 revealed that this compound produces movements and global structural variations of subunits by binding to a hydrophobic pocket at the dimer-dimer interface [133, 134]. In major detail, the HAPs binding-site is located at the dimer-dimer interface close to the C-terminal sides of the HBV core protein units and two adjacent dimers contribute to it [133].

To optimize potency and other properties such as oral bioavailability, many HAP-based HBV inhibitors were designed and synthesized by means of SAR (Structure-Activity Relationship) and SPR (Structure-Property Relationship) studies [133, 135-137]. An example is provided by 5-halogeno-heteroarylpyrimidines analogs with anti-HBV activities in the low micromolar range but also characterized by some toxicity observed in certain cell lines [136]. However, these molecules appear promising starting points in the drug design route to get original potent and safer antiviral agents for HBV [136]. Nevertheless, by modulating the dihedral angle between the C4 phenyl and the dihydropyrimidine core of HAP inhibitors it was possible to obtain novel compounds able to inhibit more efficiently the HBV capsid and to reduce the induction of CYP3A4 [119]. Induction of CYP3A4 is a crucial clinical problem as this enzyme acts in the metabolism of marketed drugs and can lead to the reduction of therapeutic efficacy [119].

The PPA (PhenylPropenAmide) and pyridazinone derivatives represent other two classes of HBV inhibitors; although chemically and structurally distinct from HAPs they share the same binding pocket on the HBV capsid [133]. PPAs work as assembly accelerators and induce a faster capsid assembly and formation of defective capsids deprived of genetic material, but without affecting capsids morphology [133]. Examples of PPA and pyridazinone derivatives are provided by the compounds AT-130 (Fig. **5**) and 3711, respectively [133]. It was shown that 3711 lowers HBV DNA by leading to generation of genome-free capsids in cell culture system [133].

Other interesting groups of HBV inhibitors include the SBA (Sulfamoy Benzamide) derivatives and the class of 2-amino-n-(2,6-dichloropyridin-3-yl) acetamide based analogues [120, 133]. A crucial function of HBV capsid is to synthesize DNA after packaging pgRNA. SBA inhibits selectively the pgRNA-containing nucleocapsids of HBV and one of the most advanced compounds of the SBA series from a clinical proof of concept is NVR 3-778. Interestingly, the activity of NVR 3-778 improves nine times if a primary amine group is located at the C-6 position of the core phenyl ring (Fig. **5**) [120]. Among 2-amino-n-(2,6-dichloropyridin-3-yl) acetamide derivatives, the compound BCM-599 blocks efficiently capsid assembly and enhances the level of free Cp dimer [133].

In addition, there are Bis-ANS (5,5'-bis[8-(phenylAmino)-1-NaphthaleneSulfonate]), that is a fluorescent probe interacting with protein hydrophobic sites and is known for its inhibition of HBV capsid assembly induced by ionic strength and, NZ-4 (isothiafludine), which is responsible for the formation of capsids lacking genetical material [133].

Searching for novel antiviral agents, a series of oxime ethers derivatives were designed based also on results from docking studies and synthetized, their anti-HBV activities were evaluated with HepG2.2.15 cells *in vitro* [138]. Results pointed out that many compounds in the oxyme analogue series presented poor cytotoxicity on HepG2.2.15 cells and could relevantly block HBsAg and HBeAg secretion [138].

Natural products represent also a reservoir of molecules with antiviral activity [139-143]. For instance, panaxadiol and panaxatriol are compounds that can be purified from ginseng, and were employed as starting point to design anti-HBV agents [139, 140]. Different analogues were synthesized and their anti-HBV activities were investigated on HBsAg and HBeAg secretions, along with HBV DNA replication on HepG 2.2.15 cells [139]. SAR studies evidenced the functional groups needed to achieve enhancement in the anti-HBV activity [139].

Natural diterpenoids (dehydroandrographolide and andrographolide) from the plant *Andrographis paniculate* possess anti-HBV activity and were used also as starting point to synthetize a novel group of antiviral agents with improved properties [144]. In addition, different molecules with anti-HBV activities were obtained from three medicinal

plants: *Wikstroemia chamaedaphne Meisn*, *Hypericum japonica* and *Herpetospermum caudigerum* [141-143].

In the studies reported here 1D [^{1}H] and [^{13}C] NMR spectra were recorded to characterize synthetized compounds (identity and purity assessment); in addition NMR spectra were acquired to analyze plant extracts and clarify the chemical structures of major components but this task required as well extensive use of 2D heteronuclear correlation experiments [133, 136, 139-144].

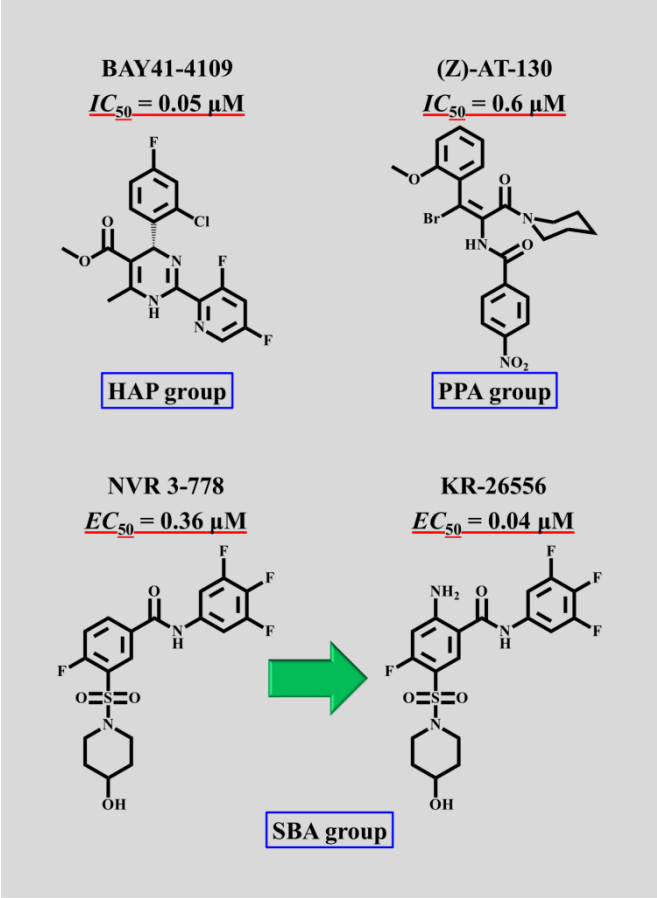


Fig. (**5**). Chemical structures of a few anti-HBV agents. IC50 values were taken from reference [133] and EC50 values from reference [120].

4. HERPES SIMPLEX VIRUS

4.1. The virus and its mechanism of action

HSV (Herpes Simplex Virus) belongs to *Herpesviridae*, a family of viruses that also comprises HCMV (Human CytoMegaloVirus), VZV (Varicella-Zoster Virus), and HHV-6 (Human HerpesVirus 6), HHV-7 and HHV-8 [145-153]. This virus possesses a "ds DNA" genome and it exists in two types: HSV-1 and HSV-2, that can be distinguished from each other for the antigenic variations in their envelope proteins [146]. Although, HSV-1 is usually associated with oral-labial disease (e.g., oral herpes) and HSV-2 with genital disease (e.g., genital ulcerations), a number of cases where genital herpes was associated with HSV-1 have been described [146]. Interestingly, the transmission of HSV-1 infection usually occurs in childhood whereas, HSV-2 infection is associated with sexual debut [152].

HSV particles are made up of an outer host-derived lipid envelope covered with different glycoproteins, an intermediate proteinaceous layer named tegument, the nucleocapsid and the viral core composed mainly by the "ds DNA" genome [145, 149]. The capsid of Herpes virions is large and icosahedral and contains a portal at a distinctive 5-fold vertex. The portal represents a sort of molecular motor that allows the viral genome to come into the capsid during virion morphogenesis and it also consents the genome to exit the capsid to be injected inside the host nucleus to start infection [149].

The binding and entry of HSV into host cells is managed by the envelope glycoproteins gB, gD, and gH-gL [147]. Following initial binding, the gD glycoprotein associates with different high-affinity receptors allowing viral entry [147]. HveA (Herpesvirus entry mediator A) – also known as TNFR14 or HVEM- represents a trimeric transmembrane receptor belonging to the tumor necrosis factor (TNF) family and was described as the first receptor associated with HSV. The second and third groups of gD receptors are nectins (i.e., different members of the immunoglobulin superfamily) such as nectin-1 (HveC), and receptors formed upon sulfation of heparan sulfate operated by specific 3-O-sulfotransferases, respectively. The next step is the fusion between viral envelop and host cell membrane followed by delivery of nucleocapsid and tegument into the cytoplasm [147, 149]. Then, tegument and nucleocapsid move to the cell nuclear membrane and anchor on the nuclear pore complex which in turn favors the entering of viral genome into the nucleus by the unique portal-vertex [145, 149]. The premade tegument proteins are as well released, and start playing crucial functions in neutralizing host defenses and inducing expression of viral genes. VP16 is a relevant tegument protein that, by engaging transcription factors and RNA polymerase II to the IE promoters, induces immediate-early (IE) gene expression thus enhancing the progressive flow of events at the bases of viral gene expression [154, 155]. One of the mechanisms by which HSV-infected cells escape from the cellular immune system relies on the ICP47 (Infected Cell Protein 47) protein [156]. Under normal conditions, portions of viral proteins are caught by TAP (Transporter associated with Antigen Processing) and carried to the ER (endoplasmic reticulum) where these peptide portions are loaded onto MHC (Major Histocompatibility Complex) class I molecules by the different chaperones [156]. Stable MHC complexes move to the cell surface where a checking of their cargo by cytotoxic T-lymphocytes brings eventually to cell lysis [156]. In the case of HSV-infected cells, ICP47 inhibits TAP and thus allows the virus to escape the immune system [156].

4.2. NMR investigations

NMR techniques gave a considerable support to improve knowledge on different aspects of HSV infection including elucidating structural features of several viral proteins and related interaction networks necessary for a proper functioning of viral machinery [157-160]. Furthermore, NMR analyses were applied during antiviral drug discovery campaigns to characterize novel anti-HSV molecules [161-165].

4.2.1. Structural and Interaction studies

NMR structural studies of different proteins involved in HSV infection have been reported, including ICP47 that, as explained in the previous paragraph, is important to escape the host immune system [156].

Interestingly, the ICP47 3-34 fragment (i.e., 3-WALEMADTFLDNMRVGPRTYADVRDEINKRGR-34) retains the activity of the entire protein [156]. Conformational analyses by CD and NMR spectroscopy revealed that the 2-34 fragment lacks an ordered structure in aqueous buffer while assuming helical structuration in presence of a membrane mimetic environment like that induced by TFE and SDS (Sodium Dodecyl Sulfate). An accurate NMR analysis of a synthetic peptide reproducing the region 2-34 of ICP47 was conducted in micellar solution of deuterated SDS; sequence-specific resonance assignments were firstly achieved by joint analysis of 2D [^{1}H, ^{1}H] TOCSY and 2D [^{1}H, ^{1}H] NOESY spectra next, distance constraints were collected from the NOESY experiment and implemented to calculate the 3D structure. The NMR solution structure in presence of SDS is made up of an helix-turn-helix motif in which the first α-helix encompasses the 4-15 region and the second α-helix includes the 22-32 segment (Fig. **6**) [156] and both helices are amphipathic. The structural features of ICP47(2-34) are in agreement with an interaction model where ICP47(2-34) binds at the surface of the membrane without undergoing a transmembrane entrance [156]. According to the observed structural features, in a first step ICP47 may be absorbed by the ER membrane, thus creating an elevated local concentration of protein with a well-organized structure. This membrane induced fold of ICP47 may be important to optimally position side chains for the interaction with TAP. In a second stage ICP47 binds TAP thus inhibiting the peptide interaction site [156]. This represents an example of how NMR techniques conducted in solution under a precise lipid-like environment can provide structural information crucial to get mechanistic insights.

In another work, the attention was focused on the region of gH glycoprotein associated with the highest fusogenic activity (i.e., 625-HGLASTLTRWAHYNALIRAF-644) [158]. Structural studies were conducted in presence of DPC (DodecylPhosphoCholine), that is able to form micelles and used again to recreate a membrane-like environment. The presence of several alanine residues along the sequence allowed to implement ^{15}N selective alanine labeling to conduct a more precise NMR investigation through heteronuclear correlation experiments [158]. In fact, assignment of ^{1}H and ^{15}N resonances could be achieved by 2D [^{1}H, ^{15}N] HSQC, ^{15}N-filtered TOCSY and NOESY experiments. The distance constraints obtained by NOESY spectra of the peptide recorded in presence of DPC were employed in the structural calculations and the results showed that DPC favors the presence of an α-helix in the peptide fragments 627-635 and 638-641 and a kink between the two helical segments [158]. A comparison between the calculated peptide structure and that of the entire protein highlighted the absence of relevant differences with the only exception of the angle between the two α-helices that diverges due to the lack in the peptide of long-range interactions that instead characterize the protein fold [158].

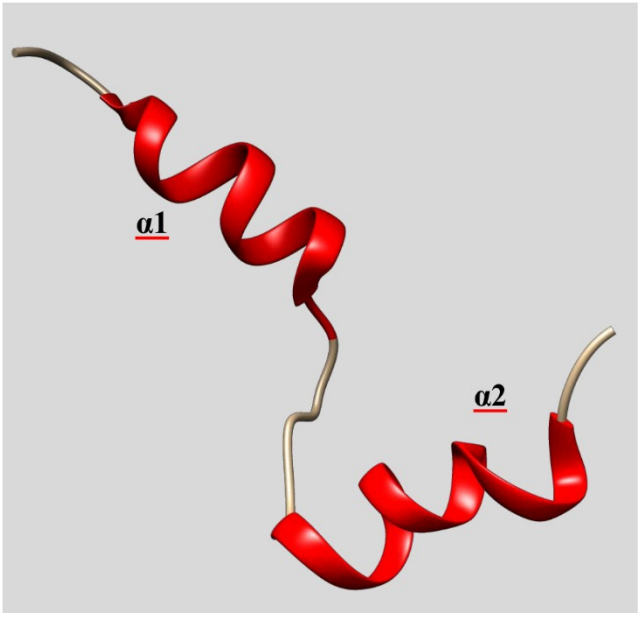


Fig. (**6**). NMR solution structure of the active domain from HSV ICP47 in presence of SDS (first conformer, PDB code: 1QLO [156]). The α-helices (α1(A4-M15) and α2 (Y22-R32)) are colored red.

Another interesting work is related to the TREX (TRanscription-EXport) complex and two of its adaptors components ALYREF (Aly/REF Export Factor) and UIF (UAP56 (U2AF65-Associated Protein 56) Interacting Factor) [166].

The TREX-NXF1 pathway is responsible for most of mRNA export into the cytoplasm where it can be translated into proteins [159]. Early proteins members of the TREX complex during transcription are deposited on the mRNA. Next, following processing (including capping and splicing) additional TREX deposition occurs. After mRNA maturation, the export receptor -known as TAP or NXF1- engages the mRNA through interactions with the TREX complex. NXF1 conducts the properly packaged mRNA through the nuclear pore [159].

To enable expression from viral intronless genes, HSV has developed a route to evade these cellular checks by

exploiting the ICP27 protein [166]. In details, one of ICP27 function consists in guiding viral mRNAs from the nucleus to the cytoplasm where it can be translated into proteins [167]. ICP27 interacts with viral RNAs and associates straight with TAP/NXF1, that is needed for the export of HSV-1 mRNAs. ICP27 also binds the export adaptor proteins Aly/REF and UIF [167]. NMR studies were conducted with three different constructs encoding for UIF N- and C-terminal regions and full length protein that were prepared and expressed in *E. coli* (UIF^{NT}, UIF^{CT} and UIF^{FL}) [166]. Inspection of 1D [^{1}H] NMR and 2D [^{1}H, ^{15}N] HSQC revealed the presence of a folded globular domain in the N-terminal region of UIF (UIF^{NT}) while the C-terminal region appeared unfolded; western blots analysis highlighted that this N-terminal region may be responsible for the binding to HSV-1 ICP27 [166]. ICP27 NMR studies revealed as well an N-terminal disordered region responsible for the interaction with RRM (RNA Recognition Motif)-domain of ALYREF [157]. To verify if the same N-terminal region of ICP27 could bind UIF, [^{13}C,^{15}N] labeled ICP27 was prepared and standard triple resonance experiments were run to get an almost complete sequence-specific backbone assignment [166]. Instead, 2D [^{1}H, ^{15}N] HSQC spectra of labeled ICP27 were acquired in absence and in presence of unlabeled UIF^{NT} thus letting identify the binding site for UIF on the ICP27 surface [166]. Indeed, all residues in the region 105-135 of ICP27 undergo broadening in 2D [^{1}H, ^{15}N] HSQC upon addition of UIF^{NT} [166]. Interestingly the region perturbed by UIF^{NT} significantly overlaps with the binding site for ALYREF (i.e., residues 103–112) [166].

Thus, NMR spectroscopy resulted crucial in identifying interaction sites that could be targeted with small molecules and peptides inhibitors thus avoiding export of viral RNA from nucleus to cytoplasm and block translation.

To support export of viral mRNAs from the nucleus to cytoplasm, ICP27 engages flexible GC rich RNA sequences through an RGG box motif [167]. Therefore, attention was given to the structural elements that drive the ICP27/RNA interaction [167]. ICP27 is made up of 512 residues and the RGG box motif is positioned within the 138 – 152 segment; NMR studies were thus conducted with the N-terminal 160 amino-acids long region. The 2D [^{1}H, ^{15}N] HSQC spectrum of ICP27 is characterized by a combination of elements (low dispersion of the resonances and high degree of peak overlaps) which indicated the presence of a random coil state, multiple conformers and flexible fold [167]. The same spectra were acquired also in presence of oligonucleotide sequences corresponding to the HSV-1 glycoprotein gC mRNA, and the results revealed that ICP27 conformational behavior was not affected by the interaction [167]. Interestingly a triple ICP27 mutant, with arginines 138, 148 and 150 replaced by lysines, was able to interact weakly to nucleotide sequences that instead associate well with the wild-type, single and double mutants. However, the [^{1}H, ^{15}N] HSQC spectrum showed that the three mutations did not change the overall unstructured protein conformation [167].

Interestingly, HSV encodes a STUbL (Small Ubiquitin-like MOdifier (SUMO)-Targeted Ubiquitin Ligase) known as ICP0 (Infected Cell Polypeptide 0) that is involved in degradation of phase-segregated nuclear protein bodies called PML NBs (ProMyelocytic Leukemia protein Nuclear Bodies), that are responsible for repression of viral genome [168]. ICP0 recognizes sumoylated proteins via one or more of its SLS (SIM (SuMO Interacting Motives)-Like-Sequences) regions, ICP0 has indeed 7 predicted SLSs. NMR interaction studies were conducted with each of the SLS and SUMO1, SUMO2. Therefore, 2D [^{1}H, ^{15}N] HSQC experiments were recorded with ^{15}N-labeled samples of both SUMO1 and SUMO2 in absence and presence of increasing amounts of ICP0 SLSs motives and the results identified SLS4 as the only region able to bind SUMO1 and SUMO2. For the interaction between SUMO2 and SLS4 by following SUMO NH chemical shift variations *vs* SLS concentration, it was possible to estimate a K_D value around 81 μM. Resonance assignments were obtained for SUMO1 and SUMO2 through a set of standard triple resonance experiments [168]. Next, CSP analysis positioned the binding site for SLS4 in the region between β2 and α1 of SUMO1 and SUMO2. Instead, 2D [^{1}H, ^{1}H] TOCSY and 2D [^{1}H, ^{1}H] NOESY spectra were employed to get assignments of the SLS4 region of ICP0; interestingly NMR data pointed out that the N- and C-terminal ends are flexible (only 12 central / 20 total residues have signals detectable in the HN region of the NMR spectra) while Hα chemical shift values indicated a propensity of SLS4 to form a β-strand [168]. The structure of SLS4, calculated through angular and distance restraints, along with the structure deposited in the PDB (Protein Data Bank) for SUMO2, and intermolecular distance restrains obtained by NMR were exploited to get a structural model of the SUMO2/SLS4 complex. The complex is stabilized by a) hydrophobic interactions between a hydrophobic shallow pocket positioned in between the strand β2 and the α1 helix in SUMO2 and the hydrophobic side chains of the SLS4 region, b) complementarities between the significant positively charged surface of SUMO2 and the negatively charged one of SLS4, and c) a salt bridges network at the N- and C-termini of the middle SLS4 region [168]. Similarly, a model of SUMO1/SLS4 complex was built based on NMR data and resulted stabilized by structural features similar to the SUMO2/SLS4 interaction. This study provides structural information essential for the design of compounds inhibitors of the interaction between ICP0 and SUMO that could work as antiviral agents by enhancing repression of viral genome.

Human USP7 (Ubiquitin Specific protease 7) represents a deubiquitinating enzyme that inhibits protein degradation through removal of polyubiquitin chains from specific substrates [169]. The USP7 C-terminal region (C-USP7) includes five UBL (UBiquitin-Like) domains that bind different substrates. Interestingly, USP7 is important for effective herpes virus infection as it is targeted by many viruses belonging to the Herpes viridae family [169]. NMR

spectroscopy was implemented to study the interaction between C-USP7 and the HSV-1 immediate-early protein ICP0 (Infected Cell Protein 0), which is involved in powerful lytic infection and virus reactivation from latency. The NMR solution structure of UBL1 domain of C-USP7 was calculated and consists of the characteristic ubiquitin-like β-grasp fold (with the β1-β2-α3-3_{10}-β3-β4-α4-α5-β5 secondary structure elements) [169]. Previous studies indicated that the ICP0 peptide 617-GPRKCARKTRH-627 encompasses the minimal region required to bind USP7, whereas, the USP7 region responsible for binding to ICP0 is located between residues 599 to 801 which includes part of UBL1, complete UBL2 and part of UBL3 domains [66]. Chemical shift perturbation experiments were conducted by running [^{1}H, ^{15}N] HSQC experiments of isolated ^{15}N-labeled UBL1, UBL12, and UBL3 domains without and with unlabeled ICP0 peptide and the results showed the involvement of residues belonging only to the first two domains of USP7 (more precisely part of UBL1 and the entire UBL2) in the interaction [169]. This study provides novel hints for structure based design of antiviral agents targeting C-USP7-ICP0 complex.

Different works were focused on glycoproteins which, as described above, are fundamentals for the binding and entry of HSV in the host cells [147].

Peptides derived from the N-terminal region of HSV-1 gD (gD-1) produce both B and T cell responses and consequently antibodies can inactivate HSV-1 *in vitro* [170].

For instance, the 9–22 region of gD-1 glycoprotein (i.e., 9-LKMADPNRFRGKDL-22) has an antigenic site assuming a β-turn conformation at residues P14–N15 [160]. Two cyclic versions of this sequence were designed and synthetized by substituting the lysine residue K11 with a cysteine (peptide C11- 9-LK[CADPNRFK]GKDL-22) or homocysteine (peptide cy11- 9-LK[HcyADPNRFK]GKDL-22), inserting chloroacetylation of the side chain of K18 and finally forming a thioether bond between the thiol group of the newly introduced residue in position 11 and the modified side chain of K18 [160]. 2D [^{1}H, ^{1}H] NOESY, 2D [^{1}H, ^{1}H] COSY and 2D [^{1}H, ^{1}H] TOCSY experiments were recorded and analyzed. The NMR solution structures of the peptides were calculated considering distance restraints collected from NOESY spectra. The results of NMR studies along with CD and FT-IR (Fourier transform-Infrared spectroscopy) data highlighted interesting differences in the structures of the two peptides C11 and cy11 [160]. In fact, the presence of an extra CH_2 group in cy11 forces the cycle to assume an altered turn structure. Interestingly, the most significant alteration of the structure occurs at the A12 and D13 positions. Next, direct and competition-type interaction assays were conducted between peptides and a gD-13 specific monoclonal antibody (A16) revealing that the peptide cy11 exhibited higher affinity to Mab A16 than peptide C11, even if the two peptides appeared provided with decreased reactivity with respect to the linear analogues [160]. This study pointed out how even small variations in a structure may induce conformational changes that can highly influence the interaction between a synthetic antigen and antibodies.

NMR studies related to the 276-SALLEDPVG-284 peptide region of the gD glycoprotein were as well conducted [170]. Immunological characterization indicated that the DPVG sequence represents the “core epitope” for antibody recognition. Different cyclic analogues based on the 276-284 linear peptide fragment were designed and synthetized and their structural preferences investigated through 2D [^{1}H, ^{1}H] TOCSY and NOESY spectra along with 2D [^{1}H,^{13}C] HSQC spectra [170]. NMR data showed that cyclic peptides in which the core -DPVG- motif was included in the cycle presented a β-turn structure provided with higher stability with respect to linear peptides or cyclic peptides with the core motif outside of the cyclic arrangement [170].

The latter examples show further how NMR spectroscopy is a really powerful instrument to study in detail the conformational properties of small peptides.

Interaction of cytoplasmic capsids with cellular organelles and consequent envelopment represent key events in the HSV life cycle. Establishment of a complex set of interactions involving the capsid surface, tegument members, and cytoplasmic tails in viral glycoproteins is at the bases of these events [171]. One of such interactions involves the tegument VP16 proteins and the cytoplasmic tail of the envelope gH glycoprotein and was characterize through NMR spectroscopy [171]. NMR studies were as well conducted at different temperatures (in the temperature range 2-45 ˚C) as this parameter influences the interaction between the two proteins. In detail conformational studies were carried out for the cytoplasmic tail of wild-type gH glycoprotein (i.e., KVLRTSVPFFWRRE) and its mutant version in which the proline residue P8 was substituted by an alanine (i.e., KVLRTSVAFFWRRE). 1D [^{1}H] and 2D [^{1}H, ^{1}H] NMR spectra were recorded to get insights on the structural features characterizing the two peptides, based on distance restraints from NOEs and on angular constraints from $^{3}J_{HNH\alpha}$ coupling constants, it was possible to calculate a 3D structure for the wild-type peptide at 10 ˚C but not at 37 ˚C [171]. In fact, at 37 ˚C the peptide is completely flexible and disordered whereas at 10 ˚C a few medium range NOEs could be detected in the region 5-10; this segment includes the proline residue which favors a tight turn-like structuration in the peptide. Based on these evidences, it was suggested that this small amount of stable structure avoids the interaction with VP16 at temperatures lower than 37 ˚C [171]. This hypothesis is strengthened by the evidence that the mutated gH peptide resulted to be unstructured at all the analyzed temperatures and always able to interact with VP16. These data let speculate an induced fit mechanism according to which the peptide needs to lack any preformed structuration to be able to bind VP16 [171]. This disorder in the gH tail may favor interactions with many different tegument partners through diverse induced conformations

thus efficiently ensuring redundancy during the particle assembly process.

Many antiviral drugs attack selectively HSV-1 TK (Thymidine kinase). Indeed many antiviral agents -like acyclovir- upon infection of cells by Herpes Viruses (HV), are transformed into mono-phosphates nearly entirely by HV TK and subsequently, into the triphosphates through action of cellular enzymes, thus producing active metabolites that work as "fake" nucleotides and are included in embryonic DNA strains inducing inhibition of HV DNA polymerase [172]. TK canonical interactors are usually flexible and gaining information on their binding site is challenging, in this case NMR spectroscopy may be particularly advantageous to get the bound ligand conformation through transferred NOE technique [38]. Such studies were not too successful for acyclovir for which a few negative NOEs, indicating exchange between bound and free states could be indeed detected, but the data were not sufficient to achieve a precise structure characterization. More insights could be retrieved thanks to a more favorable dissociation constant value that enhanced the appearance of tr-NOEs, with a tricyclic acyclovir derivative (i.e., 3,9-dihydro-3-[(2-hydroxyethoxy)methyl]-6-ethyl-9-oxo-5H-imidazo[1,2-α]purine) [172]. This compound was analyzed in the free state and bound to HSV1 TK. Transferred NOE experiments provided distance restraints for structure calculations. Obtained structures pointed out that in both free and bound forms the syn orientation of the chains respect to the aromatic rings is favored however, while in the unbound state the acyclic group appears rather disordered, upon interaction with TK only a single conformation is favored. This NMR derived information can be exploited in the structure based design of novel antiviral compounds targeting HSV TK [172].

4.2.2. Metabolomics and plant extracts analyses

NMR is the ideal technique to study complex mixtures including cellular and plant extracts. The combined use of NMR-based metabolomics studies and analyses of components in plant extracts through NMR spectroscopy resulted a valid tool to identify novel molecules with anti-HSV activity.

It is known that HSV infection is associated with metabolic changes; as HSV-1 induces a lifelong latent infection in neurons and glial cells, studies were conducted to monitor extracellular metabonomic changes caused by HSV-1 infection in human OLigodendrocytes (OL) [173]. Oligodendrocytes represent a type of glia cells and are one of the main element in the CNS (Central Nervous System) white matter [173]. An NMR-based metabonomic investigation was conducted to compare extracellular metabonomic profiles of HSV-1 infected human OLs and control cells at different time lapses (i.e., 12, 24, and 36 h following infection) [173]. This study pointed out alterations in twenty-one extracellular metabolites and revealed that the entire metabolic pattern permitted to distinguish between healthy and infected cells. In major details, the analyses carried out 36 h after-infection showed that eight extracellular metabolites, including seven amino acids (i.e., alanine, glycine, isoleucine, leucine, glutamate, glutamine, histidine, and lactate), underwent significant variation in HSV-1 targeted human oligodendrocytes with respect to control cells. These *in vitro* studies on human cultured oligodendrocytes let speculate that HSV-1 interfere with amino acid metabolism of the host cells to enhance viral replication [173].

Several studies indicated that plants represent a promising source of potential antiviral molecules against HSV infection.

Native Americans employed several plants from the boreal forest of Quebec for their beneficial effects against microbial infections [165]. A study deeply analyzed the anti-HSV properties of extracts from one of this plant: *Cornus canadensis L.* (Cornaceae). Plants extracts analogue to those implemented by Native Americans were prepared by decoction and infusion with water, ethanol and water/ethanol 1:1 mixture [165]. Different extracts were assayed for anti-HSV-1 properties through a plaque reduction test; next, fractionation procedures were implemented based on the bioassay results, to isolate bioactive ingredients. Deep analyses of 1D and 2D NMR spectra (1D ^{1}H and 2D COSY, HSQC and HMBC (Heteronuclear MultiBond Correlation)) led to clarify the chemical structures of active compounds [165]. Interestingly, hydrolysable tannins were found in the fractions with the highest activity against HSV-1 [165].

The red alga *Osmundaria obtusiloba* from Southeastern Brazilian coast was identified as source of glycolipids characterized by anti-HSV-1 and anti-HSV-2 activities [164]. Glycolipids fractions were isolated from the alga and NMR spectroscopy and mass spectrometry were exploited to elucidate their chemical structures and to better comprehend the type of functional groups responsible for the observed cytotoxicity and antiviral activity in *in vitro* assays [164]. For instance, analysis of different NMR spectra including 2D [^{1}H, ^{1}H] TOCSY and 2D [^{1}H, ^{13}C] HSQC showed pattern of signals characteristics of a glycerol moiety, an α-quinovopyranosyl group and 6-sulfo-α-quinovopyranosyl unit and thus indicated the presence of SQDG (1,2-di-*O*-acyl-3-*O*-(6-deoxy-6-sulfo-α-D-glucopyranosyl)-*sn*-glycerol) in the analyzed glycolipids extract [164]. Interestingly, SQDG was associated with potent anti-HSV activity [164].

The officinal plant *Houttuynia cordata* is well known in China for its ability to relieve fever, to reduce swelling, to drain pus, and to promote urination [174]. *H. cordata* has been employed in the treatment of different viruses such as HSV-1, influenza virus, HIV-1. As in the previous examples, NMR techniques were employed to elucidate the chemical structures of molecules contained in the plant extract. Five different Houttuynoids representing original flavonoids with houttuynin (3-xodecanal) chained to hyperoside were isolated from the plant and their structures clarified by NMR

[174]. These compounds demonstrated good HSV inhibitory properties that in a few cases resulted even better than acyclovir, that was used as a positive control [174].

Another work exploited a metabolomics-like approach to discover common compound groups or molecules derived from unrelated plants, but for which analogous biological activities have been reported [175]. This study investigated in details 67 plants active against diverse viruses including HSV, Cytomegalovirus and HIV along with a few plants, with no know anti-viral activity that were used as control. 1D ^{1}H NMR and LC (Liquid Chromatography)-MS (Mass Spectrometry) techniques were implemented to group the molecules with analogue anti-viral activity on the basis of their chemical similarity. Interestingly, the results showed the recurring presence of chlorogenic acids and its substructures within plants with anti-HSV activity and thus, suggested their use as valuable starting points for the development of antiviral agents against HSV [175].

4.2.3. Other miscellaneous studies: NMR characterization of antiviral agents based on peptide, small molecule or polymeric scaffolds

One of the strategies to obtain molecules with antiviral activities consists in the chemical modification ("optimization") of a starting peptide which reproduces protein sequences somehow linked to viral infection.

The HSV RR (Ribonucleotide Reductase) is important for viral infectivity, a peptide from the C-terminal region of HSV RR small subunit (i.e., YAGAVVNDL) is able to block the association of the two RR subunits and consequently the enzymatic activity (IC50 equal to 38 μM) [161]. A shorter peptide version (VVNDL) was used as starting point to get promising antiviral agents (Fig. **7**) [161].

Different chemical modifications were inserted including addition of a diethylacetyl group at the N-terminal side, introduction of a methyl group on the β-carbon of the second valine, dialkylation of the nitrogen of the asparagine side-chain, dimethylation of the β-carbon of aspartic acid (Fig. **7**) [161]. Indeed, these modifications led to a peptido-mimetic with a 4000 times increased potency compared to the starting VVNDL sequence [161]. In a similar work, more potent inhibitors were further obtained by introducing alkyl-ureido functionalities and thus an additional nitrogen at the N-terminal end of the starting VVNDL peptide [176]. NMR spectroscopy was employed within these studies to verify the correct chemical structure of each peptide and thus to prove the success of chemical reactions and as well as revealing the bioactive conformations.

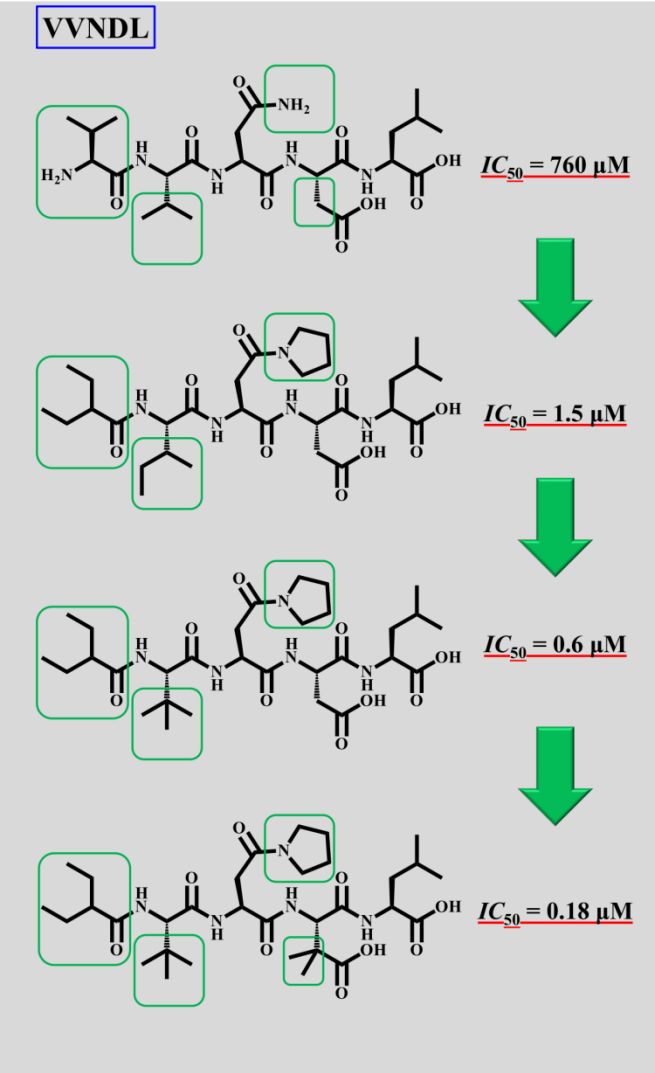


Fig. (**7**). Design of peptide inhibitors of HSV RR starting from the VVNDL peptide [161]. Green rectangles are used to highlight the modifications leading to enhanced activity.

Several structurally distinct inhibitors of herpetic DNA polymerases are employed in herpesvirus therapy and among them there are acyclic nucleoside analogues such as acyclovir [177]. Acyclovir is known to block HIV-1 RT (Reverse Transcriptase) and also to have some efficacy in HIV-infected patients that are co-infected with herpesviruses. α-CNPs (α-CarboxyNucleoside Phosphonates) represent also a group of DNA polymerase inhibitors whose chemical structure is made up of a nucleobase joint to an α-carboxyphosphonate through a proper linker such as a cyclopentyl ring, such molecules are also competitive blockers of HIV-1 RT [177]. A clever synthetic route was set up to substitute the cyclopentyl unit in the model α-CNPs by a more flexible group obtaining interestingly a relevant (~100-fold) selectivity switch towards herpetic DNA polymerases with respect to HIV-1 RT [177]. Therefore, this strategy highlighted the possibility to get selective inhibitors of a specific virus starting from

compounds more active against different pathogens. As underlined in previous paragraphs, NMR is important -as analytical technique- within this and similar studies to check the progression of a chemical reaction and verify compounds identity. In the above described work ^{1}H, ^{13}C and ^{31}P NMR spectra were acquired to characterize the synthetized compounds [177].

Chemical modification of nucleosides is considered a prolific way to obtain antiviral agents. An interesting work revealed that L-thymidine and D-thymidine (where the D-enantiomer is the natural TK substrate) possess the same ability to inhibit viral growth as they both bind and are phosphorylated by HSV1-TK [163]. In addition L-thymidine did not inhibit the human enzyme up to 1 mM concentration whereas, *in vivo* is able to block viral proliferation; interestingly, no toxicity on non-infected cells due to L-thymidine could be detected letting speculate that this enantiomer can be employed as starting point to generate novel classes of non-toxic anti HSV compounds [163].

Nevertheless, considering the bioisosterism between oxygen and selenium several seleno-purine nucleosides were designed and synthesized [162]. ^{13}C NMR was useful to distinguish different compounds isomeric forms during chemical reactions while the final reaction products were characterized by 1D [^{1}H] and ^{13}C NMR spectra [162]. A significant activity against HSV-1 and HSV-2 was detected for a seleno-acyclovir derivative accompanied by little cytotoxicity up to 100 μM concentration [162].

Instead, octAU (6-(4-Octylanilino)uracil) was described as a molecule able to block the HSV-1 UDG (Uracil-DNA Glycosylase), an enzyme involved in DNA repair and linked to exit of the virus from latency [178]. OctAU does not interfere with human UDG and a structural model of its interaction with HSV-1 UDG could be obtained through docking studies. Based on this model several analogues were designed in an attempt to improve enzyme binding affinity and physicochemical characteristics including water solubility [178]. The search of selective HSV-1,2 TK (Thymidine Kinases) inhibitors led to the synthesis of a series of N2-phenylguanine derivatives and to the discovery of HBPG (9-(4-hydroxybutyl)-N2-phenylguanine). HBPG potently behaves as a competitive inhibitor of HSV-1 TK with respect to the substrate thymidine without being itself a substrate of the enzyme [179].

Expression of HSV1-tk gene is fundamental for the activation of a prodrug through its gene product (i.e., HSV1-TK). Several HSV1-TK PET (Positron Emission Tomography) reporter probes have thus been designed based on the structures of compounds with activity against HSV infection [180]. Among them, 5-[^{18}F]fluoropropyl, [^{18}F]fluorobutyl and [^{18}F]fluoropentyl pyrimidine nucleosides were described as promising *in vivo* PET reporter probes to monitor HSV1-TK reporter gene expression [180].

It's worth noting that non-nucleoside HSV-1 inhibitors have been produced as well including those based on substituted 1,3-dihydroxyacridones [181].

The disadvantages associated with the commonly used anti-herpes drugs (e.g., acyclovir) include low bioavailability and severe side effects [182]. Cationic dextran derivatives (DEXxDSy) were proposed to overcome these issues [182]. Therefore, various DEXxDSy with a variety of MWs (Molecules Weights) and DS (Degrees of Substitution) with ammonium groups were synthesized and characterized. The results showed that higher MWs and higher degrees of substitutions were associated with higher anti-HSV activity [182]. More in detail, only polymers with MW at least equal to 100 kDa and DS values higher than about 40% were needed to achieve good antiviral potency however, higher DSs were associated as well with larger cytotoxicity. Furthermore, this study suggested a mechanism of action to explain the limited virus transmission induced by DEXxDSy in which the virus attachment and/or entry is hampered by the interaction of DEX100DS40 with HS (Heparan Sulfate) [182]. As mentioned before, both vertebrate and invertebrate species contain HS molecules that are universally expressed on the surfaces and in the extracellular matrix of nearly all types of cells functioning as receptor for different viruses including HSV [183]. Indeed, to block viral infections polyglycerol-based sulfated nanogels with sizes comprise between 100 and 200 nm and with various degrees of flexibility were implemented as mimetics of cellular HS [184]. These nanogels provided with defined flexibility possess an advantageous large contact area [184]. This feature allows a better hampering of the interaction between viral particles and host cell surfaces, which in turn lead to the inhibition of virus attachment to cells and subsequent uptake of viral particles. In addition, these nanogels presented poor cytotoxicity [184].

Other polymeric inhibitors of HSV have been reported such as the PEG_x-*b*-$PMAPTAC_y$ (Poly(Ethylene-Glycol)-*Block*-Poly(3-(Methacryloylamino)Propyl-Trimethylammonium Chloride)) copolymers, that can be produced via a simple and cost-effective synthetic route leading to water-soluble, well-defined polymers with required compositions, molecular weights, and low dispersity indices [185]. Interestingly, this co-polymer works as inhibitor of HSV-1 infection in *in vitro* and *in vivo* assays. In details it provides protection from lethal HSV infection in a mouse model and is characterized by absence of systemic or topical toxicity *in vivo*, even at considerably high concentrations [185]. The mechanism proposed to justify the co-polymer antiviral action is based on its interaction with the cell with a subsequent inhibition of the viral entry/fusion machinery although no binding to any of the principal HSV-1 receptors could be revealed [185].

5. HUMAN IMMUNODEFICIENCY VIRUS (HIV)

5.1. Mechanism of action

HIV belongs to the family of Retroviridae and possesses a ssRNA(+) genome that is reversely transcribed to DNA during the virus life course [6, 186]. The known lethal disease associated with HIV infection is AIDS (Acquired Immune Deficiency Syndrome) and consists in the progressive weakening of the immune system; since its outbreak in the early 80's AIDs has caused over 32 million of deaths [187-189]. Indeed several pathologies, like acute kidney injury, arterial stiffness, cardiovascular diseases, tuberculosis, intestinal parasite carriage and disruptive sleep apnea, can affect patients infected by HIV [189].

There are two major HIV types: HIV-1 and 2; HIV-1 is the most common one and is present all over the world whereas, HIV-2 occurs prevalently in West Africa. HIV-1 and HIV-2 are both retroviruses but they are different from a genetic point of view as their genomes possess only 55 % sequence identity [190]. HIV-1 and HIV-2 have the same basic gene organization, transmission routes, intracellular replication mechanisms and clinical outcomes as both bring to AIDS. However, a lower transmissibility and reduced risk of develop AIDS are hallmark of HIV-2 [191].

HIV differs from other retroviruses because its infection can reach an efficient and productive level only through expression of several proteins besides Gag, Pol and env. These proteins are the regulatory ones (i.e., TAT (Transcriptional AcTivator) and REV (Regulator of Expression of Viral proteins)), that manage viral gene regulation, and the accessory proteins (i.e., Vif (Virion infectivity factor), Vpr (Viral protein r), Nef (Negative factor), and Vpu (Viral protein u)), that are versatile adaptors acting in the host cells and linking viral and cellular pathways thus favoring infection and immune evasion [186, 187].

The mature HIV particle ("virion") is characterized by an external envelope that consists in a lipid bilayer coming from the host cell membrane and composed by different proteins [187]. The gp120 (glycoprotein 120) is an envelope glycoprotein of 120 kDa that is exposed on the viral surface and is engaged to the virus through interactions with the gp41 (glycoprotein 41), a transmembrane glycoprotein of 41 kDa. The envelope includes as well many other cellular membrane proteins like major histocompatibility antigens, actin and ubiquitin [187]. In addition, two thousand copies of the matrix protein p17 form the inner surface of the viral membrane whereas, a cone-shaped capsid core particle is composed by two thousand copies of capsid protein p24 and is positioned at the center of the virus [187]. The capsid includes two copies of the unspliced viral genome that is stabilized as a ribonucleoprotein complex made up by almost two thousand copies of nucleocapsid protein p7 and also includes three virally encoded enzymes (i.e., protease, reverse transcriptase and integrase) which play fundamental functions such as reverse transcription of viral RNA to viral DNA and incorporation of the viral DNA into the chromosomal DNA of the host cell [187]. In addition the proteins Nef, Vif and Vpr are packaged inside the virus particles [187].

The most canonical route of HIV transmission is the sexual diffusion through contact with cell-free or cell-associated infectious viruses in semen or mucosal surfaces [192]. Different secondary ways of transmission have also been identified (e.g., injection drug use, exposure of blood and blood products via transfusions, and exposure of fetus or infant to HIV from an infected mother) [192]. From a general point of view, three different stages characterize HIV infection: acute-, chronic-phases and AIDS [193]. An acute HIV syndrome takes place approximately between two and four weeks from the transmission of the virus and at this stage infected people may experience flu-like clinical symptoms linked to great plasma viremia and frequently fever and lymphadenopathy [192]. In the course of acute HIV phase deep reduction of $CD4^+$ T cells happens mainly in the GALT (Gut-Associated Lymphoid Tissue). At this stage, HIV reservoirs, including latently infected, quiescent $CD4^+$ T cells are generated. During the acute HIV infection phase, the virus replicates rapidly and fast destroys the CD4 cells of the immune system that are trying to fight against the infection [193]. The chronic stage is also known as clinical latency or asymptomatic HIV infection. Patients affected by chronic HIV infection may not show HIV-like symptoms. During this stage, HIV replication still occurs in the body but, at rather low levels. HIV replication takes place in all secondary lymphoid tissues, causing an indiscriminate immune activation, constant viral production, enhanced cell turnover, and in the end devastation of the host immune system and fast disease evolution [192]. Completion of this step usually takes at least ten years [193]. In the final and more aggressive phase, the $CD4^+$ T cells of immune system reach levels that are no longer sufficient to counteract the viral infection and thus AIDS occurs [192].

The life cycle of HIV is characterized by different steps and each of them gathers attention for its potential role as target for the development of chemotherapeutic routes against HIV [194]. In the first step of HIV entry into $CD4^+$ T cells three main events occur: the interaction between the gp120 protein and CD4 cell receptor, a consequent conformational transition in gp120 that favors the binding to other receptors on the cell and a conformational change in gp41 that allows the fusion of the viral envelope with the host cell membrane [194, 195]. The following step is the "decapsidation" (also named "uncoating") through which the viral RNA(+) genome is unconstrained from the capsid, released into the host cell where it is next transformed into proviral "ds DNA" by the reverse transcriptase associated with the virus [194]. Then, the HIV integrase induces the insertion of the proviral "ds DNA" into the host cell genome and the expression of viral genes occurs by exploiting the transcription and translation machineries of the host cell while being additionally modulated by viral regulatory proteins (e.g., Tat and Rev) [194]. Novel viral RNA(+) genome and proteins are moved towards the host cell

membrane where they are assembled into immature and non-infectious HIV particles [194]. The newly formed virions exit from the $CD4^+$ T cells through the budding process and undergo a maturation process in which viral proteases cleave viral precursor proteins (i.e., pp55 gag and pp160 gag-pol) and transform them in structural (gag) proteins (i.e., p17, p24, p7) and functional (pol) proteins (i.e., protease 11/p11, reverse transcriptase p66/p51 and integrase p32 tetramer) [194]. This step provides HIV virions able to infect a new $CD4^+$ T cell and thus iterate viral life cycle [194].

More than 30 years have been already spent in the development of a cure against HIV [196]. Most of the available therapeutic routes shares the idea that HIV hides within latent reservoirs whose activation leads to the destruction of reactivated cells by the action of immune system or by the activation-induced cell death associated with viral production in active $CD4^+$ T cells [196]. This approach is known as "kick and kill". In detail, "kick" agents against HIV are those that trigger the reactivation of latent HIV reservoirs [196]. A few examples of "kick" agents are provided by HMBA (HexaMethylene BisAcetamide) and Disulfram [196]. These agents annul latency by inducing the release of P-TEFb (Positive Transcription Elongation Factor b) that in turn phosphorylates diverse transcriptional modulators at the HIV promoter locus thus favoring the beginning of transcription and elongation [196]. In addition, other "kick" agents include histone deacetylase inhibitors (e.g., valproic acid, vorinostat, panobinostat, romidepsin, suberoyl bis-hydroxamic acid), histone methyl transferase inhibitors (e.g., chaetocin and BIX-01294) and the DNA cytosine demethylation agent Aza-CdR (5-aza-2-deoxycytidine) [196]. Instead, "kill" agents working as killers of reactivated cells include broadly neutralizing mAb (Monoclonal Antibodies), many of which target the HIV envelope glycoproteins gp120 coupled to gp41 and can inhibit viral entry into cells although, these agents through ADCC (Antibody-Dependent Cell-mediated Cytotoxicity) could also induce ruin of the latent reservoir. Other "kill agents" include a monoclonal antibody against the integrin α4β7, bispecific antibodies (e.g., BiTEs (Bispecific T-cell Engagers) and DARTS (Dual-Affinity Retargeting Molecules)), HIV specific T-cells engineered with CARs (Chimeric Antigen Receptors), SMAC (Second Mitochondria-derived Activator of Caspases) mimetics (e.g., birinapant, GDC-0152 and emblin), immune checkpoint antibodies (like antibodies for PD-L1 (Programmed cell Death receptor 1 Ligand)) and therapeutic vaccination [196].

Interestingly, the use of the single strategy alone (i.e., "kick" type or "kill" type) is not fully efficient in the treatment of patients affected by HIV [196]. Therefore, a combination of the two strategies ("kick and kill) increases the likelihood of getting a complete eradication of HIV infection [196].

5.2. NMR investigations

NMR resulted a valuable technique to get different structural information on several HIV related proteins (e.g., gp120, gp41, protease, reverse transcriptase, Vpu (Virus Protein U), Vpr, Vif and Tat) [197-207]. Structural features of viral RNA [208-213] as well as those of compounds with promising anti-HIV activity, in the free form and bound to protein targets, were also revealed by NMR [214, 215]**.** Nevertheless, NMR-based metabolomics studies of biofluids from HIV infected patients were described in literature too [128, 216].

5.2.1. Structural and Interaction studies

Several NMR structural studies dealing with the gp120 and gp41 subunits of HIV envelope glycoproteins have been conducted. The gp160 precursor is cleaved by furin-similar proteases to produce gp120 and gp41 that, as mentioned in the previous paragraph, represents the surface and transmembrane units that are responsible for receptor binding and fusion of virus with host cell, respectively. HIV envelope glycoprotein generates at the surface of the infectious virus trimeric arrangements made up of heterodimers that are non-covalently linked [197].

Peptides encompassing the MPER (Membrane-Proximal External Region) and the TMD (TransMembrane Domain) portions of HIV-1 gp41 were studied by solution NMR techniques as these regions play a role in the envelope-glycoprotein mediated fusion and in the regulation of the immune system answer to viral infection [197]. In detail, structural studies were centered on two peptides CpreTM, and TMDp provided with partially overlapping sequences.

The first peptide encompasses residues 671-693 representing the border between MPER and TMD regions, starting from the hinge portion and ending after the bend located into membrane whereas, TMDp includes the entire TMD region of gp41 (residues 684-704), residue K683 has been proposed as the starting site of the transmembrane region at the membrane surface. Interestingly, the border region between MPER and TMD represents the target of broadly neutralizing antibodies. Only a precise structural arrangement allows recognition of peptide epitopes by antibodies [197]. A set of 2D NMR experiments (COSY, TOCSY, NOESY and [^{1}H,^{13}C] HSQC) were recorded under structuring conditions in membrane-like environments produced by a mixture composed of water and HFIP (HexaFluoro-2-Propanol) and by DPC (DodecylPhosphoCholine) [197]. NMR structures were calculated for both peptides starting from distance constraints obtained by NOESY experiments and ϕ and ψ dihedral angle restraints obtained with the TALOS program [217] on the basis of ^{1}Hα and ^{13}Cα chemical shifts. The results revealed that CpreTM and TMDp are highly α-helical; CpreTM possesses an α-helical organization covering its entire sequence; as concerning TMDp a certain conformational flexibility around the transmembrane sequence 690-GGLV-693 was revealed [197]. In proximity

of K683 no meaningful changes in the orientation of the backbone of the helix could be detected. Thus, NMR data indicated that the HIV-1 Env(671–704) fragment is composed by two helical segments separated by a flexible membrane-inserted hinge. These data coupled with studies on antibody recognition and rabbit immunizations let delineate the following model: the N-terminal helix plays a major role in membrane fusion and has an immunogenic character while, the C-terminal helix works in an immune-suppressive way and intervenes in the final phases of the fusion process [197]. This is an example of how NMR can provide structural data useful for peptide-based vaccine design.

In another work, the interest moved towards the 12p1 peptide (1-RINNIPWSEAMM-12) representing an inhibitor of viral entry able to bind the gp120 protein [198]. Indeed, combined analysis of TOCSY, ROESY, NOESY, ^{13}C-edited HSQC, and ^{15}N-edited HSQC NMR spectra led to the assignments of almost all ^{1}H, ^{13}C, and ^{15}N resonances for the peptide in its free form. In addition inspection of ΔHα (Hα chemical shift deviations from random coil values) and ΔCα (Cα chemical shift deviations from random coil values) pointed out the absence in the free peptide of canonical secondary structures but, NMR data revealed a trans configuration for the proline P6. This proline is interestngly positioned in the region of the peptide N3-S8 which is more important for its interaction with gp120 [198]. Nevertheless, STD NMR experiments were conducted for precise peptide epitope mapping against the monomeric and trimeric forms of gp120 (i.e., in absence and presence of gp41, respectivelly). Such studies allowed identifying the atoms of the peptides in close contacts with gp120 in both states. For example STD NMR highlighted that the side chain of tryptophan in 12p1 was significantly involved in its binding to gp120. Interestingly, STD NMR experiments revealed as well small differences in the interaction of the peptide with gp120 in absence and presence of gp41 thus reflecting low variations in the structure of gp120 upon binding to gp41. The interaction between 12p1 and gp120 appeared dependent on a combination of hydrophobic and electrostatic contacts [198]. This example stresses out the importance of NMR epitope mapping to identify residues crucial for inhibition of viral entry and thus design eventually more potent antiviral agents blocking interaction between gp120 and host cell receptor.

NMR spectroscopy was exploited to get structural features of a region of the precursor gp160 glycoprotein involved in its selective proteolytic activation by the enzyme furin and other PCs (Precursor Convertases) [218]. The selective proteolytic activation of the HIV-1 envelope glycoprotein gp160 takes place at the C-terminal side of the sequence 508-REKR-511 (site 1), although the occurrence in the protein of another close similar sequence 500-KAKR-503 (site 2). This evidence lets speculate that specific structural features may play a major role in directing the enzymes to cleave just site 1. With this in mind NMR studies were conducted on a synthetic 19 residue long peptide encompassing both site 1 and site 2 (i.e., p498, sequence: 1-PTKAKRRVVQREKRAVGIG-19) and able to be recognized and correctly processed by the enzyme furin at site 1. Therefore, 2D [^{1}H, ^{1}H] NMR spectra (i.e., TOCSY, DQF-COSY and NOESY) were recorded under structuring conditions (TFE/H_2O 90/10 mixture) and deeply analyzed to unveil the tendency of different peptide regions to assume certain secondary structure elements. NMR structure calculations revealed the presence of a 3_{10} helix encompassing the 3-KAK-5 segment, an α-helix covering the 6-RRVVQR-11 fragment and the presence of a loop in the C-terminal 12-EKRAVGIG-19 portion of the peptide, containing processing site 1 [218]. Based on this NMR data molecular modeling and docking studies were conducted to build a structural model of the furin/peptide complex and get insights in the recognition process [218]. The data showed that the C-terminal peptide segment 11-REKR-14 fitted well into the catalytic site of furin. The correct processing at site 1 occurs because the corresponding peptide region is exposed in a loop whereas, the site 2 being constrained in a helical arrangement, is not easily accessible to proteases, unless a more drastic conformational change takes place [218]. This work provides important information to design peptide inhibitors of gp160 processing targeting proteases and possessing antiviral potential by avoiding formation of mature envelop glycoproteins gp120 and gp41.

Different studies focused on the regulatory Vpr protein [186, 187, 200, 208]. Vpr by means of interactions with the C-terminal part of Gag is enclosed in the virus particle; it plays several functions like enhancing viral replication by blocking cell division at the G2 stage, and participating to the nuclear transport of proviral DNA [200]. These roles of Vpr, as well as its binding to NCp7 nucleocapsid viral protein, were described as dependent on the 40 residues long C-terminal domain of Vpr [200]. Therefore, a peptide including this domain (i.e., 52-96 Vpr region) was synthetized and investigated by different solution NMR experiments, that were conducted in presence of 30% TFE, to get structural information [200]. The results showed a 53-78 region with a rather well-defined α-helix organization, and a more flexible C-terminal segment (i.e., residues 79-86) [200]. In addition, the overall helical arrangement possesses an amphipathic character with one hydrophobic face formed by V57, L60, I61, I63, L64, L67, L68, I70 and I74 and another face composed instead almost entirely by polar residues (i.e., lysine and arginines). This structural information along with mutagenesis studies suggested the possible involvement of L60 and L67, located in a leucine zipper like motif, into Vpr dimerization through hydrophobic interactions and allowed to generate a model of coiled coil Vpr dimer. The 52-96 region of Vpr but not the 1-51 is able to interact in vitro with HIV-1 RNA. Although *in vitro* experiments demonstrated that this interaction in not dependent on Vpr dimerization, the structural data let speculate the possible relevance of the leucine-zipper like

motif for the *in vivo* interaction [200]. This work further highlights the importance of NMR even when working on synthetic unlabeled peptides (as in the case of the above described Vpr fragments) to get detailed structural details useful to explain protein functions.

However, more detailed structural studies were later conducted using the entire 1-96 Vpr protein that was synthetized with 22 labeled amino acids [201]. Different 2D and 3D NMR experiments were recorded in water at low pH and in presence of acetonitrile (30%, v/v) that was employed to avoid the self-association of Vpr protein and minimize precipitation [201]. The NMR solution structure is made up of three α-helices (i.e., α1 from D17 to H33, α2 from W38 to Y50 and α3 from T55 to R77) with flexible N- and C-terminal sides (residues from M1 to E13 and from H78 to S96, respectively) [201]. Helices are amphipathic with a hydrophobic and a polar faces and connected by loops. Hydrophobic interactions occur between α1 residues L20, L23, L26, A30 and V31, α2 residues W38, L39, L42 and I46 and α3 amino acids V57, L60, I61, L64, L68 and F72 (Fig. **8A**). In addition, the mutual orientations of these α-helices is maintained in place by contacts within one set of amino acids (i.e., T19, L20 and W54) located on one face of Vpr and a second set of residues (i.e., H33, F34, H71, F72) positioned on the other face of the protein. The structure of Vpr contains as well two hydrophobic domains that stay accessible to solvent (one composed by W18, L22 and L26 and the other by I63, L67, I70 and I74) and polar accessible regions formed by amino acids from α1 (i.e., D17, E21, E24, E25, N28 and E29), residues from α3 (i.e., Q65, Q66, R62, R73 and R77) and residues from the C-terminal extended region (i.e., R80, R87, R88, R90 and K95) [201].

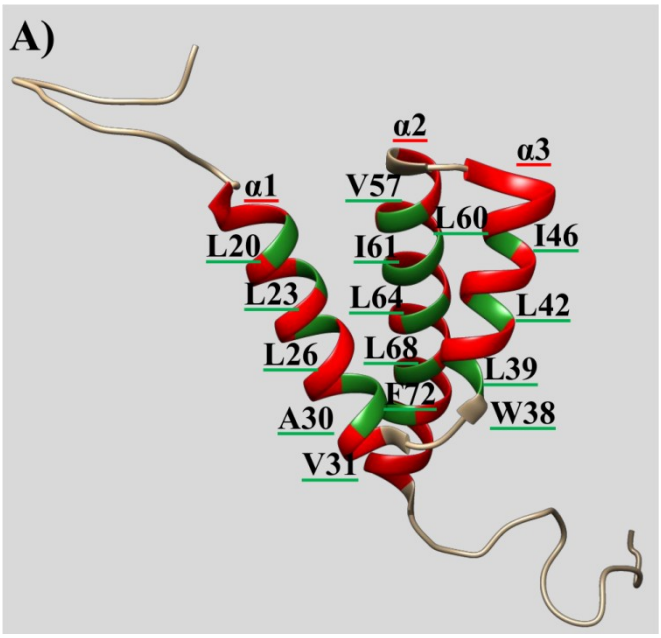


Fig. (**8**). Structural details of a few HIV-1 proteins. (**A**) NMR solution structure of Vpr (first conformer, PDB code: 1M8L [201]). The three α-helices (i.e., α1 (D17-H33), α2 (W38-Y50) and α3 (T55-R77)) are reported in red. Green is used to indicate the residues involved in the hydrophobic interactions which stabilize the structure (i.e., L20, L23, L26, A30 and V31, W38, L39, L42, I46, V57, L60, I61, L64, L68 and F72).

It's worth noting that Vpr has some cell penetrating ability besides being able to bind nucleic acids. Therefore, there has been a certain interest in the analysis of DNA transfection properties of different Vpr fragments. Efficient delivery of plasmid DNA into several cell lines can be achieved by fragments of the C-terminal helical domain of Vpr [202]. As the amphipathic character of an helix may be crucial in determining binding to membranes, it was investigated if introduction of an helix breaker Proline residue could somehow influence the cell delivery properties and interestingly it was revealed that Vpr55-82 peptide with a proline residue in position 70 instead of an isoleucine works better in gene delivery assays than the wild-type peptide. Different techniques including solid-state NMR were thus employed to study the secondary structure and interaction properties of Vpr and several proline containing mutants with model membranes made up by POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine). Solid state NMR was useful in revealing different orientations with respect to the membrane of the helical axes in regions upstream and downstream of the proline [202]. Interestingly the Vpr55-82 and Vpr55-82 Pro70 peptides assume similar structural topologies in model membranes with an alignment of the helical domain, that is positioned N-terminally of Pro70, close to completely in-planar with the membrane while, a certain conformational heterogeneity was revealed for the helical segment located C-terminally respect to proline. NMR data clearly evidenced that the improved transfection properties of the proline mutant could not be linked to divergences in membrane interactions of either the N-terminal or the C-terminal segments [202].

Solid state NMR studies were also conducted for the HIV-1 Vpu protein. Vpu is composed of a transmembrane helical domain and a cytoplasmic domain with amphipathic character made up of two additional helices [203]. Vpu interacts with different cellular protein, including CD4 (Cluster of Differentiation 4), tetherin, NTB-A (Natural Killer, T, and B cell-Antigen), and CCR7 (C-C Chemokine Receptor type 7) thus favoring the release of new virus particles and messing up the immune system. Solid state NMR studies indicated the presence of three helices in Vpu full length protein with two helices in the cytoplasmic domain assuming an U shape. Diverse lipid composition produces different length in the interhelical loop of the cytoplasmic domain and influences the orientation of the third helix. Overall NMR data pointed out the flexibility of the C-terminal helix that could allow interactors to get access more easily [203]. Structural information is also for this protein important to set up structure-based drug discovery approaches and develop inhibitors of Vpu mediated protein-protein interactions with potential antiviral properties.

Another HIV-1 protein that gathered a lot of attention was Tat (Transactivator of transcription). Tat is a transactivator

protein essential for activation and expression of HIV genes. Tat is considered a target for antiviral drug discovery for its role in defending HIV-1 infected cells from the immune system attacks [204]. Different Tat variants exist and are characterized by various lengths (i.e., 86 or 87 residues and 99 or 101 residues), the most clinically common type is the longer one [204]. Solution NMR studies were conducted with the 99-mer Tat Eli variant for which 2D [^{1}H, ^{1}H] NMR spectra were recorded and the solution structure was calculated. The Tat fold is characterized by 8 β-turns located in six different regions [204]. Indeed, there is a "region I" (residues from M1 to A21) with two β-turns, a cysteine-rich portion (also indicated as "region II", residues from 22 to 37) with two loops characterized by a considerable exposure to solvent, a "region III" (residues from 38 to 48) with a loop and a β-turn, a basic region (also indicated as "region IV", residues from 49 to 59) with an extended conformation, a glutamine-rich region (also indicated as "region V", residues from 60 to 72) with two β-turns and a C-terminal part (also indicated as "region VI", starting at residue 73) provided with three β-turns [204]. Interestingly, "region I", that is a proline rich portion, forms the core of Tat Eli and its central residue W11 creates a hydrophobic cleft with F38 and Y47 from "region III" [204]. This study pointed out that active Tat variants in spite of different sizes have a similar fold with local structural variations due to mutations [204].

Structural features of the 86 amino acid long HIV-1 Tat protein variant -Zaire 2 isolate (HVIZ2)- were as well determined by homonuclear solution NMR techniques and MD (Molecular Dynamics) calculations [205] (Fig. **8B**).

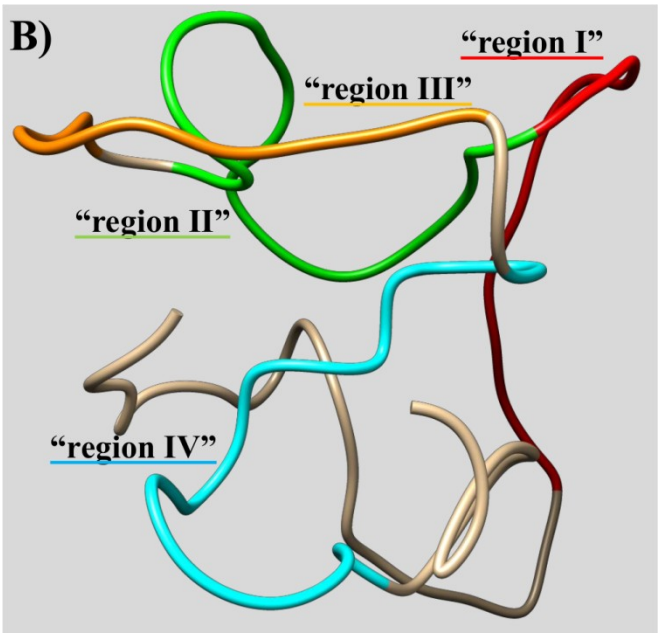


Fig. (**8**). (**B)** NMR solution structure of Tat (first conformer, PDB code: 1TIV [205]). The "region I" (cysteine-rich region, T20-C31), "region II" (core region, Y32-Y47), "region III" (basic region, R49-R57) and "region IV" (glutamine-rich region, Q60-Q76) are reported in red, light green, orange and cyan, respectively.

The above mentioned studies were conducted using chemically synthesized Tat, additional detailed solution NMR investigations were carried out by expressing in *E. coli* recombinant His-tagged-Tat(1-72) in the ^{15}N and ^{15}N/^{13}C labeled forms [206]. These studies were carried out at pH 4.1 when the protein, that as explained above, contains several cysteine residues, was mostly in the reduced state. Several triple resonance experiments were recorded and precise resonance assignments were obtained for different atoms (i.e., ^{1}HN, ^{15}N, ^{13}C', ^{13}Cα, ^{1}Hα and ^{13}Cβ) [206]. NMR data pointed out the prevalence of a random coil conformation for HIV-1 Tat protein at pH 4.1. The presence of multiple peaks and line broadening was correlated to the existence of multiple conformations largely in the cysteine-rich ("region II" in the previous example) and core regions [206]. The presence of minor cross-peaks was attributed to several effects including *cis-trans* proline isomerization, a minor content of oxidized cysteines and multiple conformers in slow equilibrium between each other [206]. However, the observed multiplicity of a few peaks along with the line-broadening and alterations of peak intensities as function of pH could as well indicate the presence of transiently folded structures for cys-rich and core regions at acidic and neutral pH [206]. Interestingly, the reduced Tat protein at pH 4.1 might mimic the Tat state in the intracellular environment where the interactions between this protein and certain binding partners (e.g., cyclin T1) occurs [206]. In this contest these data point out that the high level of disorder and the lack of a fixed conformation observed in Tat may be needed by the protein to bind a variety of different targets including proteins and nucleic acids [206].

HIV-1 RT (HIV-1 Reverse Transcriptase) is another protein playing a crucial function in HIV life cycle and has been considered a target in drug discovery as well. HIV-1 RT consists in an asymmetric heterodimer made up of the two subunits p66 and p51. Two different enzymatically active domains are present in p66 and consist in the polymerase and RNase H domains. The polymerase domain is in turn composed by fingers, palm, thumb, and connection units [207]. The thumb region was the focus of NMR solution structural studies [207]. ^{15}N- and ^{13}C, ^{15}N labeled thumb subdomain (residues 237-318) were prepared in *E.coli*; a combination of 2D and 3D NMR experiments were recorded and analyzed to get an almost complete assignment for backbone and side chain atoms [207]. The good level of dispersion of resonances in the 2D [^{1}H, ^{15}N] HSQC spectrum suggested a well-defined structure for the thumb domain. Structural calculations were conducted employing distance restrains obtained by 3D simultaneous ^{13}C- and ^{15}N-edited NOESY spectrum and from backbone torsion angles ϕ and ψ restrains generated by TALOS. The NMR solution structure of thumb domain of HIV-1 transcriptase is characterized by the presence of three α-helices (α1 from V254 to I270, α2

from Q278 to G285 and α3 from E297 to K311) (Fig. **8C**) [207].

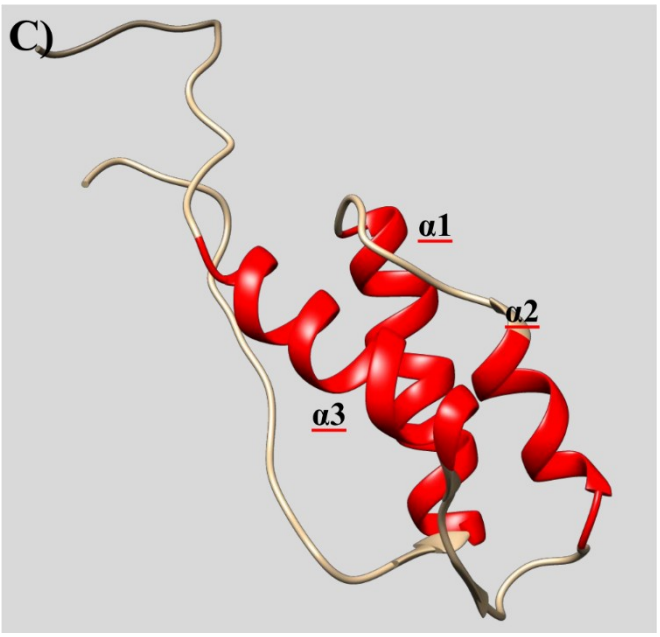


Fig. (**8**). (**C**) NMR solution structure of thumb domain from the reverse transcriptase of HIV-1 (first conformer, PDB code: 5T82 [207]). The three α-helices (i.e., α1 (V254-I270), α2 (Q278-G285) and α3 (E297-K311)) are shown in red in the ribbon representation.

NCp7 represents a nucleocapsid protein that through interactions with single stranded nucleic acids acts in packaging of genomic RNA, HIV infectivity and morphogenesis of virions [208]. Therefore, the interaction between the shortest deoxynucleotide sequence recognized by NCp7 (i.e., d(ACGCC)) and the 12-53 region of NCp7 was investigated by a combination of 2D NMR experiments [208]. NCp7 12-53 region contains two Zn knuckles of the type "CX2CX4HX4C". NMR studies showed that d(ACGCC) lies almost perpendicularly with respect to the sequence located between the two Zn knuckles (i.e., 29-RAPRKKG-35) [208]. In addition, the ribose phosphate backbone of the deoxynucleotide sequence points toward the solvent and the single ribose rings are instead oriented towards the N-terminal Zn knuckle [208]. Many hydrophobic contacts and H-bonds stabilize the complex between d(ACGCC) and the 12-53 region of NCp7. Indeed, the methyl groups of different NCp7 residues (i.e., V13, T24 and A25) provide hydrophobic contacts with the protons of ribose ring of A1 and C2 whereas, different amino acids from the C-terminal Zn knuckle (i.e., W37, Q45 and M46) are involved in interactions with the aromatic base protons of C2, G3 and C4 [208]. In addition, H-bonds are formed by the R32 side chain (positioned within the linker between Zn knuckles in NCp7) with the phosphate group of C4, by the Q45 side-chain (NCp7) with the N3 atom of C4, by the K47 side-chain (NCp7) with the 5' phosphate group of G3, by W37 (NCp7) and M46 (NCp7) amide protons with the carbonyl group of G3 [208]. Interesting large differences were found in the interaction between an inactive Cys23-(12-53)NCp7 mutant (with His23 in the first N-terminal Zn knuckle mutated to Cysteine) and d(ACGCC) with respect to what revealed for the wild-type NCp7 protein. The His to Cys mutation induces variations in the conformation of the Zn knuckle and in the internal folding of NCp7 and consequently interferes with protein function. It was not possible to obtain a precise NMR structure of the complex with the mutated NCp7 form because only too few NOEs could be observed in this case. This work further shows how NMR provides structural details essential for the design of selective antiviral agents blocking NCp7 [208].

The HIV genome, similarly to that of other retroviruses, is made up of two single-stranded RNAs forming a dimer through non-covalent contacts close to the 5′-end. The most conserved region of the HIV-1 genome is the 5′-leader and is involved in modulation of several processes during viral replication, such as RNA encapsidation in the course of virus assembly. HIV dimeric genomes are transferred from the cytoplasm to the assembly loci in the plasma membrane thanks to a few viral Gag proteins; in the plasma membrane extra Gag proteins pull together and the budding process takes place. The nucleocapsid (NC) domains of the viral Gag polyproteins and RNA portions included in the 5′-leader of the genome associate together leading to dimerization and packaging [210]. As for proteins, RNA structure modulates its function and thus, structural variations in 5'-leader are important to regulate genome function. Thus a great interest was given to NMR structural studies of the 5'-leader region [209-211]. NMR techniques were developed as well to reveal intra- and inter-molecular contacts within the entire leader, thus allowing to elucidate structural details and gain mechanistic insights about HIV-1 genome packaging and function [211]. A 356 residue long 5'-leader NMR sequence was prepared by including the complete 5′-untranslated region (5'-UTR residues 1-327) and the first 21 residues of gag. Residues encompassing the gag start codon (G328 to A356 (AUG)) are important for genome packaging and for the stability of RNA dimer. However, dimerization is also largely arbitrated by a conserved GC-rich palindromic dimer initiation site (DIS) included in the 5'-leader of the genome. By enzymatic ligation it was possible to prepare the 5' leader RNA sequence containing residues 1-327 in the unlabeled form and residues 328-356 (i.e., the AUG codon) enriched with ^{13}C for NMR studies. In this form the AUG region gives rise to ^{1}H-^{13}C heteronuclear correlations in the HMQC spectrum thus making the difficult NMR analyses of the large RNA fragment more doable. When the ionic strength is low the 5′-leader RNA exists mainly in the monomeric state and the HMQC signals of the AUG codon inserted in the long RNA fragment of the 5'-leader correspond to those of isolated AUG that is known to form an hairpin. When the ionic strength is increased and the equilibrium shifted from monomer to dimer, it happens that the AUG signals corresponding to the hairpin disappear and the ^{1}H-^{13}C

spectrum resembles a lot that of a complex between the isolated AUG portion and an isolated RNA U5 (Unique-5′) region (i.e., residues 105-115) [210]. In the dimeric form the portion A345-A356 shows sharp NMR signals characteristic of unstructured and mobile segment whereas, residues G328 to G344 of dimeric 5'-leader due to massive line broadening could not be revealed in the ^{1}H-^{13}C HMQC spectrum. Thus, NMR data clearly indicated that when the leader is in the monomeric state, the gag start codon is exposed in a dynamic hairpin while, it is buried in the dimeric RNA form by interaction with the U5 region. This structural model explains well why in the dimeric state both chemical reactivity of the gag start codon along with the *in vitro* translational activity of the genome are reduced [210]. NMR data are in agreement with a translation/packaging RNA structural switch process in which the dimer promoting GC rich loop of the DIS (Dimerization Initiation Site) hairpin is paired with U5 when the AUG is in the hairpin conformation (i.e., monomeric state) and is instead displaced by AUG as a consequence of the U5:AUG base pairing. Indeed, the conformational changes linked to the U5:AUG association at the same time capture the gag start codon and uncover the DIS and high-affinity NC interaction loci, thus decreasing translation and allowing a dimeric genome to pack [210].

Further NMR studies were conducted on the minimal 5'-leader region sufficient for packaging and known as Ψ-CES (Core Encapsidation Signal). This short 5' leader segment maintains still ability to dimerize, to interact with high affinity with NC and nevertheless, it possesses the NMR spectroscopic properties of the full length native 5'-leader [211]. In order to enhance the quality of NMR data, the native DIS region was replaced by a hairpin promoting tetraloop thus halving the dimension of the symmetrical Ψ-CES dimer. The structure of this Ψ-CES construct was obtained employing a fragmentation based ^{2}H-edited NMR techniques, which was implemented to further reduce the complexity of NMR spectra analyses and allowed to clearly identify long-range contacts in between base pairs [211]. The 5' untranslated portion of the HIV-1 viral genome contains a TAR (*trans*-activating responsive) element which is composed by 57 residues and consists in an imperfect stem-loop responsible for the viral replication [212]. TAR modulates transcription through its interactions with proteins belonging to virus and host cells [212]. In addition, the TAR element forms a kissing complex with certain RNA hairpin aptamers containing a partially complementary consensus apical loop (i.e., 5'-GUCCCAGA-3'). Interestingly, LNA (locked nucleic acid) nucleotides were inserted into the aptamer apical loop to increase the protection against nucleases as well as affinity for TAR. LNA represents nucleic acids analogues containing a 2'-O - 4'-C methylene linkage. Therefore, 2D and 3D NMR experiments were recorded to determine the structure features of a TAR/LNA modified aptamer kissing complex [212]. The results showed the presence of an unusual G-A base pairing which favored a higher level of stacking at the stem-loop junction. In addition, LNA residues improved the stability of the TAR/aptamer complex without loss of the canonical Watson–Crick base pairing and favoring a loop-loop conformation which is similar to an A-type helix [212].

In another work, the attention was focused on the DIS (Dimerization Initiation Site) of HIV-1. As also explained above, dimerization of viral RNA is important in the encapsidation and maturation of the virion [213]. Therefore, a combination of 2D NMR experiments (i.e., TOCSY, NOESY, DQF-COSY, [^{1}H, ^{15}N] HSQC, [^{1}H, ^{13}C] HSQC, HMBC and HNN-COSY) was exploited to get sequential assignment for the resonances of DIS and to determine its structure [213]. The obtained data showed the presence of two hairpins stabilized by six Watson-Crick GC base pairs. A network of intra- and intermolecular contacts formed by the adenines A8, A9 and A16 stabilizes instead the hinges which are positioned between the stems and the loops [213]. In addition, DIS contains three A-type helices which are coaxially aligned with each other and a "bulged in" conformation is found for the adenines A8 and A9. Interestingly, a previous work based on a crystallographic structure of DIS indicated an opposite conformation (i.e., "bulged out") for the same adenines, thus suggesting an equilibrium between two conformations in solution and the ability of the crystallization to move this equilibrium towards the "bulged out" form [213].

The CA (CApsid) and the SP1 (Spacer Peptide 1) are two domains of Gag, one of the main polyproteins of HIV-1 [219]. CA is composed of a NTD (N-Terminal Domain), and a CTD (C-Terminal Domain) connected by a flexible linker; a highly conserved sequence (i.e., MHR (Major Homology Region)) is present in the CTD region and plays a crucial function linked to assembly, maturation and infectivity [219]. Maturation of HIV-1 particles is accompanied by a multifaceted morphological alteration of Gag involving a coordinated set of proteolytic cleavage steps. The last stage in the maturation cleavage flow consists in the deletion of the SP1 peptide from CA, which induces reorganization of an immature lattice towards the ultimate mature conical shape. CA-SP1 tubes assembled *in vitro* are considered an intermediate assembly state of the maturation of the HIV-1 virus [219]. Therefore, different techniques (i.e., NMR, cryo-EM, and MD simulations) were employed to get information on the structural features of CA-SP1 and on the mechanism by which certain MIs (Maturation Inhibitors) act. The obtained results pointed out that in *in vitro* assembled CA-SP1 tubes SP1 is present in a dynamic equilibrium between a coil and a helical state. Studies with two MIs (i.e., Bevirimat and DFH-005 methanesulfonate) highlighted the ability of these MIs to move this equilibrium towards the helical state. Interestingly, the helical state of SP1 is favored also by a mutation (i.e., T8I), which blocks maturation [219]. The same mutation induces as well conformational and dynamical variations in CA [219]. This work with the support of solution and solid state NMR techniques clearly shows how small molecules may inhibit HIV maturation by

interfering with molecular motions and stresses out how dynamics of CA and SP1 is crucial for viral maturation [219].

NMR spectroscopy is also a valuable tool to get structural information on peptides and other compounds inhibitors of HIV and thus to better understand their mechanism of action along with the chemical modifications that could improve their antiviral potentials [214, 215, 220]. For instance, a work was focused on vMIP-II (viral Macrophage Inflammatory Protein-II), a protein found in the Kaposi's sarcoma-associated herpesvirus and able to block the entry of HIV-1 by acting on several chemokine receptors (e.g., CXCR4 and CCR5) [220]. Indeed, different peptides containing all D-amino acids (i.e., DV1 or vMIP-II (1–21) and its shorter version DV3 or vMIP-II (1–10)) were designed starting from a portion of vMIP-II (i.e., N-terminal region) known for its strong binding to CXCR4. Spin system identification and sequential backbone and side chain assignments for the designed peptides were achieved by canonical 2D homonuclear experiments (such as TOCSY, DQF-COSY and NOESY). Then, structural calculations were conducted by exploiting distance restraints obtained by NOESY spectra and backbone dihedral angle constraints retrieved by the $^{3}J_{HN\alpha}$ values [220]. The results revealed that these peptides were able to assume a partial structuration in aqueous solution with a few residues forming a turn-like structure [220]. Interestingly, DV3 assumes a structure similar to that of the corresponding vMIP-II peptide with all L-amino acids [220]. This structure information is useful as starting point to better comprehend the mechanism of interaction between receptor and peptide and of course is important to design novel series of peptide inhibitors of HIV-1 entry in the host cell provided with D-amino acids and thus with a better resistance to proteases attack.

Additional NMR work was conducted on HIV protease inhibitors. The HIV-protease is a homo-dimer and it processes gag and gag-pol polyproteins to produce the functional proteins [214]. When designing protease inhibitors it should be kept in mind that it contains an open-ended cylindrical hydrophobic cavity, the S1 and S1' sub-binding sites can well interact with hydrophobic side chains whereas, the S2 and S2' sub-sites may bind both hydrophobic and hydrophilic side chains with a preference for Glu or Ala amino acid residues [214]. PCU (PentaCycloUndecane) lactam-peptide based HIV protease inhibitors were designed [214]. A combination of 1D (i.e., ^{1}H and ^{13}C) and 2D (COSY, [^{1}H, ^{13}C] HMBC and [^{1}H, ^{13}C] HSQC) NMR experiments was exploited to get resonance assignments whereas, EASY (Efficient Adiabatic SYmmetrized)-ROESY experiment were employed for structural analysis [214]. NMR data showed that chirality of the PCU portion and its effect on the conformation of bound peptide side chain affected the inhibitor activity against HIV-1 protease. Furthermore, the structural information obtained by NMR was exploited in docking studies that in combination with analysis of IC50 values observed for the studied inhibitors allowed to identify a common binding mode for this class of compounds [214]. This interaction topology relies on a conserved network of H-bonds involving the norstatine-like functional group of the PCU hydroxylactam and active site residues (i.e., D25/D25'). This kind of inhibitors holds a certain interest as they are less toxic than the drug lopinavir [214].

Similarly NMR spectroscopy was implemented to study the interaction between HIV-1 protease and the KNI-272 inhibitor [215]. KNI-272 is a substrate-based peptidomimetic mimicking the transition state. Uniformly ^{13}C- and ^{15}N-labeled recombinant HIV-1 protease was prepared and 3D experiments of the homodimeric protease/inhibitor complex were recorded. Interestingly, the asymmetry of KNI-272 molecule led to the loss of symmetry in the homodimeric HIV-1 protease and thus allowed a differentiation of the spectra related to the two single monomers. NMR data also showed the presence of different protonation states for the catalytic Asp residues (i.e., D25 protonated state and D125 unprotonated state) which is a crucial information considering that potent protease inhibitors make contacts with the carboxyl groups of catalytic protease Asp residues [215]. Therefore, NMR led again to precise information on the interactions involved in HIV-1 protease/drug complexes thus paving the way for the design of original optimized antiviral agents targeting HIV protease [215].

5.2.2. Metabolomics studies

NMR-based metabolomics analysis of biofluids from HIV-1 infected patients pointed out differences in the metabolic profiles from healthy patients respect to those infected by HIV-1 undergoing or not ART (AntiRetroviral Treatment) [216]. From a general point of view, monitoring different metabolites, such as lipids, glucose and amino acids, can be considered a useful tool to determine the presence of HIV-1 infection and the effect of ART on infected patients [216]. For example, variations in lipids and glucose have been associated to ART-related disorders including lipodystrophy, hyperlipidaemia and hyperglycemia [216].

The levels of several molecules (i.e., sarcosine, MMA (MethylMalonic acid), D-glucose, choline and L-aspartic acid) in the plasma of HIV positive patients were found higher than those in the plasma of healthy people [128]. In addition, ART therapy favored a decrease in the amount of these metabolites but, reached levels resulted still higher than those observed in uninfected patients [128]. Similarly, the amounts of other metabolites (like 5β-cholestanol, L-lysine, acetoacetate and L-threonine) decreased in HIV patients with respect to the control group; upon ART therapy the levels of the same metabolites raised but they still resulted comprised between those found in infected patients without treatment and those relative to healthy patients [128]. Interestingly, two plasma metabolites (i.e., sarcosine

and choline) were suggested as promising biomarkers which could be used in the research on HIV/AIDS [128].

Similarly, another NMR-metabolomics work relying on analysis of 1D [^{1}H] spectra of serum samples in HIV-1 infected patients further highlighted that HIV-1 infection was associated to lower levels of a few metabolites (i.e., alanine, choline, glutamate, glycerophosphocholine, phosphocholine, taurine and valine) and increased levels of other molecules (i.e., creatine, glutamine, glycine, lactate and glucose) [128]. Therefore, these metabolites were considered promising biomarkers to be employed in studies on HIV-1 infection [128]. Indeed, the lower levels of alanine and the higher levels of glutamine and glucose can be related to the $CD4^+$ T cells count and thus to progression of viral infection [128]. A different study also based on high-resolution 1D [^{1}H] NMR spectra analyzed and compared sera samples from three different types of patients (i.e., healthy, infected and infected under ART) [221]. The described NMR metabolomics method, by comparing the spectra of sera samples from HIV positive patients (with and without ART) with those from the HIV negative control group, showed considerable differences in the regions characteristic of glucose, lipids and amino acids [216, 221].

Changes were also observed as concerning glycerol and choline which could be related to AIDS dementia complex [216, 221]. Nevertheless, NMR-based metabolomics highlighted the connection between HIV infection and the probability of incurring in cardiovascular complications especially in patients infected by HIV-1 and following ART [216, 222]. In fact NMR spectroscopy consents to get quantitative information on the number of particles of specific lipoprotein subgroups and within this context several analyses investigated the correlation between lipoprotein subclasses and cardiovascular diseases [222]. NMR-based metabolomics revealed in positive patients under ART higher levels of VLDL-p (Very Low Density Lipoprotein particles) and small LDL-particles as well as lower levels of HDL-p (High Density Lipoprotein particles) and large LDL-p than those found in healthy patients [216]. Interestingly NMR data relative to positive patients without ART revealed lower levels of small LDL-p, large LDL-p and HDL-p. These data are in agreement with increased frequency of cardiovascular disease (CVD) in HIV positive patients undergoing ART; this CVD risk could be in turn related to increased lipid levels in serum [216].

In addition, higher mortality is evidenced in HIV-infected treated patients with a poor immunological recovery [223]. NMR-based metabolomics studies were conducted to establish biomarkers for the incomplete immune reconstitution condition. Immunological recovery was represented by attaining a $CD4^+$ T-cell count equal at least to 250 cells/µl following 36 months of successful ART. Thus, plasma obtained from venous blood samples of HIV patients was analyzed and results compared [223]. Participants in this analysis satisfied the following criteria: more than 18 years, presence of HIV-1 infection and a pre-ART low CD4+ T-cell count <200 cells/µl at the lowest point. Investigations based on NMR spectroscopy, including acquisition of 2D DOSY (Diffusion-Ordered NMR Spectroscopy) spectra were exploited to evaluate the levels of lipids (i.e. triglycerides and cholesterol), sizes and particle numbers of different classes of lipoproteins (i.e., VLDL, LDL and HDL). In details, large, medium and small sizes for VLDL, LDL and HDL particles were taken into accounts [223]. High amounts of large high density lipoprotein (HDL) particles, HDL cholesterol and low density lipoprotein particles with larger sizes resulted hallmarks of a better immunological recovery after treatment. An incomplete recovery was instead associated to patients provided with high ratios of non-HDL lipoprotein particles [223].

The above described examples demonstrate the power of NMR-based metabolomics studies in the identification of biomarkers of HIV infection and how specific metabolites can be used to follow recovery after ART. NMR-based metabolomics can thus be employed for diagnostic purposes as well as to establish cellular metabolic pathways altered in HIV infection and eventually develop new antiviral agents and therapeutic routes.

5.2.3. NMR drug discovery

Different types of NMR experiments were exploited to identify new potential inhibitors able to act on HIV-1 proteins or genome and thus behave as promising antiviral molecules to be employed in novel therapeutic routes.

For instance, a fragment-based strategy relying on NMR-binding assays (i.e., STD spectra) and functional studies were conducted with a small molecule library to find out ligands of HIV reverse transcriptase [224]. This screening identified fragment-sized inhibitors targeting the reverse transcriptase of HIV-1 characterized by chemical scaffolds and mechanisms of action distinct from those of the canonical NNRTIs (Non-Nucleoside RT Inhibitors) and NRTIs (Nucleoside/Nucleotide RT Inhibitors) [224]. In details, three small molecules (i.e., *p*-hydroxyanilin, benzothiophene and oxime scaffolds) resulted able to inhibit RNA- and DNA-based DNA polymerase activities of the HIV-1 reverse transcriptase in the micromolar range [224]. In addition, these molecules blocked the activities also of the reverse transcriptase variants characterized by one of the three major mutations (i.e., K103N, Y181C, or G190A) that are resistant to the action of the canonical NNRTIs [224]. Interestingly, data showed the ability of the three fragments to block the activity of the reverse transcriptase of MoMLV (Moloney Murine Leukemia Virus) but not that of Klenow fragment of Escherichia coli DNA polymerase I, thus suggesting specificity for reverse transcriptase [224]. Steady-state kinetic DDDP (DNA-Dependent DNA Polymerase) inhibition assays showed the competition between *p*-hydroxyanilin and T/P (DNA Template/Primer) bindings to the reverse transcriptase and the competition between oxime

scaffold and dNTP (deoxyribonucleoside triphosphate) substrate interactions with reverse transcriptase [224].

This study paved the way for the design of potent antiviral agents targeting HIV reverse transcriptase and based on the three newly identified chemical scaffolds.

Another work was focused on viral entry blocking agents targeting the gp41 coiled-coil pocket [225]. To allow penetration of HIV-1 into host cells viral and cellular membranes need to fuse together and the envelope glycoprotein gp41 plays a key role within this process. Gp41 possesses an ectodomain consisting of a trimer and a central parallel coiled coil is made up by the N-terminal helices. C-terminal helices in each monomer bind together in an anti-parallel manner inside the central trimeric coiled coil organization giving rise to a six-helix bundle arrangement that likely represents the active conformation needed for fusion. Agents that interfere with this six-helix bundle organization induces inhibition of HIV fusion as well as replication *in vitro* [226]. Fusion takes place when the C-terminal helices, that are embedded in the viral membrane are brought along with the N-terminal helices in the host membrane giving rise to the fusogenic six-helix bundle [225]. Peptides encompassing the C-terminal region of gp41 are able to hamper viral fusion during infection [225]. An engineered version of gp41 was developed for drug discovery studies [226] and used for ligand screening. This gp41 version contains a N-helix/C-helix interface with a permanently open 'Trp-Trp-Ile' pocket, such cavity has been suggested as a potential target locus for small molecule inhibitors of HIV-1 entry. The pocket owes its name from three highly conserved C-helix residues: W628, W631 and I635 [225].

An NMR screening was set-up based on the acquisition of 2D [^{1}H, ^{13}C] HSQC spectra of a (^{13}C, ^{15}N) double labeled gp41 construct in its "apo" form and in presence of ligands and by monitoring shifts in methyl group resonances. This study led to identification of a potent benzamide analogue gp41 interactor [225]. Chemical shifts perturbation studies indicated that binding of the compound affected backbone H_N protons of residues G572, W571, and V570 and the methyl group resonances of L568, I573 and V570, these residues belong to the exposed coiled coil structure that the engineered gp41 construct was supposed to resemble. Then, 3D and 2D NMR experiments were recorded to get detailed information on the interaction between the identified compound and the protein [225]. 3D filter edited NOE spectra allowed detection of intermolecular NOEs between protein and ligand and thus further confirmed binding of the compound to the exposed hydrophobic pocket, highlighting that the ligand is able to provide contacts with adjacent monomers that contribute the sides of the hydrophobic cavity [225]. Based on these structural findings different chemical modifications of the starting compound were planned thus leading to the synthesis of analogues [225]. In the end such work brought to the discovery of an original group of non-peptide and low molecular weight inhibitors with micromolar activities (EC50 values range 3-41 μM) in cell fusion assay [225]. The new molecules reported here resulted however provided with lower potency with respect to a few peptide inhibitors.

In spite of these results, the proposed NMR strategy and the bound conformation of ligand retrieved thanks to NMR data can be surely employed for future further non-peptide inhibitors optimization.

To protect immunogenic peptide epitopes positioned on its envelope spike (i.e., a trimeric arrangement made up of gp120 and gp41 glycoproteins) HIV-1 exploits several host-derived N-linked glycans. 2G12 is a monoclonal antibody that recognizes clusters of high-mannose type glycans that occur on the surface of gp120. Indeed, broadly neutralizing antibodies towards gp120 and gp41 have been obtained from HIV-1-infected patients and play a protective antiviral function in animal models [227]. Getting structural information on the interaction between such glycans and antibodies is important to design novel carbohydrate-based HIV vaccines [227, 228]. With this in mind, STD and transferred NOE experiments were employed to investigate the interaction between the monoclonal antibody 2G12 and different oligomannosides. STD experiments, as explained in the Introduction section, are particularly useful to determine binding epitopes and also dissociation constant values for a carbohydrate-antibody complex [33]. In details, solution NMR studies were conducted to analyze binding of Man9GlcNAc2 of gp120 to the antibody 2G12; Man9GlcNAc2 is composed of three arms named D1, D2, D3 made up of three, two, three sugar units, respectively and a main skeleton (NN arm) with two GlcNAc moieties and one additional Man unity to which the D1 and D2 + D3 arms are linked [227, 228]. NMR data of different sugar components pointed out that the branched pentamannoside P (including the D2+D3 arms) presented two alternate binding modes for the same binding site of 2G12 involving both arms [228]. It was also shown that the minimum epitope with good affinity for 2G12 in solution was constituted by the trisaccharide sequence Manα1-2Manα1-2Man present on the D1 arm of the Man9(GlcNAc)2 glycan of gp120 [228]. Additional interaction studies were conducted on the isolated constituents of Man9GlcNAc2 including the aminoethyl α-oligomannosides Tri-1 and Tri-2, (i.e., trisaccharides including the isolated D2 and D3 arms, respectively) [227]. The results of such studies revealed similar affinities of Tri-1 and Tri-2 for the monoclonal antibody 2G12 as expected because of the bimodal interaction topology observed for the entire pentamannoside (D2+D3) [227]. Furthermore, the obtained data pointed out that the interaction between mannose-type glycans of gp120 and the monoclonal antibody 2G12 strongly depends on the mannose ring of the internal core [227, 228].

One of the NMR routes proposed for the identification of new antiviral molecules was developed focusing on the interaction between the TAR (Transactivation Response) region on the HIV-1 RNA and the tat protein [229]. The

HIV-1 TAR region is a stem-loop positioned at the 5′ extremity of growing viral transcripts [229]. Interaction between TAR element and tat induces activation of RNA transcription. Thus, molecules targeting such interaction would possess antiviral properties by interfering with viral replication. A fragment-based screening approach was set-up by introducing a chemical probe targeting the Tat binding site of TAR to favor further fragments binding to the RNA [229]. In the first step, to identify the chemical probe different NMR experiments (i.e., STD and 2D NOESY) were conducted to screen a small group of arginine analogues [229]. Interestingly, the screening led to MV2003, a molecule with the key element able to fit the TAR binding cleft (i.e., a positively charged arginine residue) and with a proton-rich group (i.e., methoxynaphthalene) which could favor a better detection of ILOEs (InterLigand NOEs) with the fragments [229]. The next phase consisted in the screening of a library of 250 small fragments against TAR in presence of MV2003. STD experiments were carried out and based on the criteria STD signal ≥ 10% and the presence of ILOEs between fragment and MV2003, 20 potential hits were identified. Next, additional NMR experiments (ligand-observe NOE spectroscopy techniques) were conducted for the putative hits alone, in presence of TAR, in presence of MV2003 and in concurrent presence of both TAR and MV2003 [229]. Interestingly, six molecules were confirmed as ligands able to interact with the viral RNA TAR region only in presence of MV2003 [229]. Furthermore, additional structural and docking studies revealed the ability of the chemical probe to favor formation and stabilize a small cavity located in the major groove of the RNA and in which the fragments could bind weakly but with increased specificity [229]. This work relying on observation of ILOEs paves the way for the development of antiviral bidentate compounds targeting with elevated specificity and increased affinity the RNA TAR element.

Another class of ligands of the TAR element of HIV-1 RNA is represented by the phenotriazines. Phenothiazine, if compared to other RNA binding drugs, is a molecule not only characterized by high bioavailability and low toxicity but also by low molecular weight and single positive charge [230]. All these features make phenothiazines promising starting compounds to target the TAR region of HIV-1 RNA [230]. With this in mind, a library of phenothiazine analogues was designed by introducing different substitution in the ring and n-alkyl linker and binding studies with TAR RNA were conducted. This NMR screening consisted mainly in STD experiments observing the ligands resonances and in Chemical Shift Perturbation (CSP) experiments looking at the shifts of the RNA imino protons induced by compound binding. DLB (Differential Line Broadening) NMR assays were performed as well and the results highlighted the structural features crucial to improve binding affinity [230]. For example NMR data pointed out that the presence of 3 or 4 methylene groups between the aromatic ring and the heteroatom favored higher affinity of binding [230]. The introduction of flexible linear chains led to the same result due to the their ability to easily fit into the minor groove of RNA [230]. Furthermore, the affinity was improved by heteroatoms or functional groups able to form H-bonds and by substitutions able to increase the number of van der Waals interactions [230]. This work stresses out the importance of NMR in suggesting to the medicinal chemists the ad hoc modifications to be inserted in the phenothiazine scaffold to generate more potent inhibitors of the TAR element of HIV-1 RNA.

The attention of another work was focused on HIV-1 Vpu, protein that favors the release of viral particles, and its interaction with the ubiquitin ligase SCF-β-TrCP that stimulates the HIV-1 receptor CD4 degradation by the proteasome [231]. Binding of the DpSGXXpS (pS=phospho Serine) consensus sequence to β-TrCP is fundamental for development of pathological conditions including HIV infection and cancer [231]. Two peptides (LIERAEDpSG or P1, and EDpSGNEpSE or P2) which encompassed partially or totally the consensus motif located in Vpu were synthetized and screened by NMR techniques against β-TrCP [231]. STD NMR experiments were exploited to prove binding and to identify the peptide residues in close contacts with the protein. In addition the transferred NOESY experiments allowed gaining the structures of the two peptides bound to β-TrCP [231]. NMR data highlighted that P1 and P2 efficiently interact with the β-TrCP WD domain; in both P1 and P2 peptides the DpSG motif forms a bend and in agreement with STD experiments, residues in close contacts with protein appear exposed on the surface. In major details, P2 assumes a large bend encompassing the DpSGNEpS segment whereas, P1 assumes a turn type tructure involving the IERAED region. The presence of an additional phosphorylated serine contributes to the binding of P2 by favoring electrostatic interactionss and H-bonds with β-TrCP whereas, the -LIER- peptide segment contributes to the binding of P1 by favoring additional hydrophobic contacts with β-TrCP. The interaction between β-TrCP and peptides may lead to different structures although the conformation of the DpSG fragment is conserved. Different structurations may modulate selectivity towards β-TrCP [231].

5.2.4. Other studies

A considerable number of HIV-1 inhibitors acting at different levels of viral infection has been reported in literature. Within such studies NMR spectroscopy was largely implemented to control the progress of synthetic routes and check compounds purity and identity similarly to what reported in previous paragraphs for inhibitors of other viruses.

Lignans, are plant secondary metabolites produced by oxidative dimerization of two phenylpropanoid units [232]. They are abundant in the plant kingdom and are present in roots, rhizomes, stems, leaves, seeds and fruits. Lignans are

involved in plant development, and in the protection of plant against several pathogens and pests. Nevertheless, lignans possess relevant pharmacological activities, such as antiviral properties [232]. Dibenzylbutanediol lignans and their analogues were synthetized and tested on HIV Tat transactivation in human epithelial cells and the results led to the identification of novel inhibitors of HIV replication [232]. Similarly, 3′-O-methylnordihydroguaiaretic acid (3′-O-methyl-NDGA) is a plant lignan isolated from *Larrea tridentata*, that is able to interfere with HIV-1 replication in infected human cells by blocking HIV Tat-modulated transactivation and the process of proviral transcription; methylated NDGA analogues were synthetized and evaluated and the results pointed out that tetramethyl NDGA (Fig. **9A**) behaves as a more potent anti-HIV agent respect to the mono- and dimethylated analogues [233].

Lectins are another class of molecules extracted from plants and characterized by the ability to inhibit viral entry by targeting the envelope glycoprotein gp120 of HIV-1 [234]. Efforts were made to overcome high manufacturing costs, formulation and potential mitogenicity associated with lectin-based inhibitors by developing polymeric synthetic lectins, based on benzoboroxole (Bzb). Antiviral activity was revealed in high molecular weight Bzb-containing lectin polymers and among them a few potent agents inhibiting HIV-1 entry could be identified [234]. Such synthetic polymeric compounds resulted provided with weak affinity towards sugars that are similar to those on the HIV envelope [234].

Further studies on plant extracts revealed that *Justicia gendarussa,* a medicinal plant from Vietnam has antiviral activity against HIV-1. Indeed, the extracts from the stems and roots of this medicinal plant contain patentiflorin A, an ANL (ArylNaphthalene Lignan) glycoside [235]. Patentiflorin A is likely involved in blockage of HIV-1 reverse transcription and its inhibition effect towards drug-resistant HIV-1 strains is higher than that of either AZT (i.e., nucleotide analogue) and nevaripine (i.e., a non-nucleotide analogue) [235].

To obtain promising small molecule inhibitors of HIV-1 entry, which hamper binding between the Phe43 pocket of HIV-1 gp120 and the host CD4 receptor, CD4 mimetics were developed (Fig. **9A**) [236]. Studies with water soluble CD4 mimetics revealed that the presence of a tetramethylpiperidine ring in their structure favored a high anti-HIV activity and synergistic anti-HIV effect with a neutralizing antibody. Similarly, molecules characterized by the coexistence of a cyclohexylpiperidine ring with a 6-fluoropyridin-3-yl ring were associated with elevated anti-HIV activity and absence of cytotoxic effect [236].

Cosalane is another molecule provided with anti-HIV properties and whose mechanism of action relies again on hampering the interaction between gp120 and CD4 and to interfere with post-attachment event preceding reverse transcription. In order to improve its inhibitory activity, cosalane derivatives were designed, synthetized and evaluated leading to discovery of compounds with increased potency against HIV-1integrase and protease [237]. Diketo acids (e.g., S-1360 and L-731988) are molecules which possess a significant and selective activity against HIV-1 integrase. A series of substituted indole-β-diketo acids were developed and resulted able to inhibit viral integrase at low micromolar concentrations [238].

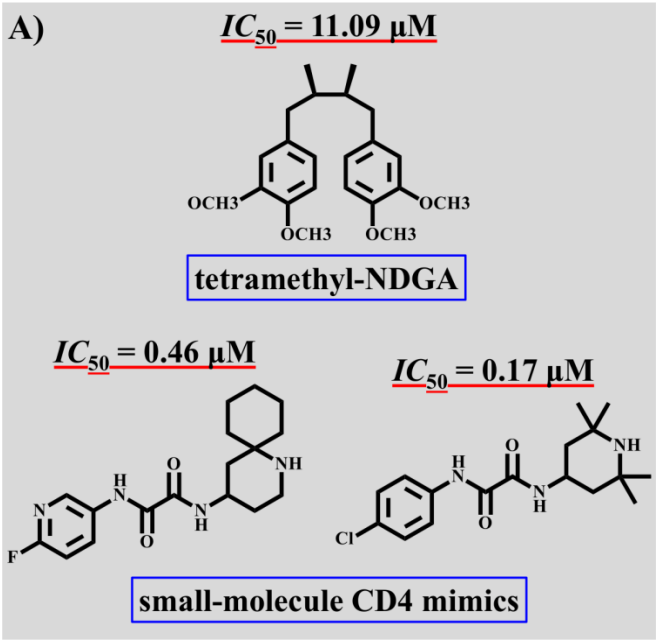


Fig. (**9**). Chemical structures of HIV-1 inhibitors. (**A**) For the tetramethyl-NDGA inhibitor the IC50 value refers to inhibition of HIV Tat-regulated transactivation in COS Cells [233]. The anti-HIV activity of CD4 mimics was determined with a TZM-bl assay employing an R5 primary isolate cYTA48P strain and TZM-bl cells (See ref. [236] for details about assay set-up and IC50 determination).

Many different series of compound inhibitors of HIV-1 protease were developed such as macrocycles with different flexibility involving P1'-P2' interactors (where P1' and P2' represent the first and second C-terminal positions following the cleavage site -P1-cleavage point-P1'- in a generic protease substrate) that were characterized by considerable enzyme inhibitory and antiviral activities [239].

Another study described design and synthesis of inhibitors of HIV-1 protease based on a hydroxyethylamine core and different phenyloxazolidinone P2 ligands (the peptide substrate cleavage site is located in between P1 and P1' going from the N-terminus to C-terminus, P2 is located immediately N-terminally with respect to P1). This study led to the identification of molecules with picomolar binding affinity and low nanomolar antiviral potency [240]. HIV-1 protease inhibitors containing a squaramide scaffold as P2 ligand were also synthetized and a molecule with an inhibitory activity in the nanomolar range was identified [241]. Nevertheless, a few non-natural peptides and peptoids including the pentacycloundecane (PCU) lactam were generated as inhibitors of the resistance-prone wild type C-

South African HIV-protease [242]. For a few of these agents IC50 values between 0.5 and 0.75 μM could be detected. NMR spectroscopy was particularly useful to investigate the structure of these peptides/peptoids [242]. Another interesting study highlighted how the insertion of *N*-phenyloxazolidinone-5-carboxamides into a (hydroxyethylamino)sulfonamide scaffold represented a useful route to get HIV-1 protease inhibitors. In detail, series of compounds containing diverse P2 phenyloxazolidinone and P2′ phenylsulfonamide groups were synthesized and a few molecules with Ki (inhibitory constant) values in the picomolar range could be obtained [243].

Synthesis and evaluation of a group of HIV-1 protease inhibitors relying on two pseudosymmetric dipeptide isosteres have been reported as well [244]. These inhibitors included, N-phenyloxazolidinone-5carboxamides inside the hydroxyethylene and (hydroxyethyl)hydrazine dipeptide isosteres as P2 and P2′ ligands. Among this series of inhibitors, compounds provided with (S)-phenyloxazolidinones linked at a site proximal to the central hydroxyl group exhibited an inhibitory action in the low nM range towards wild-type HIV-1 protease [244].

Another set of interesting molecules is composed by N-benzylated coumarin linked to the drug AZT (AZidoThimidine) that are able to simultaneously block protease and reverse transcriptase activities (Fig. **9B**) [245]. STD NMR clearly pointed out interaction of a few of such compounds to the HIV-1 protease. Although the activities of these molecules are still poorer than those required (i.e., far from the needed nanomolar IC50 values), the identified dual inhibitors can be considered as promising starting point for future optimization protocols [245].

Another work reported a new tetra-hydro-indazolyl-benzamide derivative, that was synthesized through oxazolone chemistry, and showed capacity to inhibit HIV viral replication with low cytotoxicity and high selectivity acting in the early replication phases without blocking viral integrase and likely behaving as an inhibitor of the reverse transcription step [246].

In a different work, fluorine substituted 1,2,4-triazinones were synthetized and analyzed for antiviral and anticancer activities [247]. A few of these compounds were provided with very good anti-HIV properties in MT-4 cells others, were dually inhibitors of HIV and CDK2 (Cyclin dependent kinase 2) thus being provided also with anticancer potentials [247].

Antiviral agents could be obtained by setting up a HTS-TR-FRET (High-Throughput Screening method based on a Time-Resolved Fluorescence Resonance Energy Transfer) assay employing the C-Terminal Domain (CTD) of HIV-1 capsid and a library of pharmacologically active compounds (including 1,280 drugs exhibiting *in vivo* activity) [248]. The result was the identification of an organoselenium compound (i.e., ebselen) able to block the HIV-1 capsid CTD dimerization [248] (Fig. **9B**). Interestingly ebselen resulted able to produce chemical shifts perturbations in the spectrum of the CTD domain and to destroy the dimeric arrangement when present in at least a two-fold excess with respect to the protein [248].

The inhibition of HIV transmission by polyanionic and polyzwitterionic compounds based on cyclodextrins was further analyzed [249]. The inhibitory activity of a few of such kind of molecules against replication of specific HIV-1 strains is in the low micromolar range and is associated with low toxicity towards the host cell [249].

A recent work was focused on the design and synthesis of novel molecules targeting the TAR region of HIV-1 RNA [250]. The starting point was the 6-aminoquinolone WM5, a selective ligand of TAR RNA possessing antiviral activity (Fig. **9B**). Indeed, different WM5 analogues were generated by introducing guanidine or amidine groups in precise positions (i.e., N-1, C-6 or C-7) along with diverse protonable groups with the aim to improve interactions with the phosphate groups in the TAR region of viral RNA (Fig. **9B**) [250]. This class of compounds was assayed for the ability to interfere with HIV-1 replication in MT-4 cells as well as for the capacity to inhibit formation of the Tat-TAR complex. A few of the resulting molecules were characterized by an anti-HIV-1 activity in the sub micromolar range [250]. Three WM5 derivatives could interfere with formation of the Tat-TAR complex with enhanced Ki values respect to the starting WM5 compound. Results from this work highlighted also that when at the N-1 site of the quinolone/naphthyridone core a pyridine-based protonable side chain is introduced, compounds are able to hamper formation of Tat-TAR complex and consequently block HIV-1 replication [250].

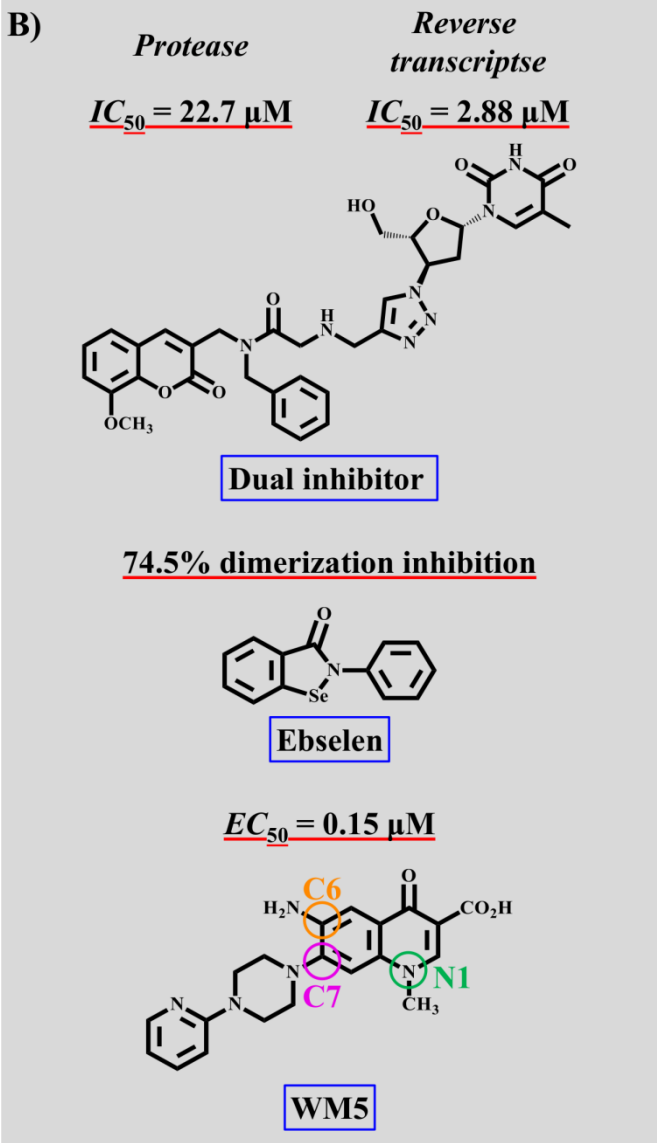


Fig. (**9**). (**B**) The capacity of the dual inhibitor to inhibit both HIV PR and RT respect to the drugs ritonavir and AZT was assessed as reported in reference [245]. The percentage of CTD dimerization inhibition was evaluated by a TR-FRET assay as described in reference [248]. EC50 value for WM5 represents the concentration needed by MT-4 cells to reach 50% protection from HIV-1 related cytopathogenicity [250].

6. CORONAVIRUSES

Coronaviruses (order Nidovirales, family Coronaviridae and subfamily Coronavirinae) are enveloped viruses characterized by a large "ssRNA(+)" genome composed by approximately 26-32 kb (kilobases) [9, 16, 251-254]. There are 26 coronavirus species that are distributed among four genera (i.e., Alpha coronaviruses, Beta coronaviruses, Gamma coronaviruses and Delta coronaviruses) [16, 251, 253, 254].

The Alpha subgroup (also indicated as α-CoV) includes different viruses isolated from felines (i.e., FECV (Feline Enteric CoronaVirus) and FIPV (Feline Infectious Peritonitis Virus)), from porcines (i.e., TGEV (Transmissible Gastro-Enteritis Virus), PEDV (Porcine Epidemic Diarrhea Virus) and PRCoV (Porcine Respiratory CoronaVirus)), from canines (i.e., CCoV (Canine CoronaVirus)) and various bat-CoVs [251]. In addition, the Alpha subgroup includes two human viruses (i.e., HCoV-229E and HCoV-NL63) and phylogenetic data pointed out that they both might derive from bat-CoVs [251]. The second subgroup (i.e., Beta coronavirus or β-CoV)) comprises viruses isolated from animals (e.g., MHV (Mouse Hepatitis Virus) from mice and BCoV (Bovine CoronaVirus) from bovines) and viruses with zoonotic origins that overcame the species barriers and infected humans (such as HCoV-OC43 and HCoV-HKU1, SARS-CoV, SARS-CoV-2 and MERS-CoV) [9, 11, 16, 251-254]. The viruses belonging to this β-CoV genus are further divided into four subsets (lineage A, lineage B, lineage C and lineage D) [16]. In addition, the three β-CoV SARS-CoV (likely originated from bat/palm civet), MERS-CoV (possibly originated from bats but next transmitted to dromedary camels that represent its largest reservoir), and the most recently identified SARS-CoV-2 (possibly originated from bats and after mutations able to infect different animals [255]) are associated with higher mortality rates in human populations [9, 11]. Instead, the gamma group (also indicated as γ-CoV) includes birds viruses like the Avian Infectious Bronchitis Virus from chicken with an exception, i.e., the coronavirus SW1 that is instead from beluga whale. Finally, the delta group (also indicated as δ-CoV) -created in 2012- regroups viruses (e.g., HKU11, HKU12 and HKU13) isolated from different mammals and birds [11, 251, 256].

From a morphological point of view, the name "coronavirus" derives from their peculiar crown-like shape when they are observed through an electron microscope; this particular shape is due to peplomers of the spike [S] glycoprotein irradiating outside from the virus envelope [9, 251]. Generally, coronaviruses genome encodes five different structural proteins: S, M (Membrane) and E (Envelope) glycoproteins, HE (Hemagglutinin Esterase) and N (Nucleocapsid) protein. Envelope proteins and N protein occur in all virions while, only a few β-CoVs possess HE [251].

The S proteins are organized in homotrimers and interact with the virion (viral particle) membrane though their transmembrane regions located at the C-termini, that are also involved in interactions with M (Membrane) proteins. Instead the N-terminal portions are responsible for the interaction between a virion and the receptors on the surface of the plasma membrane of the host cells [16, 251]. These S proteins are characterized by a S1 subunit, which is composed by a signal peptide, a NTD (N-Terminal Domain) and RBD (Receptor Binding Domain), and a S2 subunit, which is composed by FP (Fusion Peptide), HR1 and HR2 (Heptad Repeat 1 and 2), TM (Transmembrane), and CP (Cytoplasmic) domains [9]. Interestingly, SARS-CoV-2 S2 subunit has a high level of conservation with 99% similarity

with respect to SARS-CoV and two coronaviruses isolated from bats (i.e., Bat-SL-CoVZC45 and Bat-SL-CoVZC21) [9]. In many coronaviruses, the proteases of the host cells cleave the single trimer into a subunit S1, which encompasses the apex of the trimer and contains the regions that are recognized by host receptors, and in a S2 subunit, which is attached to the viral membrane [16, 257]. In addition, the two subunits have similar sizes, distinct functions and are noncovalently bound in the prefusion conformation [16, 257].

The envelope of coronaviruses also includes the M glycoproteins, which are characterized by 3 transmembrane regions and are involved in the fusion step of virus membrane with host-cell membrane [16, 251]. The function of M proteins strongly depends on glycosylation, which occurs in the Golgi apparatus [251]. Instead, the E (Envelope) glycoproteins comprise 76-109 residues, are fastened to the viral membrane by an N-terminal region, that is about 30 amino-acids long, and are involved in assembly and morphogenesis of virions within the cell [251]. The last components common to all coronaviruses particles are the N (Nucleocapsid) proteins representing phosphoproteins composed by two domains both able to bind viral RNA genomes through several mechanisms. N proteins play a fundamental role in different steps of viral life cycle including replication and transcription and in virion structure as they are positioned either at the replication/transcriptional site of the coronaviruses and also at the ERGIC (endoplasmic reticulum (ER)-Golgi intermediate compartment) that is the locus where the virus is collected [9, 251].

Interestingly, HE (Hemagglutinin Esterase) is a receptor-destroying enzyme which characterizes viruses belonging to lineage A subset of β-CoV group (e.g., HCoV-OC43, HCoV-HKU1), favors the exit of virions of new generation from the infected host cells and prevents the binding of virions to non-permissive host cells or decoys [11, 16, 251, 252].

In the next paragraphs, the attention will be focused on the three coronaviruses with the largest mortality incidence in humans (i.e., MERS-CoV, SARS-CoV and SARS-CoV-2).

6.1. Mechanism of action

The key element in the first step of coronavirus infection (i.e., virus attachment) consists in the interaction between the S protein and the receptors located on the surface of the host cell membrane [251, 258]. In the case of β-CoVs (e.g., SARS-CoV, MERS-CoV and SARS-CoV-2), the spike protein exploits its RBD to interact with the receptor and this produces a conformational change in the S-protein that induces the fusion process between viral and host-cell membranes through the S2 subunit. Indeed proteases of the host cell by performing S1/S2 cleavage work as S protein "activators" for viral entry [258]. For example, in the case of MERS-Cov, the proteolytic enzyme furin first cleaves and activates the S protein into the S1 and S2 subunits at the biosynthesis stage by interacting at the R751/S752 site [259]. In addition, after viral entry a second step of cleavage occurs at the R887/S888 site, that is positioned close to the fusion peptide, on the S2 subunit thus converting S2 into S2'[259].

The S proteins of SARS-CoV and SARS-CoV-2 recognize the ACE2 (Angiotensin-Converting Enzyme 2) receptor on the surface of the host cell membrane whereas, the S protein of MERS-CoV recognizes the DPP4 (DiPeptidyl Peptidase 4) receptor [9, 258, 260, 261].

Interestingly, DPP4 is described as responsible for the host species restriction [261]. Indeed, MERS-CoV can infect bat, human, camel, non-human primate and swine cells but not the hamster and ferret cells due to the differences in five amino acids involved in the recognition of DPP4 [261]. Furthermore, glycosylation of DPP4 found in mouse is considered as an additional barrier against the infection by MERS-CoV [261].

After binding of the spike protein to the receptor, the next step of coronavirus life cycle includes entering of the virion in the host cell and may occur by two alternative mechanisms [262]. The first one (i.e., membrane fusion) is based on the fusion of the viral and host membranes starting from outside the cell and the following entry of the viral RNA genome into the cytosol [262]. Instead, the second mechanism (i.e., endocytosis) consists in the formation of vesicles through envelopment of the receptor-attached virion by host cell membrane and the consequent entrance into the cytosol [262]. The result associated with both mechanisms is however the release of viral RNA genome into the cytoplasm and subsequent uncoating of the RNA [262]. Then, two replicase polyproteins (i.e., pp1a and pp1ab) are formed by the translation of two genes on the viral RNA strand (i.e., ORF1a (Open Reading Frame 1a) and ORF1ab) and are cleaved by viral proteases into 16 non-structural proteins which in turn associate with each other and form the RNA replicase-transcriptase complex [262, 263]. This complex is responsible for the formation of full-length RNA(-) by replication and transcription. Replication leads to copies of full-length RNA(-) which, in turn, are exploited as template for the production of full-length RNA(+) genomes [262, 263]. Transcription consists in an intermittent/discontinous process which leads to a subgroup composed by 7-9 sub-genomic RNAs including those encoding for all structural proteins [263, 264]. Interestingly, different RNA sequences of the viral genome (e.g., 5' and 3' UTRs (UnTranslated Regions) and TRS-L (Transcription Regulating Sequences of the Leader) along with those preceding each gene TRS-Bs) are responsible for the modulation of both replication and transcription [265]. In addition, the switch between these processes is possibly modulated also by interaction with proteins that are engaged specifically by the replication or transcription complexes [265]. The next phase is the translocation of newly generated RNA(+) genome and proteins (i.e., structural- and nonstructural ones) to the lumen of ERGIC (Endoplasmic Reticulum (ER)-Golgi Intermediate Compartment) in which the assembly of virions occurs [262-264]. The viral particles undergo a maturation process which

is characterized by different post-assembly maturation mechanisms (e.g., proteolytic and oligo saccharide processing) [266]. The new virions bud off from the Golgi as vesicles that then, are moved towards the host cell membrane where they fuse with cell membrane and are next released in the extracellular region by a mechanism through which the integrity of host cell is not damaged (i.e., nonlytic exocytosis) [262].

MERS-CoV, following its entering in the lung alveoles, causes an infection that is associated with a significant answer of the immune system and high levels of cytokine release which in turn favor inflammation and a large amount of fluid in lungs [267]. The result is the emergence of different diseases (e.g., high fever, chills, rigors, severe cough, dyspnea, and hypovolemic shock) that in a few cases are accompanied by symptoms of pneumonia whereas, in others are accompanied by symptoms of COPD (Chronic Obstructive Pulmonary Syndrome) like disease [267]. The transmission of the MERS-CoV infection is characterized by different routes such as "dromedary camels to humans", "bats to camels", "camels to camels", "cattle to man", "dogs to humans", "cats to man", "bats to humans" and "humans to humans" [267].

As concerning SARS-CoV, symptoms include fever, malaise, myalgia, headache, diarrhoea, and rigors, however, no specific symptom or group of symptoms can be related explicitly to a clear diagnosis [268]. Generally, fever is one of the most frequent symptom; cough (initially dry), shortness of breath, and diarrhoea also occur usually during the first and/or second week of illness and those in more severe conditions quickly experience respiratory distress. Bats are considered as potential first reservoir of SARS-CoV from which the virus moved to civet cats and humans. However, infection transmission mostly occurs by the "human to human" route [268].

The most diffused symptoms in patients affected by SARS-CoV-2 infection are fever, dry cough and tiredness whereas, the less diffused ones are aches, pains, sore throat, diarrhoea, conjunctivitis, headache, loss of taste or smell, a rash on skin, or discoloration of fingers or toes [269]. In addition, patients in severe conditions are affected by breathing difficulties / shortness of breath, chest pain or pressure and loss of speech or movement [269]. Studies of genomic sequencing highlight that the Receptor Binding Domain (RBD) of SARS-CoV-2 represents a sort of mutated form of the RBD of its most closely related virus (i.e., RaTG13). Since RaTG13 is characteristic of a bat species (i.e., *Rhinolophus affinis*), it was suggested that SARS-CoV-2 originated from bat and that after mutating, it acquired ability to attack other animals. Interestingly, mutations may increase the affinity of the RBD for ACE-2 receptor in humans and other animals [255].

6.2. NMR investigations

This paragraph deals with NMR studies of MERS-CoV, SARS-CoV and SARS-CoV-2 proteins and RNA regions playing different roles during the course of viral life cycle.

6.2.1. Structural and Interaction studies

At the best of our knowledge, literature search seems to point out that only a limited number of NMR structural studies on MERS-CoV have been conducted thus far. The non-structural protein 3 of MERS-CoV contains a highly conserved N-terminal region with a macrodomain. This macrodomain interacts with ADP-ribose and other derivatives. To better understand structure-function relationship and get eventually insights in the protein possible roles, preliminary solution NMR structural studies were carried out [255]. MERS-CoV macrodomain (residues 1110-1274) was expressed as recombinant protein in *E. coli* in the ^{15}N, ^{13}C double labeled form. The 2D [^{1}H, ^{15}N] HSQC spectrum and a combination of 3D experiments (i.e., HNCACB, CBCA(CO)NH, HNCO and HN(CA)CO) were acquired to achieve almost complete resonance assignments for backbone atoms whereas, another set of 3D experiments (i.e., HCCH-TOCSY, HCCH-COSY, HCC(CO)NH and HBHA(CO)NH) were recorded to get resonance assignments for aliphatic side-chains [270]. Then, the analysis of the secondary chemical shifts highlighted the presence of six helices and seven β-strands in the macrodomain of MERS-CoV [270]. Resonance assignments may be of course employed in future in ligand-screening approaches.

Differently from MERS-CoV, many NMR structural studies related to SARS-CoV proteins, including the S2 subunit of S protein, E protein, N protein and nonstructural proteins, are available along with conformational analyses on frameshifting pseudoknot from viral mRNA and UTR region from viral genome [271-281].

NMR techniques were employed to study peptide regions of the S2 subunit involved in the interaction with host cell membrane and in the fusion step of viral life cycle [271, 272, 282]. Indeed SARS-CoV fusion protein or the S2 subunit represents a class I type fusion system that includes two heptad repeats, HR1 and HR2, along with a C-terminal TM domain. A six helix bundle topology or trimer-of-hairpins structure is stabilized by the HR1/HR2 complex. Several S2 peptide segments located mainly in the upstream portion of HR1 have been identified as provided with fusogenic properties [271, 272].

In a first study, NMR was exploited to determine the structures of three peptides denoted as FP (N-terminal Fusion Peptide), encompassing the 770-788 region of S2, PTM (Pre-TransMembrane), encompassing the 1185-1202 portion of S2 located near the transmembrane region, and IFP (Internal Fusion Peptide), consisting in the 873-888 region of S2 that is positioned between the other two peptide portions [271]. DPC micelles were used to mimic a membrane-like environment and different NMR experiments were recorded. For instance, 2D TOCSY and NOESY spectra were analyzed to achieve almost complete proton resonance assignments for the three peptides whereas, $^{13}C\alpha$

chemical shifts were gained by natural abundance 2D [^{1}H,^{13}C] HSQC spectra [271]. Then, the solution structures of the peptides were calculated in DPC starting from distance restraints, obtained by NOE contacts, and angular backbone dihedral angles ϕ and ψ restraints, generated by the Preditor webserver [283] starting from ^{1}Hα and ^{13}Cα chemical shifts [271]. Results pointed out that the FP peptide (sequence: 1-MWKTPTLKYFGGFNFSQIL-19) possesses a "V-shaped" helical conformation with a bend in the 10-FGG-12 portion, two hydrophobic patches formed by the side-chains of two sets of residues, one subgroup comprising L7, Y9, F10, and F13 and the other including F15, I18, and L19 and a polar surface formed by the side-chains of T6, N14, and Q17 [271]. Instead, the IFP peptide (sequence: 1-GAALQIPFAMQMAYRF-16) assumes an α-helix conformation which is characterized by a mostly hydrophobic face, composed by five residues (i.e., I6, A9, M10, A13 and F16), and by an opposite face which consists in a combination of non-polar (i.e., L4, F8 and M12) and polar amino acids (i.e., Q11, Y14, and R15) [271]. The PTM peptide (sequence: 1-LGKYEQYIKWPWYVWLGF-18) is characterized by a helix-loop-helix type conformation in which the N-terminal α-helix, formed by a largely polar sequence (i.e., 3-KYEQY-7), is separated from the C-terminal α-helix, by a loop including the 8-IK-9 dipeptide segment [271]. The C-terminal α-helix is composed by two hydrophobic faces, one formed by the sidechains of W10, V14 and F18 and the other by I8, W12, L16 and W15. In addition, PRE (Paramagnetic Relaxation Enhancement) studies were conducted to verify localization and insertion in micelles by acquiring TOCSY spectra of the three peptides in the absence or in the presence of paramagnetic lipids made up by 5-DSA (Doxyl Stearic Acid) and 16-DSA and the results highlighted how PTM positions on the surface of lipids while FP and IFP localize within the micelles, thus providing preliminary information on the fusion mechanism [271, 272]. In fact based on the structural features of the peptides FP, IFP and PTM, it was speculated a potential mechanism of membrane fusion between SARS-CoV and host cells. Following binding of the S1 subunit of the spike protein to its receptor, the FP and IFP regions of the S2 subunit penetrate in the host cells inducing HR1 and HR2 to form a six helix bundle. The PTM peptide with its aromatic residues along with the TM domain that are positioned on the viral membrane form an extended helical hydrophobic arrangement with the HR domains and FP, IFP peptides thus enhancing fusion between viral and host membranes [271].

In a second work, the attention was focused on a 64-mer LFP (Long Fusion Peptide) construct encompassing residues R758-E821 of the S2 subunit of the spike protein of SARS-CoV and including the FP peptide region (i.e., 13-MYKTPTLKYFGGFNFSQIL-31) [272]. NMR characterization was achieved in presence of a membrane-like background made up of DPC micelles; 3D NMR experiments were first acquired to get resonance assignments. Next, the solution structure of the peptide was calculated employing distance and dihedral angle (i.e., ϕ and ψ) constraints. The solution structure of LFP is composed of an amphipathic α-helix encompassing the region preceding the FP portion (i.e., 4-REVFAQVKQ-12) and including a hydrophilic face, made up by the side-chains of a group of residues (i.e., R4, E5, Q9 and Q12), and by a hydrophobic face, composed by packing of the side chains of a second set of residues (i.e., T3, V6, F7, V10) [272]. Instead, a relative longer α-helix is formed by the residues located at the C-terminal end (i.e., 43-IEDLLFNKVTLADAGFMKQY-62) and characterized by a kink at the D55 residue [272]. In addition, the region between the two α-helical segments is composed by a long loop including residues from M13 to F25, a single turn of helix consisting of residues from F27 to I30, and an extended conformation in the prolines rich sequence 32-PSPLKP-37 [272]. Interestingly, ^{15}N-^{1}H heteronuclear NOE measurements showed for the LFP peptide in presence of DPC micelles a dynamical motion occurring in the ps-ns timescale involving the helical regions [272]. Nevertheless, PRE studies were carried out by comparing 2D [^{1}H, ^{15}N] HSQC spectra of LFP alone and in presence of spin-labeled lipids (i.e., 5-DSA and 16-DSA) and the results indicated a deeper insertion of the LFP middle region in the hydrophobic core of the micelle and a low insertion into micelles of the amino acids belonging to the N- and C-terminal ends [272]. As it has been demonstrated that FPs (Fusion Peptides) from SARS-CoV can block viral fusion possibly by binding to membranes, the NMR structural information for the long FP construct could indeed support the design of antiviral agents against SARS.

In analogy with other viral fusion proteins belonging to class 1, the HR1 and HR2 repeat regions of SARS-CoV spike protein are believed to undergo a conformational change from a pre-fusion to a subsequent post-fusion state, that is necessary for the fusion process to take place [273]. Thus, further solution NMR studies in presence of 30% TFE were conducted on the HR2 domain of S2 subunit (i.e.,1-GSHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLN ESLIDLQELGKYEQYIK-55) to determine its structural features in absence of the HR1 domain and thus in a pre-fusion state [273]. First, a combination of 3D experiments (i.e., HNCO, CBCA(CO)NH and CBCANH) was exploited to get backbone resonance assignments; a second set of NMR experiments (i.e., ^{15}N-edited TOCSY-HSQC, ^{15}N-edited NOESY-HSQC, 3D HCC(CO)NH, and 3D CC(CO)NH) was employed to gain side chains assignments whereas, analysis of the 3D HNHA spectrum led to $^{3}J_{HNH\alpha}$ coupling constant values [273]. Furthermore, the CLEANEX-PM (Phase-Modulated CLEAN chemical EXchange) experiment was recorded to identify exchangeable amide protons and revealed a relatively high solvent exposure and fast exchange for the amide protons of residues belonging to the 7-DVDLGDI-13 and 51-EQYIK-55 regions as well as for K49 (Fig. **10A**) [273, 284]. Indeed NMR structure calculations pointed out the presence of a continuous α-helix covering the segment 17-NASVVNIQKEIDRLNEVAKNLNESLIDLQEL-47 in the isolate HR2 monomer (Fig. **10A**) [273]. Instead, additional experiments (i.e., 3D F1-filtered F2-edited [^{1}H, ^{13}C] NOESY-HSQC) were recorded with a sample containing 50% ^{14}N/^{12}C-labeled and 50% ^{15}N/^{13}C-labeled monomers, to determine intermolecular NOE contacts between HR2 units

[273]. The resulting data pointed out a coiled coil structure which consisted in a bundle of three α-helices characterized by inter-molecular contacts involving different hydrophobic residues (i.e., V20, I23, I27, L30, A34, L37, L41 and L44) (Fig. **10A**) [273]. Interestingly, residues V21 and I42, that are highly conserved among coronaviruses, are surface exposed in this pre-fusion state of HR2 and are also the residues involved in interactions with HR1 in the post-fusion state. This study further corroborates the importance of these two residues in stabilizing a six bundle helical arrangement in the post fusion state [273].

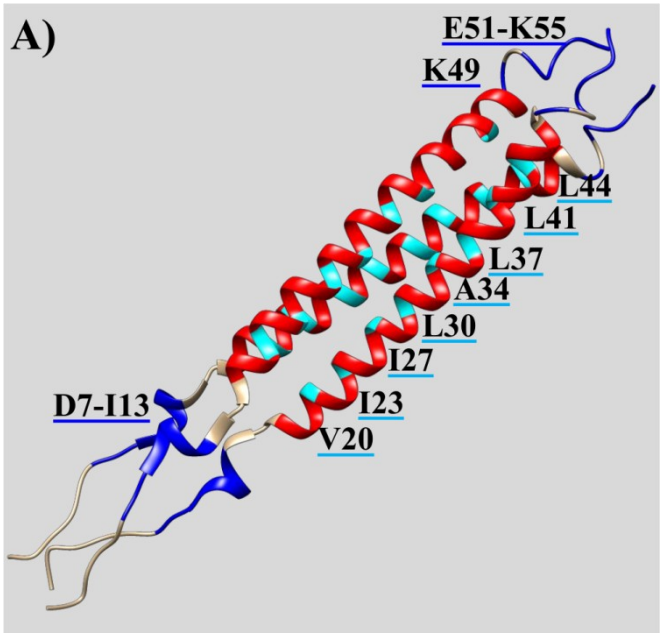


Fig. (**10**). Structural details related to different SARS-CoV and SARS-CoV-2 proteins. (**A**) NMR solution structure of the trimer formed by the HR2 Domain of S2 subunit from the SARS-CoV S protein (first conformer, PDB code: 2FXP [273]). The α-helix of each monomer (i.e., N17-L47) is reported in red. Cyan is used to indicate the residues involved in the intermolecular hydrophobic interactions (i.e., V20, I23, I27, L30, A34, L37, L41 and L44). Solvent-exposed residues are reported in blue (i.e., D7-I13, E51-K55 and K49).

Coronavirus envelope (CoV E) proteins are made up by roughly 100-residue polypeptides and contain at least one α-helical TM domain with channel-forming properties. The extra-membrane C-terminal ends include a conserved proline, positioned in the middle of a predicted β-coil-β motif that is important to target the Golgi-complex [282]. Solution NMR studies were performed for the entire E protein of SARS-CoV (E_{FL}) and a shorter version encompassing its TM domain (E_{TR} residues 8-65) in the membrane mimetic environment created by SDS, that produced high quality NMR spectra [282]. Comparison of [^{1}H-^{15}N]-TROSY-HSQC spectra acquired in water and in 99% D_2O pointed out a low solvent exposure for residues from L18 to L39 in both E_{FL} and E_{TR} thus letting speculate that both peptides possess only one single TM region. Further PRE experiments were conducted with E_{TR} in presence of 5-DSA and 16-DSA and indicated that residues within the peptide segments 11-20 and 40-55 were located near the micelles surfaces whereas, residues 19-40, and partially the portion 48-61 were inserted in the micelles; similarly such studies highlighted that the region 55-65 could be engaged to the surface of the micelles. Heteronuclear ^{1}H-^{15}N NOE measurements pointed out that E_{TR} has a well folded conformation with more dynamic regions encompassing the N-terminus and 46-55 fragment that links the TM domain to the membrane interacting region [282]. Resonance assignments for E_{TR} were obtained by combined inspection of 2D and 3D spectra while it was not possible to conduct a complete structure calculation for E_{FL} due to the low expression yields. In detail, the structure of E_{TR} was calculated in presence of SDS on the basis of the distance restrains obtained by a combination of ^{15}N-resolved NOESY-HSQC, ^{13}C-resolved NOESY-HSQC and PRE studies and employing backbone dihedral angle restraints obtained by TALOS+ [285] starting from the chemical shifts of ^{13}C', ^{13}Cα, ^{13}Cβ, ^{1}Hα atoms. NMR data revealed the presence of a long α-helix in the region 15-NSVLLFLAFVVFLLVTLAILTALRLAAYAAN-45, that contains the TM domain of E_{TR} and includes a slight bend (peptide segment 26-FLLVT-30), a shorter α-helix in the 55-TVYVYSRVKNL-65 segment which encompasses the extra-membrane region of E_{TR}, and a flexible linker covering 46-IVNVSLVKP-54, that is located between the two α-helices (Fig. **10B**) [282]. In summary this work indicated that the predicted β-loop-β motif forms indeed a membrane-bound short α-helix suggesting a possible dynamical exchange between multiple conformations. Comparison of ^{13}Cα chemical shift in SDS indicated almost identical values for E_{FL} and E_{TR} in the overlapping peptide region, while additional C-terminal residues in E_{FL} have canonical chemical shifts of a random coil specie.

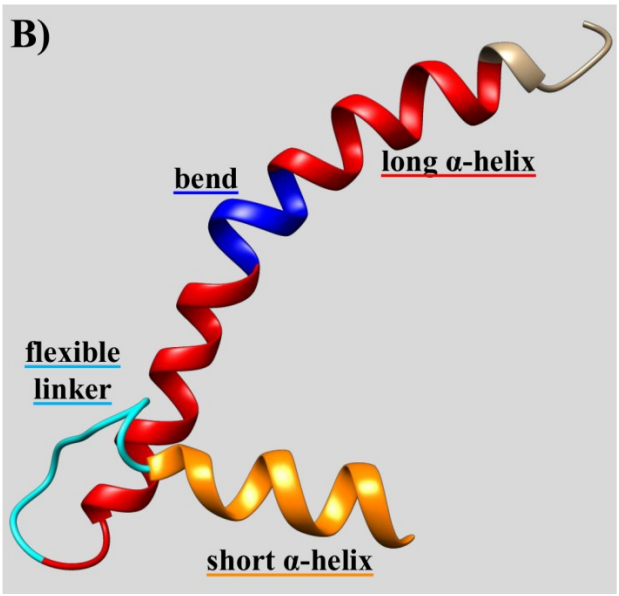


Fig. (**10**). (**B**) NMR solution structure of a truncated version of E protein from SARS-CoV (E_{TR} first conformer, PDB code: 2MM4 [282]). The long α-helix (i.e., N15-N45) is reported in red, the slight bend (i.e., F26-T30) in blue, the flexible linker (i.e., I46-P54) in cyan and the short α-helix (i.e., T55-L65) in orange.

A similar work was conducted by using LMPG (Lyso-Myristoyl PhosphatidylGlycerol) micelles to reproduce a membrane-like environment [274]. The obtained data revealed for E_{TR} monomer the presence of an N-terminal helical region encompassing the TM domain (12-LIVNSVLLFLAFVVFLLVTLAILTAL-37, α1 in Fig. **10C**), an intermediate region with a second α-helix (39-LAAYAANIV-47, α2 in Fig. **10C**) and a third α-helix towards the C-terminus (i.e., 52-VKPTVYVYSRVKNL-65, α3 in Fig. **10C**) [274]. Compared to the results in SDS (Fig. **10B**) [282], in LMPG α3 is 3 residues longer at its N-terminal end, and a new helical segment, α2, is present (Fig. **10C**). By observing migration of the protein by native gel electrophoresis at different protein-lipid ratios a pentameric arrangement was speculated. So, a structural model of oligomeric E_{TR} was built by employing 10 intermolecular NOE contacts collected by 3D NMR ^{15}N-resolved NOESY-HSQC experiments recorded for two different asymmetrically deuterated samples: a) ^{15}N/^{2}H-labeled E_{TR}, and b) an equimolar mixture of ^{15}N/^{2}H-labeled and a non-deuterated ^{13}C-labeled E_{TR} [274]. In detail the structure was built with Haddock [286] starting from the monomeric units. This structure consists in an α-helical bundle characterized by a right-hand sense of screwing (see the A-B-C-D-E order in Fig. **10C**). In the oligomeric topology, each subunit presents a better defined structure if compared to that of the single E_{TR} monomers [274].

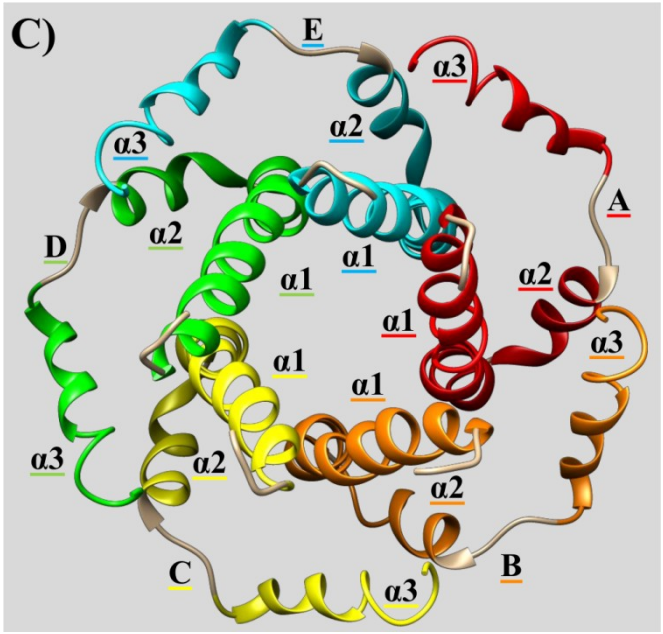


Fig. (**10**). (**C**) NMR solution structure of the pentameric state of E protein from SARS-CoV (first conformer, PDB code: 5X29 [274]). The three helices of each monomer (i.e., α1 from L12 to L37, α2 from L39 to V47 and α3 from V52 to L65) are reported in red, orange, yellow, light green and cyan for the monomers A, B, C, D and E, respectively.

The interaction between membranes and E protein was investigated in another work focused on the structural details of the 9-mer peptide TK9, including a portion of the E protein C-terminal end (i.e., 55-TVYVYSRVK-63) in aqueous solution, in presence of zwitterionic model membrane micelles (consisting of DPC) or in presence of negatively charged model membrane micelles (consisting of SDS) [275]. A set of 2D NMR experiments (TOCSY, NOESY and [^{1}H,^{13}C] HSQC) were acquired to get resonance assignments including those of the ^{13}CαH and determine the amino acids involved in interaction with DPC or SDS micelles [275]. The structures of the peptide in aqueous solution, in DPC or SDS micelles were calculated through distance restraints obtained from NOESY spectra and torsional angle restraints generated from Hα and ^{13}Cα chemical shift values with the PREDITOR webserver [275]. These NMR analyses indicated for the peptide in aqueous solution a mostly flexible structure without a well-defined conformation and for the peptide in both DPC or SDS micelles an α-helical conformation that is stabilized by a set of hydrophobic and hydrophilic interactions and is more favored in DPC micelles. This result points out a higher specificity of TK9 peptide towards eukaryotic cell membrane (that are mainly made up of zwitterionic lipids) than bacterial anionic cell membranes. In addition, in the structure of TK9 in presence of DPC residues V2 and V4 are close to Y5 and Y3, respectively and consequently involved in CH_3-π (aromatic) interactions. Instead, in the structure of TK9 in SDS residues Y5 and Y3 are positioned near K9 and V2, respectively and consequently involved in cation-π and CH_3-π interactions in between each other [275]. 1D R2 (transverse relaxation rate) and R1 (longitudinal relaxation rates) relaxation experiments were conducted for the peptide in absence and presence of micelles and the resulting data highlighted higher ΔR2 values in SDS micelles respect to DPC micelles possibly reflecting a different binding propensity of TK9 towards the two types of micelles formed by DPC respect to SDS. Furthermore, PRE studies related to the peptide in presence of the paramagnetic probe 16-DSA indicated in DPC that the hydrophobic and/or aromatic residues in the N-terminal regions of TK9 encompassing the segment T1-Y5 inserts deeply into the DPC micelle whereas, the positively charged residues in the C-terminal part S6-K9 stay close to the zwitterionic head groups of the micelles. On the contrary, in SDS only an association with the micelles head groups on the surface is observed [275].

NMR structural studies were conducted also for the C-terminal dimerization domain of the nucleocapsid (NP) protein of SARS-CoV [276, 277]. The NP protein is crucial for assembly of infectious viral particles and interacts with the viral RNA creating a ribonucleoprotein core. NP protein possesses an N-terminal domain (residues 45–181) and a C-terminal domain (CTD; residues 248–365) edged by

disordered N- and C-terminal ends and separated by a linker sequence. The NTD is the RNA-binding domain, and the CTD is a dimerization domain but it has also ability to interact with nucleic acids with large affinity [276, 277]. The SAIL (stereo-array isotope labeling) technique [287] was implemented to get high quality NMR spectra for the C-terminal dimerization domain of NP and conduct detailed NMR studies [276]. SAIL allows to overcome the size limit issue of solution NMR protein studies by optimal isotope labeling scheme including a stereo-selective substitution of one ^{1}H in CH_2 groups by ^{2}H; substitution of two ^{1}H in CH_3 groups by ^{2}H; prochiral CH_3 groups of Leu and Val are stereo-selectively changed in order to have one $^{12}C(^{2}H)_3$ and another $^{13}C^{1}H(^{2}H)_2$ methyl group; six-membered aromatic rings are labeled to get occurrence of sequential $^{12}C–^{2}H$ and $^{13}C–^{1}H$ groups [287]. The NMR solution structure of the CTD domain of SARS-CoV NP protein consists of a domain-swapped homodimer conformation; a disordered region (i.e., 248-TKKSAAEASKKP-259) sticks out from the core of the homodimer. In addition, the fold of the single monomer includes different α-helical segments and two β-strands forming an antiparallel β-sheet (Fig. **10D**) [276]. CSP (Chemical Shift Perturbation) studies were further conducted to analyze the interaction between nucleic acids and the C-terminal dimerization domain of the SARS-CoV NP protein by employing two different poly-dT (poly-deoxythymine) ssDNA sequences, one composed by 10 deoxynucleotides (dT10) and another by 20 deoxynucleotides (dT20), to mimic the ssRNA characterizing the viral genome. CSP data, combined with the results of mutational analyses, revealed the involvement of the N-terminal disordered region (i.e., K250, E253, A254, S256, K257, and K258) and the β-sheet (i.e., R320, H335, and A337) in the interaction of CTD homodimer with ssDNA (Fig. **10D**) [276]. These structural information can be implemented to set up structure-based drug discovery approaches to design inhibitors of CTD mediated interactions as novel antiviral agents interfering with SARS-CoV assembly.

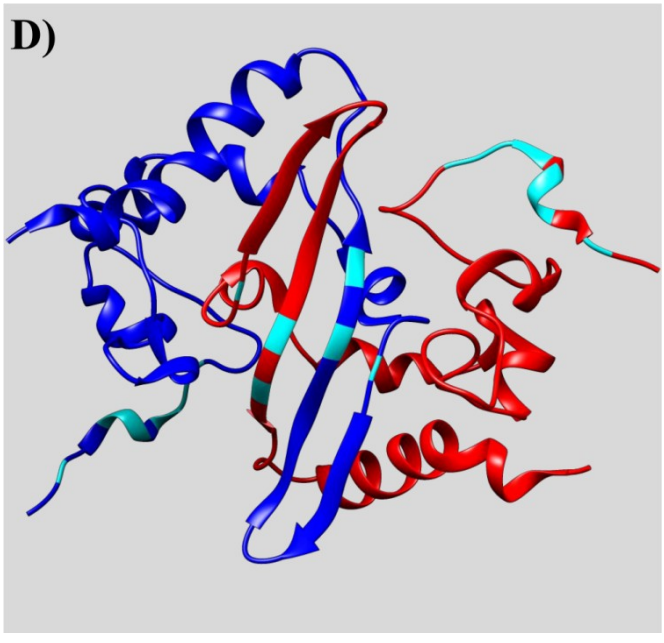


Fig. (**10**). (**D**) NMR solution structure of the C-terminal dimerization domain of the SARS-CoV NP protein (first conformer, PDB code: 2JW8 [276]). The two protein units composing the homodimer are colored blue and red, respectively. Helical regions encompass segments S251-S256, N270-F275, D291-G296, T297-Y299, H301-Q307, S311-F316, Q346-D359 and A360-F364, the β1 strand includes the G322-T326 region and β2 the G329-K339 segment, the disorder region includes the T248-P259 segment. Residues involved in the interaction with ssDNA are colored cyan (i.e., K250, E253, A254, S256, K257, K258, R320, H335, and A337) [276].

Solution NMR techniques were implemented as well to study fragments of viral RNA genome such as the region responsible for the -1 PRF (Programmed -1 Ribosomal Frameshifting) that is a process exploited by the virus to produce fusion proteins. Even small variations in frameshift frequencies can have a tremendous negative outcome on virus propagation, thus targeting -1 PRF is considered a drug discovery strategy. SARS-CoV -1 PRF signal includes a heptameric slippery site'' (i.e., UUUAAAC), a 5-nt spacer, and an H-form mRNA pseudoknot [278]. A combination of NMR experiments (i.e., [^{1}H, ^{1}H] NOESY, ^{15}N-HMQC and quantitative J(N,N) HNN-COSY) led to establish a complete sequential NOE path linking the majority of base paired imino protons; the NOE contacts confirmed that the SARS-CoV mRNA pseudoknot includes three double-stranded RNA stem structures where interactions in between the stems produce indeed distortions of the A-form RNA [278]. Further NMR studies of the mRNA pseudoknot from SARS-CoV revealed a dimerization mechanism involving a palindromic sequence belonging to the loop (i.e., 5'-ACUAGU-3', named "L2") capping the third stem of the mRNA pseudoknot [279].

The NMR solution structure of a portion of NSP1 (NonStructural Protein 1) replicase (i.e., 13–128) has also been reported by canonical analysis of 2D and 3D NMR experiments [288, 289]. NSP1 is a member of the SARS-CoV replicase polyprotein that regulates RNA replication and processing [288, 289]. NMR studies revealed indeed a fold including six β-strands, one α-helix and a 3_{10} helix [289]. Interestingly, these secondary structure elements organize themselves and provide a mixed parallel/antiparallel six-stranded β-barrel in which one opening is covered by α1, that appears like a lid whereas, the 3_{10} helix runs on one side of the barrel [289].

The SARS-CoV NSP3 is a multi-domain protein composed by at least seven modules (i.e., NSP3A, NSP3B, NSP3C, NSP3D, NSP3E, NSP3F and NSP3G). NSP3A (i.e., 1-183 region) is the domain responsible for the binding to ssRNA, NSP3B (i.e., 184-351 portion) is the module characterized by ADP-ribose-1''-phosphatase activity, and NSP3D (i.e., 723-1037 region) is a papain-like protease responsible for the cleavage of the polyproteins pp1a and pp1ab [280]. An interesting work was focused on the middle region of NSP3C domain, termed SUD ("SARS-Unique Domain")-M (residues 513-651) and on a shorter version (i.e., SUD-M(527-651)); nucleic acid binding propensity mediated by the C-terminal region that interestingly is also

conserved among coronaviruses, has been attributed to the SUD [280]. Resonance assignments for both the long and shorter protein constructs were obtained by 2D, 3D and 4D-APSY (Automated Projection Spectroscopy) NMR experiments and the NMR structures were calculated [280, 290]. Results showed in the 528-648 region the presence of a globular domain characterized by six β-strands (i.e., β1 for 528-530 region, β2 for 547-550 portion, β3 for 572-574 region, β4 for 580-583 portion, β5 for 603-605 region and β6 for 631-634 portion), five α-helices (i.e., α1 for 533-543 region, α2 for 554-563 portion, α3 for 589-599 region, α4 for 617-624 segment and α5 for 638-649 fragment) and one 3_{10} helix in 611-613 region which are arranged in the β1-α1-β2-α2-β3-β4-α3-β5-3_{10}-α4-β6-α5 fold [280]. In addition, the structures of both the long and short constructs have two well-defined long loops without regular secondary structures in the regions between α2 and β3 (residues 564-571) and between β5 and α4 (residues 606-616) whereas, the structure of SUD-M(513-651) includes also a flexible extended tail at the N-terminus (i.e., 513-527 region) and a flexible tail at the C-terminus (i.e., 649-651 region) [280]. To verify ability to bind nucleic acids, CSP studies were conducted by looking for differences between 2D [^{15}N,^{1}H]-HSQC spectra of SUD-M(527-651) alone and in presence of ssRNA (homodecamers of uridin (poly(U_{10})), guanosine (poly(G_{10})) and adenosine (poly(A_{10})) or ADP-ribose that was tested as well based on the structural homology between SUD-M and several NTPases, including a high structural similarity to SARS-CoV NSP3b, which is also able to interact with ADP-ribose. These interaction studies highlighted binding of SUD-M(527-651) to poly(A_{10}) and major perturbations for residues G527 from the N-terminal flexible extended tail, W531 to L533 between β1 and α1, I556-Q561 region from α2 and V611 from the 3_{10} helix, thus allowing to identify a binding pocket [280].

The nonstructural proteins NSP7-NSP10 are conserved among all CoVs and are translated as part of the pp1a and pp1ab viral polyproteins, and then, through action of the NSP5 protease are released in the form of mature proteins. The SARS-CoV NSP7 protein (NonStructural Protein7) plays a function in the transcription and replication of the viral genome and has been object of NMR structural studies at pH 6.5 [281] and at pH 7.5 [291]. The structure of NSP7 calculated at pH 6.5 consists in an antiparallel four-helix bundle characterized by a disordered tail in the N-terminal region (i.e., 1-10), a small and extended portion (i.e., 11-12), α1-helix (i.e., 13-20), a well-defined loop (i.e., 21-28), α2-helix (i.e., 29-42), a small region devoid of regular secondary structure (i.e., 43-46), α3-helix (i.e., 47–65), a short loop (i.e., 66-69) and α4-helix (i.e., 70-82) [281]. Comparison with the structure of NSP7 obtained at pH 7.5 [291] revealed still the presence of a fold with four helices but highlighted as well differences in the lengths and sequence positioning of the helices and variations in the tertiary folds.

Thus far at the best of our knowledge, only a reduced number of NMR structural studies for proteins of SARS-CoV-2 has been reported.

The structure of the NTD (N-Terminal Domain) RNA binding module of the N protein from SARS-CoV-2 and its interaction with RNA have been studied by solution NMR techniques [12]. The nucleocapsid phosphoprotein N works to engage the viral genome to the viral membrane. It contains an RNA binding domain at the N-terminal side that interacts with the RNA genome; the C-terminal domain through binding to the M protein links the ribonucleoprotein complex to the viral membrane [12]. After achieving resonance assignments through a canonical protocol relying on analysis of 2D and 3D NMR spectra, recorded for a uniformly ^{15}N, ^{13}C double labeled protein sample (residues 44-180), the structure of the RNA binding domain was calculated by employing interproton distance restraints obtained by 3D [^{15}N, ^{1}H] NOESY-HSQC and 3D [^{13}C, ^{1}H] NOESY-HMQC experiments and backbone torsion angle restraints gained through TALOS+ [285] starting from assigned chemical shifts [12]. The results show an overall fold that includes a core characterized by an antiparallel five-stranded β-sheet platform with a U-shape right-hand appearance. The β-core looks like the palm of an hand (Fig. **10E**) [12]. In addition, the β-strands of this β-sheet platform organize themselves in the sequential order β4-β2-β3-β1-β5 and are flanked by two short α-helices, one preceding β1 (i.e., α1) and the other (i.e., α2) coming after β4 (Fig. **10E**) [12]. The calculated structure also reveals the presence of large bulging loop between β2 and β3 strands that comprises a basic β-hairpin made up by short β2' and β3' strands, forming a structural topology similar to a finger (Fig. **10E**) [12]. Interestingly, 2D [^{1}H, ^{15}N] HSQC and 2D [^{1}H, ^{13}C] HSQC spectra were acquired for the protein alone and in presence of a short RNA duplex made up by the 5'-CACUGAC-3' and 5'-GUCAGUG-3' strands, and compared. The results of these CSP studies indicated that the residues most affected by the interaction are A50, T57, H59, R92, I94, S105, R107, R149 and Y172 and belong to the basic finger or are located in proximity of the region between the basic finger and the palm of the hand (Fig. **10E**) [12]. These CSP data were exploited to get a model of the RNA-NTD complex with Haddock [286].

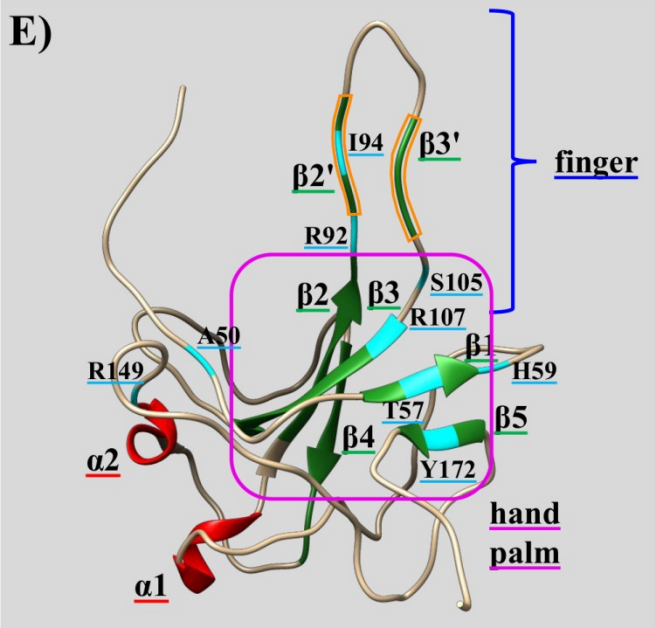


Fig. (**10**). (**E**) NMR solution structure of the NTD (N-Terminal Domain) RNA binding module of the nucleocapsid phosphoprotein from SARS-CoV-2 (first conformer, PDB code: 6YI3 [12]). The β-strands (i.e., β1 for L56-H59 region, β2 for Y86-T91 segment, β2' R93-R95 segment, β3' M101-D103 segment, β3 for R107-Y112 portion, β4 for I130-T135 region and β5 for F171-A173 segment) are reported in green and the α-helices (i.e., α1 for S79-I84 portion and α2 for D144-T148 region) are reported in red. Residues largely affected by interaction with RNA are colored cyan (i.e., A50, T57, H59, R92, I94, S105, R107, R149 and Y172).

6.2.2. Other NMR studies

An interesting work was conducted to discover antiviral agents against MERS-CoV acting as inhibitors of viral replication by targeting its papain-like protease (PLpro) [292]. A screening of 30,000 compounds was carried out based on a fluorescence enzymatic assay [292]. The hits from this first HTS (High Throughput Screening) were confirmed through direct interaction assays by SPR (Surface Plasmon Resonance). The hit "Compound 6" (Fig. **11A**) was further analyzed by NMR spectroscopy thus, 2D [^{1}H, ^{15}N] TROSY (Transverse Relaxation Optimized Spectroscopy) NMR spectra were acquired for uniformly ^{15}N-labeled His-tagged MERS-PLpro alone and in the presence of the inhibitor. NMR CSP data indicated that Compound 6 was able to produce a large number of variations in the spectrum of the protein [292]. In detail 13 peaks were affected by binding of inhibitor 6 to the protein with a few peaks undergoing chemical shifts variations and others completely disappearing. Disappearance of a few peaks could be connected to exchange broadening linked to the compound interacting and dissociating from the protease or to conformational changes affecting the protease backbone upon inhibitor binding [292]. This work highlights how NMR spectroscopy even in absence of protein resonance assignments is able to provide useful information on the interaction mechanism.

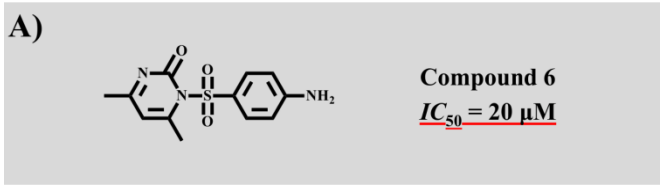


Fig. (**11**). (**A**) Chemical structures of an inhibitor of the Papain-Like protease from MERS-CoV. The IC50 value is relative to inhibition of the MERS-PLpro protease activity [292].

NMR spectroscopy was also employed to analyze peptide inhibitors of viral entry. SARS-CoV utilizes the ACE2 as entry receptor by engaging it through the Receptor Binding Domain (RBD) of its S protein [293]. Thus peptide libraries were generated by including 12-mer peptides (i.e., RBD-1 – RBD-16) derived from the RBD of S protein (i.e., covering all together the N318-T509 region) and screened by SPR against ACE2 to discover RBD interaction epitopes. This first screening delivered three peptide hits (i.e., RBD-11 covering the Y438-R449 region, RBD-14 the S474-T485 portion and RBD-15 the T486-V497 segment). These hits were used to design a second library of 6-mer peptides (i.e., peptides RBD-11a to RBD-11e, RBD-14a to RBD-14d and RBD-15a to RBD-15e), containing an overlap of three residues within each sequence, that were employed in a second round of SPR screening [293]. The RBD-11b (corresponding to 438-YKYRYL-443 RBD region) (Fig. **11B**) gave the largest SPR response and was further analyzed by deepest SPR studies and STD NMR [293]. SPR indicated a K_D equal to 46 μM for the binding to ACE2; STD data also allowed to estimate a K_D value of 60 ± 25 μM. In addition STD NMR indicated the proximity of all tyrosine residues in the peptide to the ACE2 receptor, strong interactions between Hα of three residues (i.e., Y440, R441, and L443) and the receptor, interactions of moderate strength between positively charged side-chains of lysines and arginines and the receptor surface, hydrophobic interactions involving the side-chain of the leucine at the C-terminus [293].

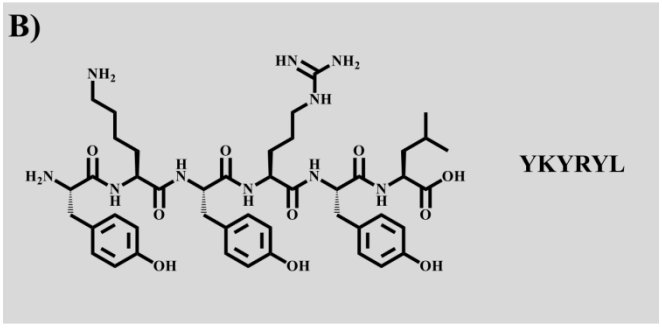


Fig. (**11**). (**B**) Peptide inhibitor of SARS-CoV replication. The hexapeptide "Tyr-Lys-Tyr-Arg-Tyr-Leu" is specific for coronaviruses. It inhibits *in vitro* the SARS-CoV infection of Vero E6 cells. Nevertheless, the same peptide blocks proliferation of coronavirus NL63 but does not inhibit infection of Vero cells attacked by alpha virus Sindbis [293].

An intriguing work further highlights the relevance of NMR technique to establish binding epitopes and thus gain structural information for the design of novel antiviral compounds [294]. Indeed, the study focuses on DC-SIGN (Dendritic Cell-Specific ICAM-3 Grabbing Non-integrin) representing a lectin characterized by a C-terminal CRD (Carbohydrate Recognition Domain) which can bind the highly glycosylated proteins found on the surface of many pathogens including SARS-CoV [294]. DC-SIGN works indeed as a kind of universal pathogen receptor expressed on the surface of immature dendritic cells and is an attractive target in drug-discovery. Binding of mannosyl di- and trisaccharides to the CRD of DC-SIGN was investigated by NMR studies through STD and transferred NOESY experiments. NMR data highlight the fundamental role of the terminal mannoses at positions 2 (for disaccharides) or 3 (for trisaccharides) that establish the strongest interaction with the binding site of DC-SIGN [294]. This information could be exploited for the development of molecules able to block the interaction between the C-terminal CRD of DC-SIGN and highly glycosylated proteins of SARS-CoV and other pathogens [294].

In addition NMR was indeed employed similarly to what described in previous paragraphs related to other viruses, as analytical tool for the characterization of newly synthesized inhibitors including those targeting the MERS-CoV helicase [295]. Sixteen halogenated triazole compounds were synthetized, these molecules were analyzed by 1D [^{1}H] and 1D [^{13}C] NMR to verify their identities and by a FRET assay to evaluate their interference with MERS-CoV helicase activity. This study led to the identification of two potent helicase inhibitors: Compound 12 and Compound 16 (Fig. **11C**) [295].

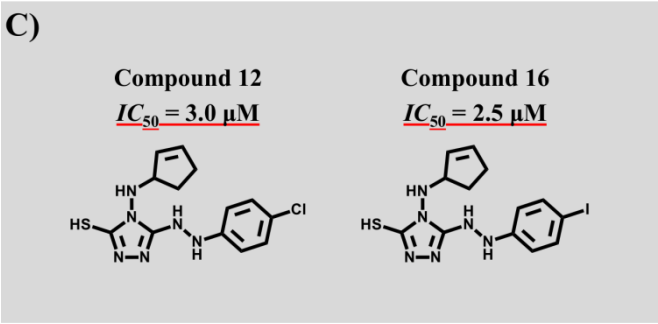


Fig. (**11**). (**C**) Inhibitors of MERS-CoV helicase [295]. Antiviral properties were tested through a helicase specific assay by employing a purified M-nsp13 cloned helicase domain [295].

Licorice root (Glycyrrhiza radix) contains the triterpene glycoside glycyrrhizic acid (glycyrrhizin, GL) and its aglycone 18β-glycyrrhetinic acid (GLA) that represent its most bioactive compounds. Interestingly GL also works *in vitro* against SARS-CoV replication [296]. With this in mind, 15 GL analogues were synthetized and characterized by several analytical techniques including ^{1}H and ^{13}C NMR and their activities against SARS-CoV were evaluated. Enhanced anti-SARS-CoV activity was achieved by insertion of 2-acetamido-β-D-glucopyranosylamine into the glycoside moiety of GL thus producing a 10-fold improvement with respect to GL. Up to 70 fold enhanced anti-SARS-CoV activity was obtained instead through amide derivatives of GL and with GL derivatives conjugated with two amino acid residues and a free -COOH group although the increase in activity was also accompanied by higher cytotoxicity due to a poorer selectivity [296].

SARS-CoV has an RNA genome encoding for the replicase polyproteins pp1a and pp1b. Functional proteins necessary for viral propagation are generated through proteolytic processing of the nonstructural polyproteins pp1a and pp1b. The principal protease regulating this processing is Mpro, which is also known as 3CLpro (i.e., dimeric chymotrypsin-like protease). The main protease of SARS-CoV-2 related to the Wuhan outbreak of respiratory disease has 96% sequence identity with SARS-CoV Mpro [297]. Much efforts have been directed towards generating inhibitors of this protease as potential anti-Coronaviruses agents [298, 299]. For instance, a varied library of peptide anilides was synthetized, and inhibition power against the SARS-CoV 3CL protease was investigated. Through this approach an anilide derived from 2-chloro-4-nitroaniline, L-phenylalanine and 4-(dimethylamino)benzoic acid was identified as a potent protease inhibitor (Ki 0.03 µM) (Fig. **12A**) [299].

Similarly, isatin derivatives (Fig. **12B**) were synthetized and analyzed as effective inhibitors of 3CL protease of SARS-CoV, this work led to the development of potent compounds also provided in a few cases (compound 4o in Fig. **12B**) with selectivity against SARS-CoV main protease with respect to other enzymes like papain or chymotrypsin and trypsin [298].

Indeed the main protease of coronaviruses and the 3C protease of enteroviruses possess a similar active-site topology and both needs a glutamine in the P1 position of the substrate [297]. This observation was exploited in an original structure-based approach to find protease inhibitors. In details, to generate potent antiviral agents targeting α-coronaviruses, β-coronaviruses, and enteroviruses, peptidomimetic α-ketoamides were designed to target and block the main and 3C proteases. All the α-ketoamides contained a 5-membered ring (γ-lactam) derivative of glutamine as P1 residue [297]. Optimization of the P2 substituent of the α-ketoamides resulted essential for providing near-equipotency towards the three diverse viruses genera. Compounds 11r (with cyclohexylmethyl moiety in P2) and 11u (with cyclopentylmethyl moiety in P2) resulted the best inhibitors possessing low-micromolar EC50 values towards enteroviruses, α-coronaviruses, and β-coronaviruses in cell cultures (Fig. **12C**). Compound (11r) showed also very good anti-MERS-CoV activity in Huh7 cells that were infected by the virus. It is expected that this compound could be active against SARS-CoV-2 main protease due to sequence similarity of the protease from SARS-CoV-2 and SARS-CoV as mentioned before [297].

In the last months, many efforts were spent in the research on SARS-CoV-2 and in the identification of its potential

inhibitors. For instance, NMR was employed to obtain the structure and absolute configuration of different molecules (diketopiperazines) extracted from the solid culture of *Aspergillus versicolor* [300]. HMBC (Heteronuclear Multiple Bond Correlation) and ROESY (Rotating frame Overhauser Effect Spectroscopy) spectra provided large structural information; these compounds were implemented in a virtual screening approach through docking studies against the Mpro protease of SARS-CoV-2. Compounds with the best docking scores could work as antiviral agents blocking protease activity but they need of course experimental validation [300].

^{1}H and ^{13}C NMR spectra were analyzed also to characterize newly synthetized Norcantharimide-based compounds. These molecules represent derivatives of cantharidin, a toxic monoterpene extracted from the dried body of Chinese blister beetle and used in Chinese medicine against different malignant tumors (e.g., hepatoma, breast cancer and colorectal cancer) [301]. Cantharidin was exploited as antiviral agent against MC (Molluscum Contagiosum) infection of the skin and Norcantharidin was suggested as potential *in vitro* inhibitor against HIV [301]. ADME (Absorption, Distribution, Metabolism and Excretion) analysis suggested the potential use of Norcantharimide-based compounds as orally active drugs. Docking studies against the main protease Mpro of SARS-CoV-2 indicated good theorical affinity of these type of compounds for this protease thus suggesting that they could be explored as anti-SARS-CoV-2 agents [301].

In addition, a recent work stressed out how NMR and crystallography could work together to design antiviral agents targeting SARS-CoV-2 protease. A crystallographic screening identified fragments binding to the protease Mpro of SARS-CoV-2 at the active site or at other potential allosteric pockets. STD NMR experiments were next exploited to evaluate the affinities of molecular fragments for the protease [302]. This work shows further how STD-NMR is powerful to rapidly screening fragment libraries against Mpro and identify ligands.

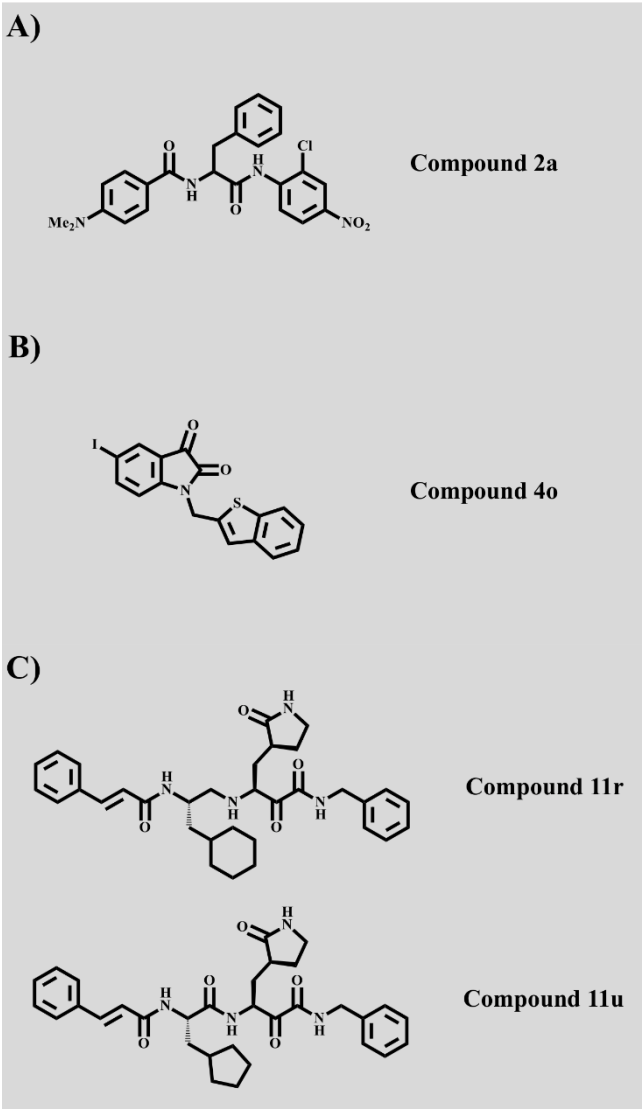


Fig. (**12**). Protease inhibitors. (**A**) Anilide Inhibitor [299] and (**B**) Isantin derivative [298] targeting the 3CL protease of SARS-CoV. (**C**) Peptidomimetic α-ketoamides inhibitors active against α-coronaviruses, β-coronaviruses, and enteroviruses [297].

7. CONCLUSION

This review reports on the principal applications of NMR spectroscopy in the study of human viruses and in the antiviral drug discovery field. It focuses mainly on the following subsets of viruses: HCV, HBV, HSV, HIV, that have been largely studied through NMR techniques. Examples of NMR studies linked to MERS-CoV, SARS-CoV and SARS-CoV-2 are described as well [12, 276, 292].

Intriguingly, solution or solid-state NMR thus far has provided structural information on viral proteins and viral genomes either in isolation or bound to diverse binding partners including nucleic acids, lipids, proteins and compound inhibitors or drugs [63, 124, 303]. Interestingly, NMR is a pivotal technique to study dynamic systems

undergoing conformational movements and has proved also powerful to discover and analyze structural disorder in viral proteins [79]. NMR-based information has been employed to set up structure-based approaches to design efficient antiviral agents inhibiting crucial steps of viral life cycle including but not limited to viral entry, fusion, and replication. Indeed, cleaver NMR screening assays have been set up to find out ligands of essential viral proteins and genomes [89, 229].

NMR-STD has as well proved useful to determine epitope mapping for glycans or peptides binding to antibodies thus paving the way for the design of novel vaccines [227, 228].

In addition, NMR-based metabolomics approaches have allowed analysis of patients affected by viral infections and identification of metabolites in biological fluids to be considered biomarkers of the disease. NMR-metabolomics has revealed also changes in metabolic profiles associated with antiviral therapy that could be in future exploited to generate novel diagnostic tools as well as suggest ways to follow progression of viral infection and/or improve infection treatments [127, 128, 216].

NMR spectroscopy as an analytical instrument during the syntheses of antiviral agents has played a crucial role to assess compounds purity and identity [177, 298, 299].

Due to the versatility of NMR technique we can envision that in the next few years much more efforts will be devoted in setting up NMR-based drug discovery campaigns targeting SARS-CoV-2.

*Address correspondence to Marilisa Leone at the Institute of Biostructures and Bioimaging, National Research Council of Italy, Via Mezzocannone 16, 80134, Naples, Italy; Tel: +39-081-253-4512; E-mail: marilisa.leone@cnr.it

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

ACKNOWLEDGEMENTS

Fondazione Umberto Veronesi is acknowledged for a grant to Marian Vincenzi (FUV Post-Doctoral fellowship 2019-2020).

REFERENCES

[1] Holmes, E.C. Evolutionary history and phylogeography of human viruses. *Annu. Rev. Microbiol.,* **2008**, *62*, 307-328.

[2] Woolhouse, M.; Scott, F.; Hudson, Z.; Howey, R.; Chase-Topping, M. Human viruses: discovery and emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.,* **2012**, *367*(1604), 2864-2871.

[3] Parvez, M.K.; Parveen, S. Evolution and Emergence of Pathogenic Viruses: Past, Present, and Future. *Intervirology,* **2017**, *60*(1-2), 1-7.

[4] Xue, B.; Blocquel, D.; Habchi, J.; Uversky, A.V.; Kurgan, L.; Uversky, V.N.; Longhi, S. Structural disorder in viral proteins. *Chem. Rev.,* **2014**, *114*(13), 6880-6911.

[5] ViralZone. https://viralzone.expasy.org/295

[6] Tong, L. Viral proteases. *Chem. Rev.,* **2002**, *102*(12), 4609-4626.

[7] World Health Organization. Coronavirus disease (COVID-2019) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[8] Hemida, M.G.; Chu, D.K.; Poon, L.L.; Perera, R.A.; Alhammadi, M.A.; Ng, H.Y.; Siu, L.Y.; Guan, Y.; Alnaeem, A.; Peiris, M. MERS coronavirus in dromedary camel herd, Saudi Arabia. *Emerg. Infect. Dis.,* **2014**, *20*(7), 1231-1234.

[9] Yang, Y.; Peng, F.; Wang, R.; Guan, K.; Jiang, T.; Xu, G.; Sun, J.; Chang, C. The deadly coronaviruses: The 2003 SARS pandemic and the 2020 novel coronavirus epidemic in China. *J. Autoimmun.,* **2020**, *109*, 102434.

[10] Peeri, N.C.; Shrestha, N.; Rahman, M.S.; Zaki, R.; Tan, Z.; Bibi, S.; Baghbanzadeh, M.; Aghamohammadi, N.; Zhang, W.; Haque, U. The SARS, MERS and novel coronavirus (COVID-19) epidemics, the newest and biggest global health threats: what lessons have we learned? *Int. J. Epidemiol.,* **2020**, 1-10.

[11] Malik, Y.S.; Sircar, S.; Bhat, S.; Sharun, K.; Dhama, K.; Dadar, M.; Tiwari, R.; Chaicumpa, W. Emerging novel coronavirus (2019-nCoV)-current scenario, evolutionary perspective based on genome analysis and recent developments. *Vet. Q.,* **2020**, *40*(1), 68-76.

[12] Dinesh, D.C.; Chalupska, D.; Silhan, J.; Veverka, V.; Boura, E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *bioRxiv,* **2020**.

[13] De Clercq, E.; Li, G. Approved Antiviral Drugs over the Past 50 Years. *Clin. Microbiol. Rev.,* **2016**, *29*(3), 695-747.

[14] da Costa, V.G.; Moreli, M.L.; Saivish, M.V. The emergence of SARS, MERS and novel SARS-2 coronaviruses in the 21st century. *Arch. Virol.,* **2020**, 1517-1526.

[15] Le Page, M. Viruses from animals. *New Sci.,* **2020**, *245*(3268), 10.

[16] Tortorici, M.A.; Veesler, D. Structural insights into coronavirus entry In: *Complementary Strategies to Understand Virus Structure and Function*. Rey, F.A., Ed., **2019**; Vol. *105*, pp 93-116.

[17] Zhou, Y.; Agudelo, J.; Lu, K.; Goetz, D.H.; Hansell, E.; Chen, Y.T.; Roush, W.R.; McKerrow, J.; Craik, C.S.; Amberg, S.M.; Simmons, G. Inhibitors of SARS-CoV entry - Identification using an internally-controlled dual envelope pseudovirion assay. *Antiviral Res.,* **2011**, *92*(2), 187-194.

[18] Ghongade, K.D.; Biluri, L.A.; Mali, A.S. A review on current approaches of the development of antiviral drugs. *Int. J. Univers. Pharm. Bio Sci.,* **2018**, *7*(1), 1-21.

[19] Jones, P.S. Strategies for antiviral drug discovery. *Antivir. Chem. Chemother.,* **1998**, *9*(4), 283-302.

[20] Tyndall, J.D.; Nall, T.; Fairlie, D.P. Proteases universally recognize beta strands in their active sites. *Chem. Rev.,* **2005**, *105*(3), 973-999.

[21] Madala, P.K.; Tyndall, J.D.; Nall, T.; Fairlie, D.P. Update 1 of: Proteases universally recognize beta strands in their active sites. *Chem. Rev.,* **2010**, *110*(6), PR1-31.

[22] Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G.F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U.S.A.,* **2003**, *100*(23), 13190-13195.

[23] Yao, N.; Hong, Z.; Lau, J.Y. Application of structural biology tools in the study of viral hepatitis and the design of antiviral therapy. *Gastroenterology,* **2002**, *123*(4), 1350-1363.

[24] Scott, C.; Griffin, S. Viroporins: structure, function and potential as antiviral targets. *J. Gen. Virol.,* **2015**, *96*(8), 2000-2027.

[25] Wang, S.; Fogeron, M.L.; Schledorn, M.; Dujardin, M.; Penzel, S.; Burdette, D.; Berke, J.M.; Nassal, M.; Lecoq, L.; Meier, B.H.; Bockmann, A. Combining Cell-Free Protein Synthesis and NMR Into a Tool to Study Capsid Assembly Modulation. *Front. Mol. Biosci.,* **2019**, *6*, 67.

[26] Fu, D.-Y.; Xue, Y.-R.; Yuc, X.; Wu, Y. Anti-virus reagents targeting the capsid protein assembly. *J. Mater. Chem. B,* **2019**, *7*, 3331-3340.

[27] Gogineni, V.; Schinazi, R.F.; Hamann, M.T. Role of Marine Natural Products in the Genesis of Antiviral Agents. *Chem. Rev.,* **2015**, *115*(18), 9655-9706.

[28] Frick, D.N.; Lam, A.M. Understanding helicases as a means of virus control. *Curr. Pharm. Des.,* **2006**, *12*(11), 1315-1338.

[29] Hermann, T. Small molecules targeting viral RNA. *Wiley Interdiscip. Rev. RNA,* **2016**, *7*(6), 726-743.

[30] Callaway, E. The race for coronavirus vaccines: a graphical guide. *Nature,* **2020**, *580*(7805), 576-577.

[31] Tse, L.V.; Meganck, R.M.; Graham, R.L.; Baric, R.S. The Current and Future State of Vaccines, Antivirals and Gene Therapies Against Emerging Coronaviruses. *Front. Microbiol.,* **2020**, *11*, 658.

[32] Mercorelli, B.; Palu, G.; Loregian, A. Drug Repurposing for Viral Infectious Diseases: How Far Are We? *Trends Microbiol.,* **2018**, *26*(10), 865-876.

[33] Vincenzi, M.; Mercurio, F.A.; Leone, M. NMR spectroscopy in the conformational analyses of peptides: an overview. *Curr. Med. Chem.,* **2020**.

[34] Pellecchia, M.; Bertini, I.; Cowburn, D.; Dalvit, C.; Giralt, E.; Jahnke, W.; James, T.L.; Homans, S.W.; Kessler, H.; Luchinat, C.; Meyer, B.; Oschkinat, H.; Peng, J.; Schwalbe, H.; Siegal, G. Perspectives on NMR in drug discovery: a technique comes of age. *Nat. Rev. Drug Discov.,* **2008**, *7*(9), 738-745.

[35] Pellecchia, M.; Sem, D.S.; Wuthrich, K. NMR in drug discovery. *Nat. Rev. Drug Discov.,* **2002**, *1*(3), 211-219.

[36] Li, Y.; Kang, C. Solution NMR Spectroscopy in Target-Based Drug Discovery. *Molecules,* **2017**, *22*(9), 1399.

[37] Ishima, R. Protein-Inhibitor Interaction Studies Using NMR. *Appl. NMR Spectrosc.,* **2015**, *1*, 143-181.

[38] Leone, M.; Freeze, H.H.; Chan, C.S.; Pellecchia, M. The Nuclear Overhauser Effect in the lead identification process. *Curr. Drug Discov. Technol.,* **2006**, *3*(2), 91-100.

[39] Pellecchia, M.; Becattini, B.; Crowell, K.J.; Fattorusso, R.; Forino, M.; Fragai, M.; Jung, D.; Mustelin, T.; Tautz, L. NMR-based techniques in the hit identification and optimisation processes. *Expert Opin. Ther. Targets,* **2004**, *8*(6), 597-611.

[40] Mayer, M.; Meyer, B. Characterization of Ligand Binding by Saturation Transfer Difference NMR Spectroscopy. *Angew. Chem. Int. Ed. Engl.,* **1999**, *38*(12), 1784-1788.

[41] Mayer, M.; Meyer, B. Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc.,* **2001**, *123*(25), 6108-6117.

[42] Walpole, S.; Monaco, S.; Nepravishta, R.; Angulo, J. STD NMR as a Technique for Ligand Screening and Structural Studies. *Methods Enzymol.,* **2019**, *615*, 423-451.

[43] Benie, A.J.; Moser, R.; Bauml, E.; Blaas, D.; Peters, T. Virus-ligand interactions: identification and characterization of ligand binding by NMR spectroscopy. *J. Am. Chem. Soc.,* **2003**, *125*(1), 14-15.

[44] Raingeval, C.; Cala, O.; Brion, B.; Le Borgne, M.; Hubbard, R.E.; Krimm, I. 1D NMR WaterLOGSY as an efficient method for fragment-based lead discovery. *J. Enzyme Inhib. Med. Chem.,* **2019**, *34*(1), 1218-1225.

[45] Huang, R.; Leung, I.K.H. Protein-Small Molecule Interactions by WaterLOGSY. *Methods Enzymol.,* **2019**, *615*, 477-500.

[46] Singh, M.; Tam, B.; Akabayov, B. NMR-Fragment Based Virtual Screening: A Brief Overview. *Molecules,* **2018**, *23*(2), 233.

[47] Becattini, B.; Culmsee, C.; Leone, M.; Zhai, D.; Zhang, X.; Crowell, K.J.; Rega, M.F.; Landshamer, S.; Reed, J.C.; Plesnila, N.; Pellecchia, M. Structure-activity relationships by interligand NOE-based design and synthesis of antiapoptotic compounds targeting Bid. *Proc. Natl. Acad. Sci. U.S.A.,* **2006**, *103*(33), 12602-12606.

[48] Giannangelo, C.R.; Ellis, K.M.; Sexton, A.E.; Stoessel, D.; Creek, D.J. The Role of Metabolomics in Antiparasitic Drug Discovery In: *Comprehensive Analysis of Parasite Biology: From Metabolism to Drug Discovery*. Müller, S.; Cerdan, R.; Radulescu, O., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA, **2016**, pp 321-342.

[49] Emwas, A.H.; Roy, R.; McKay, R.T.; Tenori, L.; Saccenti, E.; Gowda, G.A.N.; Raftery, D.; Alahmari, F.; Jaremko, L.; Jaremko, M.; Wishart, D.S. NMR Spectroscopy for Metabolomics Research. *Metabolites,* **2019**, *9*(7), 123.

[50] Sheedy, J.R. Metabolite analysis of biological fluids and tissues by proton nuclear magnetic resonance spectroscopy. *Methods Mol. Biol.,* **2013**, *1055*, 81-97.

[51] Emwas, A.H.; Luchinat, C.; Turano, P.; Tenori, L.; Roy, R.; Salek, R.M.; Ryan, D.; Merzaban, J.S.; Kaddurah-Daouk, R.; Zeri, A.C.; Nagana Gowda, G.A.; Raftery, D.; Wang, Y.; Brennan, L.; Wishart, D.S. Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics,* **2015**, *11*(4), 872-894.

[52] Le Moyec, L.; Triba, M.N.; Nahon, P.; Bouchemal, N.; Hantz, E.; Goossens, C.; Amathieu, R.; Savarin, P. Nuclear magnetic resonance metabolomics and human liver diseases: The principles and evidence associated with protein

and carbohydrate metabolism. *Biomed. Rep.,* **2017**, *6*(4), 387-395.
[53] Rabinowitz, J.D.; Purdy, J.G.; Vastag, L.; Shenk, T.; Koyuncu, E. Metabolomics in drug target discovery. *Cold Spring Harb. Symp. Quant. Biol.,* **2011**, *76*, 235-246.
[54] Jang, C.; Chen, L.; Rabinowitz, J.D. Metabolomics and Isotope Tracing. *Cell,* **2018**, *173*(4), 822-837.
[55] Lane, A.N.; Fan, T.W. NMR-based Stable Isotope Resolved Metabolomics in systems biochemistry. *Arch. Biochem. Biophys.,* **2017**, *628*, 123-131.
[56] Fradet-Turcotte, A.; Archambault, J. Recent advances in the search for antiviral agents against human papillomaviruses. *Antivir. Ther.,* **2007**, *12*(4), 431-451.
[57] Franco, R.A.; Galbraith, J.W.; Overton, E.T.; Saag, M.S. Direct-acting antivirals and chronic hepatitis C: towards elimination. *Hepatoma Res.,* **2018**, *4*, 1-18.
[58] Zarebska-Michaluk, D. Viral hepatitis C treatment shortening - what is the limit? *Clin. Exp. Hepatol.,* **2019**, *5*(4), 265-270.
[59] Spearman, C.W.; Dusheiko, G.M.; Hellard, M.; Sonderup, M. Hepatitis C. *Lancet,* **2019**, *394*(10207), 1451-1466.
[60] Niepmann, M.; Gerresheim, G.K. Hepatitis C Virus Translation Regulation. *Int. J. Mol. Sci.,* **2020**, *21*(7), 2328.
[61] Mailly, L.; Baumert, T.F. Hepatitis C virus infection and tight junction proteins: The ties that bind. *Biochim. Biophys. Acta Biomembr.,* **2020**, *1862*(7), 183296.
[62] Moriishi, K.; Matsuura, Y. Mechanisms of hepatitis C virus infection. *Antivir. Chem. Chemother.,* **2003**, *14*(6), 285-297.
[63] Gouttenoire, J.; Castet, V.; Montserret, R.; Arora, N.; Raussens, V.; Ruysschaert, J.M.; Diesis, E.; Blum, H.E.; Penin, F.; Moradpour, D. Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B. *J. Virol.,* **2009**, *83*(12), 6257-6268.
[64] Lukavsky, P.J. Structure and function of HCV IRES domains. *Virus Res.,* **2009**, *139*(2), 166-171.
[65] Locker, N.; Easton, L.E.; Lukavsky, P.J. HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J.,* **2007**, *26*(3), 795-805.
[66] Lukavsky, P.J.; Kim, I.; Otto, G.A.; Puglisi, J.D. Structure of HCV IRES domain II determined by NMR. *Nat. Struct. Biol.,* **2003**, *10*(12), 1033-1038.
[67] Kim, I.; Lukavsky, P.J.; Puglisi, J.D. NMR study of 100 kDa HCV IRES RNA using segmental isotope labeling. *J. Am. Chem. Soc.,* **2002**, *124*(32), 9338-9339.
[68] Shin, J.Y.; Bang, K.-M.; Song, H.K.; Kim, N.-K. Structural Studies of Peptide Binding Interaction of HCV IRES Domain IV. *J. Korean. Soc. Magn. Reson.,* **2017**, *21*(3), 109-113.
[69] LaPlante, S.R.; Aubry, N.; Bolger, G.; Bonneau, P.; Carson, R.; Coulombe, R.; Sturino, C.; Beaulieu, P.L. Monitoring drug self-aggregation and potential for promiscuity in off-target in vitro pharmacology screens by a practical NMR strategy. *J. Med. Chem.,* **2013**, *56*(17), 7073-7083.
[70] LaPlante, S.R.; Aubry, N.; De ´ziel, R.; Ni, F.; Xu, P. Transferred 13C T1 Relaxation at Natural Isotopic Abundance: A Practical Method for Determining Site-Specific Changes in Ligand Flexibility upon Binding to a Macromolecule. *J. Am. Chem. Soc.,* **2000**, *122*, 12530-12535.
[71] Wyss, D.F.; Arasappan, A.; Senior, M.M.; Wang, Y.S.; Beyer, B.M.; Njoroge, F.G.; McCoy, M.A. Non-peptidic small-molecule inhibitors of the single-chain hepatitis C virus NS3 protease/NS4A cofactor complex discovered by structure-based NMR screening. *J. Med. Chem.,* **2004**, *47*(10), 2486-2498.
[72] Zhao, L.; Wang, S.; Du, L.; Dev, J.; Zhou, L.; Liu, Z.; Chou, J.J.; OuYang, B. Structural basis of interaction between the hepatitis C virus p7 channel and its blocker hexamethylene amiloride. *Protein Cell,* **2016**, *7*(4), 300-304.
[73] OuYang, B.; Xie, S.; Berardi, M.J.; Zhao, X.; Dev, J.; Yu, W.; Sun, B.; Chou, J.J. Unusual architecture of the p7 channel from hepatitis C virus. *Nature,* **2013**, *498*(7455), 521-525.
[74] Montserret, R.; Saint, N.; Vanbelle, C.; Salvay, A.G.; Simorre, J.P.; Ebel, C.; Sapay, N.; Renisio, J.G.; Bockmann, A.; Steinmann, E.; Pietschmann, T.; Dubuisson, J.; Chipot, C.; Penin, F. NMR structure and ion channel activity of the p7 protein from hepatitis C virus. *J. Biol. Chem.,* **2010**, *285*(41), 31446-31461.
[75] Cook, G.A.; Dawson, L.A.; Tian, Y.; Opella, S.J. Three-dimensional structure and interaction studies of hepatitis C virus p7 in 1,2-dihexanoyl-sn-glycero-3-phosphocholine by solution nuclear magnetic resonance. *Biochemistry,* **2013**, *52*(31), 5295-5303.
[76] Cook, G.A.; Opella, S.J. Secondary structure, dynamics, and architecture of the p7 membrane protein from hepatitis C virus by NMR spectroscopy. *Biochim. Biophys. Acta,* **2011**, *1808*(6), 1448-1453.
[77] Paulsen, R.B.; Seth, P.P.; Swayze, E.E.; Griffey, R.H.; Skalicky, J.J.; Cheatham, T.E., 3rd; Davis, D.R. Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc. Natl. Acad. Sci. U.S.A.,* **2010**, *107*(16), 7263-7268.
[78] Pai, J.; Yoon, T.; Kim, N.D.; Lee, I.S.; Yu, J.; Shin, I. High-throughput profiling of peptide-RNA interactions using peptide microarrays. *J. Am. Chem. Soc.,* **2012**, *134*(46), 19287-19296.
[79] Dujardin, M.; Madan, V.; Gandhi, N.S.; Cantrelle, F.X.; Launay, H.; Huvent, I.; Bartenschlager, R.; Lippens, G.; Hanoulle, X. Cyclophilin A allows the allosteric regulation of a structural motif in the disordered domain 2 of NS5A and thereby fine-tunes HCV RNA replication. *J. Biol. Chem.,* **2019**, *294*(35), 13171-13185.
[80] Barbato, G.; Cicero, D.O.; Cordier, F.; Narjes, F.; Gerlach, B.; Sambucini, S.; Grzesiek, S.; Matassa, V.G.; De Francesco, R.; Bazzo, R. Inhibitor binding induces active site stabilization of the HCV NS3 protein serine protease domain. *EMBO J.,* **2000**, *19*(6), 1195-1206.
[81] Lemke, C.T.; Goudreau, N.; Zhao, S.; Hucke, O.; Thibeault, D.; Llinas-Brunet, M.; White, P.W. Combined X-ray, NMR, and kinetic analyses reveal uncommon binding characteristics of the hepatitis C virus NS3-NS4A protease inhibitor BI 201335. *J. Biol. Chem.,* **2011**, *286*(13), 11434-11443.

[82] Rajesh, S.; Sridhar, P.; Tews, B.A.; Feneant, L.; Cocquerel, L.; Ward, D.G.; Berditchevski, F.; Overduin, M. Structural basis of ligand interactions of the large extracellular domain of tetraspanin CD81. *J. Virol.,* **2012**, *86*(18), 9606-9616.

[83] Spadaccini, R.; D'Errico, G.; D'Alessio, V.; Notomista, E.; Bianchi, A.; Merola, M.; Picone, D. Structural characterization of the transmembrane proximal region of the hepatitis C virus E1 glycoprotein. *Biochim. Biophys. Acta,* **2010**, *1798*(3), 344-353.

[84] Meoni, G.; Lorini, S.; Monti, M.; Madia, F.; Corti, G.; Luchinat, C.; Zignego, A.L.; Tenori, L.; Gragnani, L. The metabolic fingerprints of HCV and HBV infections studied by Nuclear Magnetic Resonance Spectroscopy. *Sci. Rep.,* **2019**, *9*(1), 4128.

[85] Gabbani, T.; Marsico, M.; Bernini, P.; Lorefice, E.; Grappone, C.; Biagini, M.R.; Milani, S.; Annese, V. Metabolomic analysis with (1)H-NMR for non-invasive diagnosis of hepatic fibrosis degree in patients with chronic hepatitis C. *Dig. Liver Dis.,* **2017**, *49*(12), 1338-1344.

[86] Godoy, M.M.; Lopes, E.P.; Silva, R.O.; Hallwass, F.; Koury, L.C.; Moura, I.M.; Goncalves, S.M.; Simas, A.M. Hepatitis C virus infection diagnosis using metabonomics. *J. Viral. Hepat.,* **2010**, *17*(12), 854-858.

[87] Safaei, A.; Rezaei-Tavirani, M.; Oskouie, A.A.; Mohebbi, S.R.; Shabani, M.; Sharifian, A. Serum Metabolic Profiling of Advanced Cirrhosis Based on HCV. *Hepat. Mon.,* **2017**, *17*(3), e44431.

[88] Embade, N.; Marino, Z.; Diercks, T.; Cano, A.; Lens, S.; Cabrera, D.; Navasa, M.; Falcon-Perez, J.M.; Caballeria, J.; Castro, A.; Bosch, J.; Mato, J.M.; Millet, O. Metabolic Characterization of Advanced Liver Fibrosis in HCV Patients as Studied by Serum 1H-NMR Spectroscopy. *PLoS One,* **2016**, *11*(5), e0155094.

[89] LaPlante, S.R.; Padyana, A.K.; Abeywardane, A.; Bonneau, P.; Cartier, M.; Coulombe, R.; Jakalian, A.; Wildeson-Jones, J.; Li, X.; Liang, S.; McKercher, G.; White, P.; Zhang, Q.; Taylor, S.J. Integrated strategies for identifying leads that target the NS3 helicase of the hepatitis C virus. *J. Med. Chem.,* **2014**, *57*(5), 2074-2090.

[90] LaPlante, S.R.; Nar, H.; Lemke, C.T.; Jakalian, A.; Aubry, N.; Kawai, S.H. Ligand bioactive conformation plays a critical role in the design of drugs that target the hepatitis C virus NS3 protease. *J. Med. Chem.,* **2014**, *57*(5), 1777-1789.

[91] LaPlante, S.R.; Bos, M.; Brochu, C.; Chabot, C.; Coulombe, R.; Gillard, J.R.; Jakalian, A.; Poirier, M.; Rancourt, J.; Stammers, T.; Thavonekham, B.; Beaulieu, P.L.; Kukolj, G.; Tsantrizos, Y.S. Conformation-based restrictions and scaffold replacements in the design of hepatitis C virus polymerase inhibitors: discovery of deleobuvir (BI 207127). *J. Med. Chem.,* **2014**, *57*(5), 1845-1854.

[92] LaPlante, S.R.; Aubry, N.; Bonneau, P.R.; Kukolj, G.; Lamarre, D.; Lefebvre, S.; Li, H.; Llinas-Brunet, M.; Plouffe, C.; Cameron, D.R. NMR line-broadening and transferred NOESY as a medicinal chemistry tool for studying inhibitors of the hepatitis C virus NS3 protease domain. *Bioorg. Med. Chem. Lett.,* **2000**, *10*(20), 2271-2274.

[93] Toenjes, S.T.; Gustafson, J.L. Atropisomerism in medicinal chemistry: challenges and opportunities. *Future Med. Chem.,* **2018**, *10*(4), 409-422.

[94] Laplante, S.R.; D. Fader, L.; Fandrick, K.R.; Fandrick, D.R.; Hucke, O.; Kemper, R.; Miller, S.P.; Edwards, P.J. Assessing atropisomer axial chirality in drug discovery and development. *J. Med. Chem.,* **2011**, *54*(20), 7005-7022.

[95] LaPlante, S.R.; Forgione, P.; Boucher, C.; Coulombe, R.; Gillard, J.; Hucke, O.; Jakalian, A.; Joly, M.A.; Kukolj, G.; Lemke, C.; McCollum, R.; Titolo, S.; Beaulieu, P.L.; Stammers, T. Enantiomeric atropisomers inhibit HCV polymerase and/or HIV matrix: characterizing hindered bond rotations and target selectivity. *J. Med. Chem.,* **2014**, *57*(5), 1944-1951.

[96] Pouliot, J.J.; Thomson, M.; Xie, M.; Horton, J.; Johnson, J.; Krull, D.; Mathis, A.; Morikawa, Y.; Parks, D.; Peterson, R.; Shimada, T.; Thomas, E.; Vamathevan, J.; Van Horn, S.; Xiong, Z.; Hamatake, R.; Peat, A.J. Preclinical Characterization and In Vivo Efficacy of GSK8853, a Small-Molecule Inhibitor of the Hepatitis C Virus NS4B Protein. *Antimicrob. Agents Chemother.,* **2015**, *59*(10), 6539-6550.

[97] Gopalsamy, A.; Lim, K.; Ciszewski, G.; Park, K.; Ellingboe, J.W.; Bloom, J.; Insaf, S.; Upeslacis, J.; Mansour, T.S.; Krishnamurthy, G.; Damarla, M.; Pyatski, Y.; Ho, D.; Howe, A.Y.; Orlowski, M.; Feld, B.; O'Connell, J. Discovery of pyrano[3,4-b]indoles as potent and selective HCV NS5B polymerase inhibitors. *J. Med. Chem.,* **2004**, *47*(26), 6603-6608.

[98] Jonckers, T.H.; Vandyck, K.; Vandekerckhove, L.; Hu, L.; Tahri, A.; Van Hoof, S.; Lin, T.I.; Vijgen, L.; Berke, J.M.; Lachau-Durand, S.; Stoops, B.; Leclercq, L.; Fanning, G.; Samuelsson, B.; Nilsson, M.; Rosenquist, A.; Simmen, K.; Raboisson, P. Nucleotide prodrugs of 2'-deoxy-2'-spirooxetane ribonucleosides as novel inhibitors of the HCV NS5B polymerase. *J. Med. Chem.,* **2014**, *57*(5), 1836-1844.

[99] Clark, J.L.; Hollecker, L.; Mason, J.C.; Stuyver, L.J.; Tharnish, P.M.; Lostia, S.; McBrayer, T.R.; Schinazi, R.F.; Watanabe, K.A.; Otto, M.J.; Furman, P.A.; Stec, W.J.; Patterson, S.E.; Pankiewicz, K.W. Design, synthesis, and antiviral activity of 2'-deoxy-2'-fluoro-2'-C-methylcytidine, a potent inhibitor of hepatitis C virus replication. *J. Med. Chem.,* **2005**, *48*(17), 5504-5508.

[100] Lee, H.; Zhu, T.; Patel, K.; Zhang, Y.Y.; Truong, L.; Hevener, K.E.; Gatuz, J.L.; Subramanya, G.; Jeong, H.Y.; Uprichard, S.L.; Johnson, M.E. High-throughput screening (HTS) and hit validation to identify small molecule inhibitors with activity against NS3/4A proteases from multiple hepatitis C virus genotypes. *PLoS One,* **2013**, *8*(10), e75144.

[101] Ismail, M.A.; Abouzid, K.A.; Mohamed, N.S.; Dokla, E.M. Ligand design, synthesis and biological anti-HCV evaluations for genotypes 1b and 4a of certain 4-(3- & 4-[3-(3,5-dibromo-4-hydroxyphenyl)-propylamino]phenyl) butyric acids and 3-(3,5-dibromo-4-hydroxyphenyl)-propylamino-acetamidobenzoic acid esters. *J. Enzyme Inhib. Med. Chem.,* **2013**, *28*(6), 1274-1290.

[102] VanCompernolle, S.E.; Wiznycia, A.V.; Rush, J.R.; Dhanasekaran, M.; Baures, P.W.; Todd, S.C. Small molecule

inhibition of hepatitis C virus E2 binding to CD81. *Virology,* **2003**, *314*(1), 371-380.
[103] Ruggiero, E.; Richter, S.N. G-quadruplexes and G-quadruplex ligands: targets and tools in antiviral therapy. *Nucleic Acids Res.,* **2018**, *46*(7), 3270-3283.
[104] Yang, L.; Liu, F.; Tong, X.; Hoffmann, D.; Zuo, J.; Lu, M. Treatment of Chronic Hepatitis B Virus Infection Using Small Molecule Modulators of Nucleocapsid Assembly: Recent Advances and Perspectives. *ACS Infect. Dis.,* **2019**, *5*(5), 713-724.
[105] Prescott, N.A.; Bram, Y.; Schwartz, R.E.; David, Y. Targeting Hepatitis B Virus Covalently Closed Circular DNA and Hepatitis B Virus X Protein: Recent Advances and New Approaches. *ACS Infect. Dis.,* **2019**, *5*(10), 1657-1667.
[106] Mehmankhah, M.; Bhat, R.; Anvar, M.S.; Ali, S.; Alam, A.; Farooqui, A.; Amir, F.; Anwer, A.; Khan, S.; Azmi, I.; Ali, R.; Ishrat, R.; Hassan, M.I.; Minuchehr, Z.; Kazim, S.N. Structure-Guided Approach to Identify Potential Inhibitors of Large Envelope Protein to Prevent Hepatitis B Virus Infection. *Oxid. Med. Cell. Longev.,* **2019**, *2019*, 1297484.
[107] Tan, W.S.; Ho, K.L. Phage display creates innovative applications to combat hepatitis B virus. *World J. Gastroenterol.,* **2014**, *20*(33), 11650-11670.
[108] Crosby, I.T.; Bourke, D.G.; Jones, E.D.; Jeynes, T.P.; Cox, S.; Coates, J.A.; Robertson, A.D. Antiviral agents 3. Discovery of a novel small molecule non-nucleoside inhibitor of hepatitis B virus (HBV). *Bioorg. Med. Chem. Lett.,* **2011**, *21*(6), 1644-1648.
[109] Sebastien, B. Towards Elimination of Hepatitis B Virus Using Novel Drugs, Approaches and Combined Modalities. *Clin. Liver Dis.,* **2016**, *20*(4), 737-749.
[110] Sarrazin, S.; Lamanna, W.C.; Esko, J.D. Heparan sulfate proteoglycans. *Cold Spring Harb. Perspect. Biol.,* **2011**, *3*(7), a004952.
[111] Gong, Q.; Wu, J.; Liu, Q.; Shi, Y.; Cui, D.; Xu, L.; Zhang, Y. Solution structure of N-terminal segment of hepatitis B virus surface antigen Pre-S1. *Sci. China C Life Sci.,* **1998**, *41*(5), 530-541.
[112] Chi, S.W.; Kim, D.H.; Lee, S.H.; Chang, I.; Han, K.H. Pre-structured motifs in the natively unstructured preS1 surface antigen of hepatitis B virus. *Protein Sci.,* **2007**, *16*(10), 2108-2117.
[113] Chi, S.W.; Kim, D.H.; Kim, J.S.; Lee, M.K.; Han, K.H. Solution conformation of an immunodominant epitope in the hepatitis B virus preS2 surface antigen. *Antiviral Res.,* **2006**, *72*(3), 207-215.
[114] Lee, S.H.; Cha, E.J.; Lim, J.E.; Kwon, S.H.; Kim, D.H.; Cho, H.; Han, K.H. Structural characterization of an intrinsically unfolded mini-HBX protein from hepatitis B virus. *Mol. Cells,* **2012**, *34*(2), 165-169.
[115] Lee, Y.-T.; Kim, B.; Kim, K.-S.a.; Choi, B.-S. Structural characterization of HBx-interacting protein using NMR spectroscopy. *J. Korean. Soc. Magn. Reson.,* **2005**, *9*, 122-137.
[116] Bishop, K.D.; Blocker, F.J.; Egan, W.; James, T.L. Hepatitis B virus direct repeat sequence: imino proton exchange rates and distance and torsion angle restraints from NMR. *Biochemistry,* **1994**, *33*(2), 427-438.
[117] Flodell, S.; Schleucher, J.; Cromsigt, J.; Ippel, H.; Kidd-Ljunggren, K.; Wijmenga, S. The apical stem-loop of the hepatitis B virus encapsidation signal folds into a stable tri-loop with two underlying pyrimidine bulges. *Nucleic Acids Res.,* **2002**, *30*(21), 4803-4811.
[118] Muhamad, A.; Ho, K.L.; Rahman, M.B.; Tejo, B.A.; Uhrin, D.; Tan, W.S. Hepatitis B virus peptide inhibitors: solution structures and interactions with the viral capsid. *Org. Biomol. Chem.,* **2015**, *13*(28), 7780-7789.
[119] Lin, X.; Shi, H.; Zhang, W.; Qiu, Z.; Zhou, Z.; Dey, F.; Zhong, S.; Qiu, H.; Xie, J.; Zhou, X.; Yang, G.; Tang, G.; Shen, H.C.; Zhu, W. A New Approach of Mitigating CYP3A4 Induction Led to the Discovery of Potent Hepatitis B Virus (HBV) Capsid Inhibitor with Optimal ADMET Profiles. *J. Med. Chem.,* **2019**, *62*(22), 10352-10361.
[120] Na, H.G.; Imran, A.; Kim, K.; Han, H.S.; Lee, Y.J.; Kim, M.J.; Yun, C.S.; Jung, Y.S.; Lee, J.Y.; Han, S.B. Discovery of a New Sulfonamide Hepatitis B Capsid Assembly Modulator. *ACS Med. Chem. Lett.,* **2020**, *11*(2), 166-171.
[121] Gilany, K.; Mohamadkhani, A.; Chashmniam, S.; Shahnazari, P.; Amini, M.; Arjmand, B.; Malekzadeh, R.; Nobakht Motlagh Ghoochani, B.F. Metabolomics analysis of the saliva in patients with chronic hepatitis B using nuclear magnetic resonance: a pilot study. *Iran J. Basic Med. Sci.,* **2019**, *22*(9), 1044-1049.
[122] Guo, W.; Tan, H.Y.; Wang, N.; Wang, X.; Feng, Y. Deciphering hepatocellular carcinoma through metabolomics: from biomarker discovery to therapy evaluation. *Cancer Manag. Res.,* **2018**, *10*, 715-734.
[123] van Ditzhuijsen, T.J.; Kuijpers, L.P.; Koens, M.J.; Rijntjes, P.J.; van Loon, A.M.; Yap, S.H. Hepatitis B pre-S1 and pre-S2 proteins: clinical significance and relation to hepatitis B virus DNA. *J. Med. Virol.,* **1990**, *32*(2), 87-91.
[124] Lecoq, L.; Wang, S.; Wiegand, T.; Bressanelli, S.; Nassal, M.; Meier, B.H.; Bockmann, A. Localizing Conformational Hinges by NMR: Where Do Hepatitis B Virus Core Proteins Adapt for Capsid Assembly? *Chemphyschem,* **2018**, *19*(11), 1336-1340.
[125] Bottcher, B.; Wynne, S.A.; Crowther, R.A. Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature,* **1997**, *386*(6620), 88-91.
[126] Wynne, S.A.; Crowther, R.A.; Leslie, A.G. The crystal structure of the human hepatitis B virus capsid. *Mol. Cell,* **1999**, *3*(6), 771-780.
[127] Zheng, H.; Chen, M.; Lu, S.; Zhao, L.; Ji, J.; Gao, H. Metabolic characterization of hepatitis B virus-related liver cirrhosis using NMR-based serum metabolomics *Metabolomics,* **2017**, *13*, 121.
[128] Song, Z.; Wang, H.; Yin, X.; Deng, P.; Jiang, W. Application of NMR metabolomics to search for human disease biomarkers in blood. *Clin. Chem. Lab. Med.,* **2019**, *57*(4), 417-441.
[129] Cox, I.J.; Aliev, A.E.; Crossey, M.M.; Dawood, M.; Al-Mahtab, M.; Akbar, S.M.; Rahman, S.; Riva, A.; Williams, R.; Taylor-Robinson, S.D. Urinary nuclear magnetic resonance spectroscopy of a Bangladeshi cohort with hepatitis-B hepatocellular carcinoma: A biomarker

corroboration study. *World J. Gastroenterol.,* **2016**, *22*(16), 4191-4200.
[130] Shi, C.; Wu, C.Q.; Cao, A.M.; Sheng, H.Z.; Yan, X.Z.; Liao, M.Y. NMR-spectroscopy-based metabonomic approach to the analysis of Bay41-4109, a novel anti-HBV compound, induced hepatotoxicity in rats. *Toxicol. Lett.,* **2007**, *173*(3), 161-167.
[131] Li, H.; Zhu, W.; Zhang, L.; Lei, H.; Wu, X.; Guo, L.; Chen, X.; Wang, Y.; Tang, H. The metabolic responses to hepatitis B virus infection shed new light on pathogenesis and targets for treatment. *Sci. Rep.,* **2015**, *5*, 8421.
[132] Dan, Y.; Zhang, Y.; Cheng, L.; Ma, J.; Xi, Y.; Yang, L.; Su, C.; Shao, B.; Huang, A.; Xiang, R.; Cheng, P. Hepatitis B virus X protein (HBx)-induced abnormalities of nucleic acid metabolism revealed by (1)H-NMR-based metabonomics. *Sci. Rep.,* **2016**, *6*, 24430.
[133] Yang, L.; Lu, M. Small Molecule Inhibitors of Hepatitis B Virus Nucleocapsid Assembly: A New Approach to Treat Chronic HBV Infection. *Curr. Med. Chem.,* **2018**, *25*(7), 802-813.
[134] Bourne, C.R.; Finn, M.G.; Zlotnick, A. Global structural changes in hepatitis B virus capsids induced by the assembly effector HAP1. *J. Virol.,* **2006**, *80*(22), 11055-11061.
[135] Qiu, Z.; Lin, X.; Zhou, M.; Liu, Y.; Zhu, W.; Chen, W.; Zhang, W.; Guo, L.; Liu, H.; Wu, G.; Huang, M.; Jiang, M.; Xu, Z.; Zhou, Z.; Qin, N.; Ren, S.; Qiu, H.; Zhong, S.; Zhang, Y.; Zhang, Y.; Wu, X.; Shi, L.; Shen, F.; Mao, Y.; Zhou, X.; Yang, W.; Wu, J.Z.; Yang, G.; Mayweg, A.V.; Shen, H.C.; Tang, G. Design and Synthesis of Orally Bioavailable 4-Methyl Heteroaryldihydropyrimidine Based Hepatitis B Virus (HBV) Capsid Inhibitors. *J. Med. Chem.,* **2016**, *59*(16), 7651-7666.
[136] Boucle, S.; Lu, X.; Bassit, L.; Ozturk, T.; Russell, O.O.; Amblard, F.; Coats, S.J.; Schinazi, R.F. Synthesis and antiviral evaluation of novel heteroarylpyrimidines analogs as HBV capsid effectors. *Bioorg. Med. Chem. Lett.,* **2017**, *27*(4), 904-910.
[137] Ren, Q.; Liu, X.; Luo, Z.; Li, J.; Wang, C.; Goldmann, S.; Zhang, J.; Zhang, Y. Discovery of hepatitis B virus capsid assembly inhibitors leading to a heteroaryldihydropyrimidine based clinical candidate (GLS4). *Bioorg. Med. Chem.,* **2017**, *25*(3), 1042-1056.
[138] Tan, J.; Zhou, M.; Cui, X.; Wei, Z.; Wei, W. Discovery of Oxime Ethers as Hepatitis B Virus (HBV) Inhibitors by Docking, Screening and In Vitro Investigation. *Molecules,* **2018**, *23*(3), 637.
[139] Chen, H.; Wang, L.J.; Ma, Y.B.; Huang, X.Y.; Geng, C.A.; Zhang, X.M.; Chen, J.J. Panaxadiol and panaxatriol derivatives as anti-hepatitis B virus inhibitors. *Nat. Prod. Bioprospect.,* **2014**, *4*(3), 163-174.
[140] Li, X.L.; Wang, C.Z.; Mehendale, S.R.; Sun, S.; Wang, Q.; Yuan, C.S. Panaxadiol, a purified ginseng component, enhances the anti-cancer effects of 5-fluorouracil in human colorectal cancer cells. *Cancer Chemother. Pharmacol.,* **2009**, *64*(6), 1097-1104.
[141] Chen, G.-y.; Liu , M.; Yu, M.-b.; Wei, S.-c. In *MATEC Web of Conferences - ICCBS 2016*, **2016**; Vol. *60*, pp 1-5.
[142] Li, S.F.; Jiao, Y.Y.; Zhang, Z.Q.; Chao, J.B.; Jia, J.; Shi, X.L.; Zhang, L.W. Diterpenes from buds of Wikstroemia chamaedaphne showing anti-hepatitis B virus activities. *Phytochemistry,* **2018**, *151*, 17-25.
[143] Yang, F.; Zhang, H.J.; Zhang, Y.Y.; Chen, W.S.; Yuan, H.L.; Lin, H.W. A hepatitis B virus inhibitory neolignan from Herpetospermum caudigerum. *Chem. Pharm. Bull. (Tokyo),* **2010**, *58*(3), 402-404.
[144] Chen, H.; Ma, Y.B.; Huang, X.Y.; Geng, C.A.; Zhao, Y.; Wang, L.J.; Guo, R.H.; Liang, W.J.; Zhang, X.M.; Chen, J.J. Synthesis, structure-activity relationships and biological evaluation of dehydroandrographolide and andrographolide derivatives as novel anti-hepatitis B virus agents. *Bioorg. Med. Chem. Lett.,* **2014**, *24*(10), 2353-2359.
[145] Waxman, L.; Darke, P.L. The herpesvirus proteases as targets for antiviral chemotherapy. *Antivir. Chem. Chemother. ,* **2000**, *11*, 1-22.
[146] Azwa, A.; Barton, S.E. Aspects of herpes simplex virus: a clinical review. *J. Fam. Plann. Reprod. Health Care,* **2009**, *35*(4), 237-242.
[147] Koelle, D.M.; Corey, L. Recent progress in herpes simplex virus immunobiology and vaccine research. *Clin. Microbiol. Rev.,* **2003**, *16*(1), 96-113.
[148] Crimi, S.; Fiorillo, L.; Bianchi, A.; D'Amico, C.; Amoroso, G.; Gorassini, F.; Mastroieni, R.; Marino, S.; Scoglio, C.; Catalano, F.; Campagna, P.; Bocchieri, S.; De Stefano, R.; Fiorillo, M.T.; Cicciu, M. Herpes Virus, Oral Clinical Signs and QoL: Systematic Review of Recent Data. *Viruses,* **2019**, *11*(5), 463.
[149] McElwee, M.; Vijayakrishnan, S.; Rixon, F.; Bhella, D. Structure of the herpes simplex virus portal-vertex. *PLoS Biol.,* **2018**, *16*(6), e2006191.
[150] Looker, K.J.; Garnett, G.P. A systematic review of the epidemiology and interaction of herpes simplex virus types 1 and 2. *Sex. Transm. Infect.,* **2005**, *81*(2), 103-107.
[151] Nicoll, M.P.; Proenca, J.T.; Efstathiou, S. The molecular basis of herpes simplex virus latency. *FEMS Microbiol. Rev.,* **2012**, *36*(3), 684-705.
[152] Chemaitelly, H.; Nagelkerke, N.; Omori, R.; Abu-Raddad, L.J. Characterizing herpes simplex virus type 1 and type 2 seroprevalence declines and epidemiological association in the United States. *PLoS One,* **2019**, *14*(6), e0214151.
[153] Suzich, J.B.; Cliffe, A.R. Strength in diversity: Understanding the pathways to herpes simplex virus reactivation. *Virology,* **2018**, *522*, 81-91.
[154] Xing, J.; Ni, L.; Wang, S.; Wang, K.; Lin, R.; Zheng, C. Herpes simplex virus 1-encoded tegument protein VP16 abrogates the production of beta interferon (IFN) by inhibiting NF-kappaB activation and blocking IFN regulatory factor 3 to recruit its coactivator CBP. *J. Virol.,* **2013**, *87*(17), 9788-9801.
[155] Hancock, M.H.; Cliffe, A.R.; Knipe, D.M.; Smiley, J.R. Herpes simplex virus VP16, but not ICP0, is required to reduce histone occupancy and enhance histone acetylation on viral genomes in U2OS osteosarcoma cells. *J. Virol.,* **2010**, *84*(3), 1366-1375.
[156] Pfander, R.; Neumann, L.; Zweckstetter, M.; Seger, C.; Holak, T.A.; Tampe, R. Structure of the active domain of

the herpes simplex virus protein ICP47 in water/sodium dodecyl sulfate solution determined by nuclear magnetic resonance spectroscopy. *Biochemistry,* **1999**, *38*(41), 13692-13698.

[157] Tunnicliffe, R.B.; Hautbergue, G.M.; Kalra, P.; Jackson, B.R.; Whitehouse, A.; Wilson, S.A.; Golovanov, A.P. Structural basis for the recognition of cellular mRNA export factor REF by herpes viral proteins HSV-1 ICP27 and HVS ORF57. *PLoS Pathog.,* **2011**, *7*(1), e1001244.

[158] Galdiero, S.; Russo, L.; Falanga, A.; Cantisani, M.; Vitiello, M.; Fattorusso, R.; Malgieri, G.; Galdiero, M.; Isernia, C. Structure and orientation of the gH625-644 membrane interacting region of herpes simplex virus type 1 in a membrane mimetic system. *Biochemistry,* **2012**, *51*(14), 3121-3128.

[159] Lesbirel, S.; Wilson, S.A. The m(6)Amethylase complex and mRNA export. *Biochim. Biophys. Acta Gene Regul. Mech.,* **2019**, *1862*(3), 319-328.

[160] Schlosser, G.; Mezo, G.; Kiss, R.; Vass, E.; Majer, Z.; Feijlbrief, M.; Perczel, A.; Bosze, S.; Welling-Wester, S.; Hudecz, F. Synthesis, solution structure analysis and antibody binding of cyclic epitope peptides from glycoprotein D of Herpes simplex virus type I. *Biophys. Chem.,* **2003**, *106*(2), 155-171.

[161] Moss, N.; Deziel, R.; Adams, J.; Aubry, N.; Bailey, M.; Baillet, M.; Beaulieu, P.; DiMaio, J.; Duceppe, J.S.; Ferland, J.M.; et al. Inhibition of herpes simplex virus type 1 ribonucleotide reductase by substituted tetrapeptide derivatives. *J. Med. Chem.,* **1993**, *36*(20), 3005-3009.

[162] Sahu, P.K.; Umme, T.; Yu, J.; Kim, G.; Qu, S.; Naik, S.D.; Jeong, L.S. Structure-Activity Relationships of Acyclic Selenopurine Nucleosides as Antiviral Agents. *Molecules,* **2017**, *22*(7), 1167.

[163] Spadari, S.; Maga, G.; Focher, F.; Ciarrocchi, G.; Manservigi, R.; Arcamone, F.; Capobianco, M.; Carcuro, A.; Colonna, F.; Iotti, S.; et al. L-thymidine is phosphorylated by herpes simplex virus type 1 thymidine kinase and inhibits viral growth. *J. Med. Chem.,* **1992**, *35*(22), 4214-4220.

[164] de Souza, L.M.; Sassaki, G.L.; Romanos, M.T.; Barreto-Bergter, E. Structural characterization and anti-HSV-1 and HSV-2 activity of glycolipids from the marine algae Osmundaria obtusiloba isolated from Southeastern Brazilian coast. *Mar. Drugs,* **2012**, *10*(4), 918-931.

[165] Lavoie, S.; Cote, I.; Pichette, A.; Gauthier, C.; Ouellet, M.; Nagau-Lavoie, F.; Mshvildadze, V.; Legault, J. Chemical composition and anti-herpes simplex virus type 1 (HSV-1) activity of extracts from Cornus canadensis. *BMC Complement. Altern. Med.,* **2017**, *17*(1), 123.

[166] Tunnicliffe, R.B.; Tian, X.; Storer, J.; Sandri-Goldin, R.M.; Golovanov, A.P. Overlapping motifs on the herpes viral proteins ICP27 and ORF57 mediate interactions with the mRNA export adaptors ALYREF and UIF. *Sci. Rep.,* **2018**, *8*(1), 15005.

[167] Corbin-Lickfett, K.A.; Souki, S.K.; Cocco, M.J.; Sandri-Goldin, R.M. Three arginine residues within the RGG box are crucial for ICP27 binding to herpes simplex virus 1 GC-rich sequences and for efficient viral RNA export. *J. Virol.,* **2010**, *84*(13), 6367-6376.

[168] Hembram, D.S.S.; Negi, H.; Biswas, P.; Tripathi, V.; Bhushan, L.; Shet, D.; Kumar, V.; Das, R. The Viral SUMO-Targeted Ubiquitin Ligase ICP0 is Phosphorylated and Activated by Host Kinase Chk2. *J. Mol. Biol.,* **2020**, *432*(7), 1952-1977.

[169] Pozhidaeva, A.K.; Mohni, K.N.; Dhe-Paganon, S.; Arrowsmith, C.H.; Weller, S.K.; Korzhnev, D.M.; Bezsonova, I. Structural Characterization of Interaction between Human Ubiquitin-specific Protease 7 and Immediate-Early Protein ICP0 of Herpes Simplex Virus-1. *J. Biol. Chem.,* **2015**, *290*(38), 22907-22918.

[170] Mezo, G.; Majer, Z.; Vass, E.; Jimenez, M.A.; Andreu, D.; Hudecz, F. Conformational study of linear and cyclic peptides corresponding to the 276-284 epitope region of HSV gD-1. *Biophys. Chem.,* **2003**, *103*(1), 51-65.

[171] Kamen, D.E.; Gross, S.T.; Girvin, M.E.; Wilson, D.W. Structural basis for the physiological temperature dependence of the association of VP16 with the cytoplasmic tail of herpes simplex virus glycoprotein H. *J. Virol.,* **2005**, *79*(10), 6134-6141.

[172] Czaplicki, J.; Bohner, T.; Habermann, A.K.; Folkers, G.; Milon, A. A transferred NOE study of a tricyclic analog of acyclovir bound to thymidine kinase. *J. Biomol. NMR,* **1996**, *8*(3), 261-272.

[173] Huang, R.; Gao, H.; Ma, L.; Wang, X.; Jia, J.; Wang, M.; Zhang, L.; Liu, X.; Zheng, P.; Yang, L.; Yang, L.; Dan, L.; Peng, X. Dynamic 1H NMR-based extracellular metabonomic analysis of oligodendroglia cells infected with herpes simplex virus type 1. *Metabolomics,* **2014**, *10*, 33–41

[174] Chen, S.D.; Gao, H.; Zhu, Q.C.; Wang, Y.Q.; Li, T.; Mu, Z.Q.; Wu, H.L.; Peng, T.; Yao, X.S. Houttuynoids A-E, anti-herpes simplex virus active flavonoids with novel skeletons from Houttuynia cordata. *Org. Lett.,* **2012**, *14*(7), 1772-1775.

[175] Prinsloo, G.; Vervoort, J. Identifying anti-HSV compounds from unrelated plants using NMR and LC-MS metabolomic analysis. *Metabolomics,* **2018**, *14*(10), 134.

[176] Moss, N.; Beaulieu, P.; Duceppe, J.S.; Ferland, J.M.; Gauthier, J.; Ghiro, E.; Goulet, S.; Guse, I.; Llinas-Brunet, M.; Plante, R.; Plamondon, L.; Wernic, D.; Deziel, R. Ureido-based peptidomimetic inhibitor of herpes simplex virus ribonucleotide reductase: an investigation of inhibitor bioactive conformation. *J. Med. Chem.,* **1996**, *39*(11), 2178-2187.

[177] John, J.; Kim, Y.; Bennett, N.; Das, K.; Liekens, S.; Naesens, L.; Arnold, E.; Maguire, A.R.; Gotte, M.; Dehaen, W.; Balzarini, J. Pronounced Inhibition Shift from HIV Reverse Transcriptase to Herpetic DNA Polymerases by Increasing the Flexibility of alpha-Carboxy Nucleoside Phosphonates. *J. Med. Chem.,* **2015**, *58*(20), 8110-8127.

[178] Sun, H.; Zhi, C.; Wright, G.E.; Ubiali, D.; Pregnolato, M.; Verri, A.; Focher, F.; Spadari, S. Molecular modeling and synthesis of inhibitors of herpes simplex virus type 1 uracil-DNA glycosylase. *J. Med. Chem.,* **1999**, *42*(13), 2344-2350.

[179] Xu, H.; Maga, G.; Focher, F.; Smith, E.R.; Spadari, S.; Gambino, J.; Wright, G.E. Synthesis, properties, and pharmacokinetic studies of N2-phenylguanine derivatives as

inhibitors of herpes simplex virus thymidine kinases. *J. Med. Chem.,* **1995**, *38*(1), 49-57.
[180] Chacko, A.M.; Qu, W.; Kung, H.F. Synthesis and in vitro evaluation of 5-[(18)f]fluoroalkyl pyrimidine nucleosides for molecular imaging of herpes simplex virus type 1 thymidine kinase reporter gene expression. *J. Med. Chem.,* **2008**, *51*(18), 5690-5701.
[181] Lowden, C.T.; Bastow, K.F. Anti-herpes simplex virus activity of substituted 1-hydroxyacridones. *J. Med. Chem.,* **2003**, *46*(23), 5015-5020.
[182] Pachota, M.; Klysik, K.; Synowiec, A.; Ciejka, J.; Szczubialka, K.; Pyrc, K.; Nowakowska, M. Inhibition of Herpes Simplex Viruses by Cationic Dextran Derivatives. *J. Med. Chem.,* **2017**, *60*(20), 8620-8630.
[183] Zhu, W.; Li, J.; Liang, G. How does cellular heparan sulfate function in viral pathogenicity? *Biomed. Environ. Sci.,* **2011**, *24*(1), 81-87.
[184] Dey, P.; Bergmann, T.; Cuellar-Camacho, J.L.; Ehrmann, S.; Chowdhury, M.S.; Zhang, M.; Dahmani, I.; Haag, R.; Azab, W. Multivalent Flexible Nanogels Exhibit Broad-Spectrum Antiviral Activity by Blocking Virus Entry. *ACS Nano,* **2018**, *12*(7), 6429-6442.
[185] Pachota, M.; Klysik-Trzcianska, K.; Synowiec, A.; Yukioka, S.; Yusa, S.I.; Zajac, M.; Zawilinska, B.; Dzieciatkowski, T.; Szczubialka, K.; Pyrc, K.; Nowakowska, M. Highly Effective and Safe Polymeric Inhibitors of Herpes Simplex Virus in Vitro and in Vivo. *ACS Appl. Mater. Interfaces,* **2019**, *11*(30), 26745-26752.
[186] Marcsisin, S.R.; Narute, P.S.; Emert-Sedlak, L.A.; Kloczewiak, M.; Smithgall, T.E.; Engen, J.R. On the solution conformation and dynamics of the HIV-1 viral infectivity factor. *J. Mol. Biol.,* **2011**, *410*(5), 1008-1022.
[187] Turner, B.G.; Summers, M.F. Structural biology of HIV. *J. Mol. Biol.,* **1999**, *285*(1), 1-32.
[188] Ghosn, J.; Taiwo, B.; Seedat, S.; Autran, B.; Katlama, C. Hiv. *Lancet,* **2018**, *392*(10148), 685-697.
[189] Lingan, K. A Review on Various Aspects of HIV Infection. *HIV Curr. Res.,* **2018**, *3*(1).
[190] Differences between HIV-1 and HIV-2. https://www.medicalnewstoday.com/articles/323893
[191] Nyamweya, S.; Hegedus, A.; Jaye, A.; Rowland-Jones, S.; Flanagan, K.L.; Macallan, D.C. Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis. *Rev. Med. Virol.,* **2013**, *23*(4), 221-240.
[192] Moir, S.; Chun, T.W.; Fauci, A.S. Pathogenic mechanisms of HIV disease. *Annu. Rev. Pathol.,* **2011**, *6*, 223-248.
[193] U.S. Department of Health and Human Services. The stages of HIV Infection. https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/46/the-stages-of-hiv-infection
[194] De Clercq, E. HIV LIFE CYCLE: TARGETS FOR ANTI-HIV AGENTS. In: *HIV-1 Integrase: Mechanism and Inhibitor Design.*, 1 ed. Neamati, N., Ed.; John Wiley & Sons, Inc. , **2011**, pp 1-14.
[195] Tsou, L.K.; Chen, C.H.; Dutschman, G.E.; Cheng, Y.C.; Hamilton, A.D. Blocking HIV-1 entry by a gp120 surface binding inhibitor. *Bioorg. Med. Chem. Lett.,* **2012**, *22*(9), 3358-3361.
[196] Thorlund, K.; Horwitz, M.S.; Fife, B.T.; Lester, R.; Cameron, D.W. Landscape review of current HIV 'kick and kill' cure research - some kicking, not enough killing. *BMC Infect. Dis.,* **2017**, *17*(1), 595.
[197] Apellaniz, B.; Rujas, E.; Serrano, S.; Morante, K.; Tsumoto, K.; Caaveiro, J.M.; Jimenez, M.A.; Nieva, J.L. The Atomic Structure of the HIV-1 gp41 Transmembrane Domain and Its Connection to the Immunogenic Membrane-proximal External Region. *J. Biol. Chem.,* **2015**, *290*(21), 12999-13015.
[198] Celigoy, J.; Ramirez, B.; Tao, L.; Rong, L.; Yan, L.; Feng, Y.R.; Quinnan, G.V.; Broder, C.C.; Caffrey, M. Probing the HIV gp120 envelope glycoprotein conformation by NMR. *J. Biol. Chem.,* **2011**, *286*(27), 23975-23981.
[199] Oliva, R.; Falcigno, L.; D'Auria, G.; Dettin, M.; Scarinci, C.; Pasquato, A.; Di Bello, C.; Paolillo, L. Structural investigation of the HIV-1 envelope glycoprotein gp160 cleavage site, 2: relevance of an N-terminal helix. *ChemBioChem,* **2003**, *4*(8), 727-733.
[200] Schuler, W.; Wecker, K.; de Rocquigny, H.; Baudat, Y.; Sire, J.; Roques, B.P. NMR structure of the (52-96) C-terminal domain of the HIV-1 regulatory protein Vpr: molecular insights into its biological functions. *J. Mol. Biol.,* **1999**, *285*(5), 2105-2117.
[201] Morellet, N.; Bouaziz, S.; Petitjean, P.; Roques, B.P. NMR structure of the HIV-1 regulatory protein VPR. *J. Mol. Biol.,* **2003**, *327*(1), 215-227.
[202] Marquette, A.; Leborgne, C.; Schartner, V.; Salnikov, E.; Bechinger, B.; Kichler, A. Peptides derived from the C-terminal domain of HIV-1 Viral Protein R in lipid bilayers: Structure, membrane positioning and gene delivery. *Biochim. Biophys. Acta Biomembr.,* **2020**, *1862*(2), 183149.
[203] Zhang, H.; Lin, E.C.; Das, B.B.; Tian, Y.; Opella, S.J. Structural determination of virus protein U from HIV-1 by NMR in membrane environments. *Biochim. Biophys. Acta,* **2015**, *1848*(11 Pt A), 3007-3018.
[204] Watkins, J.D.; Campbell, G.R.; Halimi, H.; Loret, E.P. Homonuclear 1H NMR and circular dichroism study of the HIV-1 Tat Eli variant. *Retrovirology,* **2008**, *5*, 83.
[205] Bayer, P.; Kraft, M.; Ejchart, A.; Westendorp, M.; Frank, R.; Rosch, P. Structural studies of HIV-1 Tat protein. *J. Mol. Biol.,* **1995**, *247*(4), 529-535.
[206] Shojania, S.; O'Neil, J.D. HIV-1 Tat is a natively unfolded protein: the solution conformation and dynamics of reduced HIV-1 Tat-(1-72) by NMR spectroscopy. *J. Biol. Chem.,* **2006**, *281*(13), 8347-8356.
[207] Sharaf, N.G.; Brereton, A.E.; Byeon, I.L.; Karplus, P.A.; Gronenborn, A.M. NMR structure of the HIV-1 reverse transcriptase thumb subdomain. *J. Biomol. NMR,* **2016**, *66*(4), 273-280.
[208] Morellet, N.; Demene, H.; Teilleux, V.; Huynh-Dinh, T.; de Rocquigny, H.; Fournie-Zaluski, M.C.; Roques, B.P. Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC). *J. Mol. Biol.,* **1998**, *283*(2), 419-434.

[209] Ganser, L.R.; Al-Hashimi, H.M. HIV-1 leader RNA dimeric interface revealed by NMR. *Proc. Natl. Acad. Sci. U.S.A.,* **2016**, *113*(47), 13263-13265.
[210] Lu, K.; Heng, X.; Garyu, L.; Monti, S.; Garcia, E.L.; Kharytonchyk, S.; Dorjsuren, B.; Kulandaivel, G.; Jones, S.; Hiremath, A.; Divakaruni, S.S.; LaCotti, C.; Barton, S.; Tummillo, D.; Hosic, A.; Edme, K.; Albrecht, S.; Telesnitsky, A.; Summers, M.F. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science,* **2011**, *334*(6053), 242-245.
[211] Keane, S.C.; Summers, M.F. NMR Studies of the Structure and Function of the HIV-1 5'-Leader. *Viruses,* **2016**, *8*(12), 338.
[212] Lebars, I.; Richard, T.; Di Primo, C.; Toulme, J.J. NMR structure of a kissing complex formed between the TAR RNA element of HIV-1 and a LNA-modified aptamer. *Nucleic Acids Res.,* **2007**, *35*(18), 6103-6114.
[213] Kieken, F.; Paquet, F.; Brule, F.; Paoletti, J.; Lancelot, G. A new NMR solution structure of the SL1 HIV-1Lai loop-loop dimer. *Nucleic Acids Res.,* **2006**, *34*(1), 343-352.
[214] Makatini, M.M.; Petzold, K.; Sriharsha, S.N.; Ndlovu, N.; Soliman, M.E.; Honarparvar, B.; Parboosing, R.; Naidoo, A.; Arvidsson, P.I.; Sayed, Y.; Govender, P.; Maguire, G.E.; Kruger, H.G.; Govender, T. Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. *Eur. J. Med. Chem.,* **2011**, *46*(9), 3976-3985.
[215] Wang, Y.X.; Freedberg, D.I.; Yamazaki, T.; Wingfield, P.T.; Stahl, S.J.; Kaufman, J.D.; Kiso, Y.; Torchia, D.A. Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry,* **1996**, *35*(31), 9945-9950.
[216] Sitole, L.J.; Williams, A.A.; Meyer, D. Metabonomic analysis of HIV-infected biofluids. *Mol. Biosyst.,* **2013**, *9*(1), 18-28.
[217] Cornilescu, G.; Delaglio, F.; Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR,* **1999**, *13*(3), 289-302.
[218] Oliva, R.; Leone, M.; Falcigno, L.; D'Auria, G.; Dettin, M.; Scarinci, C.; Di Bello, C.; Paolillo, L. Structural investigation of the HIV-1 envelope glycoprotein gp160 cleavage site. *Chemistry,* **2002**, *8*(6), 1467-1473.
[219] Wang, M.; Quinn, C.M.; Perilla, J.R.; Zhang, H.; Shirra, R., Jr.; Hou, G.; Byeon, I.J.; Suiter, C.L.; Ablan, S.; Urano, E.; Nitz, T.J.; Aiken, C.; Freed, E.O.; Zhang, P.; Schulten, K.; Gronenborn, A.M.; Polenova, T. Quenching protein dynamics interferes with HIV capsid maturation. *Nat. Commun.,* **2017**, *8*(1), 1779.
[220] Mori, M.; Liu, D.; Kumar, S.; Huang, Z. NMR structures of anti-HIV D-peptides derived from the N-terminus of viral chemokine vMIP-II. *Biochem. Biophys. Res. Commun.,* **2005**, *335*(3), 651-658.
[221] Philippeos, C.; Steffens, F.E.; Meyer, D. Comparative 1H NMR-based metabonomic analysis of HIV-1 sera. *J. Biomol. NMR,* **2009**, *44*(3), 127-137.
[222] Riddler, S.A.; Li, X.H.; Otvos, J.; Post, W.; Palella, F.; Kingsley, L.; Visscher, B.; Jacobson, L.P.; Sharrett, A.R. Antiretroviral therapy is associated with an atherogenic lipoprotein phenotype among HIV-1-infected men in the multicenter AIDS cohort study. *J. Acquir. Immun. Defic. Syndr.,* **2008**, *48*(3), 281-288.
[223] Rodriguez-Gallego, E.; Gomez, J.; Pacheco, Y.M.; Peraire, J.; Vilades, C.; Beltran-Debon, R.; Mallol, R.; Lopez-Dupla, M.; Veloso, S.; Alba, V.; Blanco, J.; Canellas, N.; Rull, A.; Leal, M.; Correig, X.; Domingo, P.; Vidal, F. A baseline metabolomic signature is associated with immunological CD4(+) T-cell recovery after 36 months of antiretroviral therapy in HIV-infected patients. *Aids,* **2018**, *32*(5), 565-573.
[224] La, J.; Latham, C.F.; Tinetti, R.N.; Johnson, A.; Tyssen, D.; Huber, K.D.; Sluis-Cremer, N.; Simpson, J.S.; Headey, S.J.; Chalmers, D.K.; Tachedjian, G. Identification of mechanistically distinct inhibitors of HIV-1 reverse transcriptase through fragment screening. *Proc. Natl. Acad. Sci. U.S.A.,* **2015**, *112*(22), 6979-6984.
[225] Stewart, K.D.; Huth, J.R.; Ng, T.I.; McDaniel, K.; Hutchinson, R.N.; Stoll, V.S.; Mendoza, R.R.; Matayoshi, E.D.; Carrick, R.; Mo, H.M.; Severin, J.; Walter, K.; Richardson, P.L.; Barrett, L.W.; Meadows, R.; Anderson, S.; Kohlbrenner, W.; Maring, C.; Kempf, D.J.; Molla, A.; Olejniczak, E.T. Non-peptide entry inhibitors of HIV-1 that target the gp41 coiled coil pocket. *Biorg. Med. Chem. Lett.,* **2010**, *20*(2), 612-617.
[226] Stewart, K.D.; Steffy, K.; Harris, K.; Harlan, J.E.; Stoll, V.S.; Huth, J.R.; Walter, K.A.; Gramling-Evans, E.; Mendoza, R.R.; Severin, J.M.; Richardson, P.L.; Barrett, L.W.; Matayoshi, E.D.; Swift, K.M.; Betz, S.F.; Muchmore, S.W.; Kempf, D.J.; Molla, A. Design and characterization of an engineered gp41 protein from human immunodeficiency virus-1 as a tool for drug discovery. *J. Comput. Aided Mol. Des.,* **2007**, *21*(1-3), 121-130.
[227] Enriquez-Navas, P.M.; Chiodo, F.; Marradi, M.; Angulo, J.; Penades, S. STD NMR Study of the Interactions between Antibody 2G12 and Synthetic Oligomannosides that Mimic Selected Branches of gp120 Glycans. *ChemBioChem,* **2012**, *13*(9), 1357-1365.
[228] Enriquez-Navas, P.M.; Marradi, M.; Padro, D.; Angulo, J.; Penades, S. A solution NMR study of the interactions of oligomannosides and the anti-HIV-1 2G12 antibody reveals distinct binding modes for branched ligands. *Chemistry,* **2011**, *17*(5), 1547-1560.
[229] Davidson, A.; Begley, D.W.; Lau, C.; Varani, G. A Small-Molecule Probe Induces a Conformation in HIV TAR RNA Capable of Binding Drug-Like Fragments. *J. Mol. Biol.,* **2011**, *410*(5), 984-996.
[230] Mayer, M.; Lang, P.T.; Gerber, S.; Madrid, P.B.; Pinto, I.G.; Guy, R.K.; James, T.L. Synthesis and testing of a focused phenothiazine library for binding to HIV-1 TAR RNA. *Chem. Biol.,* **2006**, *13*(9), 993-1000.
[231] Evrard-Todeschi, N.; Gharbi-Benarous, J.; Bertho, G.; Coadou, G.; Megy, S.; Benarous, R.; Girault, J.P. NMR studies for identifying phosphopeptide ligands of the HIV-1 protein Vpu binding to the F-box protein beta-TrCP. *Peptides,* **2006**, *27*(1), 194-210.

[232] Xia, Y.M.; Bi, W.H.; Zhang, Y.Y. Synthesis of Dibenzylbutanediol Lignans and Their Anti-Hiv, Anti-Hsv, Anti-Tumor Activities. *J. Chil. Chem. Soc.,* **2009**, *54*(4), 428-431.

[233] Hwu, J.R.; Tseng, W.N.; Gnabre, J.; Giza, P.; Huang, R.C.C. Antiviral activities of methylated nordihydroguaiaretic acids. 1. Synthesis, structure identification, and inhibition of Tat-regulated HIV transactivation. *J. Med. Chem.,* **1998**, *41*(16), 2994-3000.

[234] Mahalingam, A.; Geonnotti, A.R.; Balzarini, J.; Kiser, P.F. Activity and Safety of Synthetic Lectins Based on Benzoboroxole-Functionalized Polymers for Inhibition of HIV Entry. *Mol. Pharm.,* **2011**, *8*(6), 2465-2475.

[235] Zhang, H.J.; Rumschlag-Booms, E.; Guan, Y.F.; Wang, D.Y.; Liu, K.L.; Li, W.F.; Nguyen, V.H.; Cuong, N.M.; Soejarto, D.D.; Fong, H.H.S.; Rong, L.J. Potent Inhibitor of Drug-Resistant HIV-1 Strains Identified from the Medicinal Plant Justicia gendarussa. *J. Nat. Prod.,* **2017**, *80*(6), 1798-1807.

[236] Kobayakawa, T.; Konno, K.; Ohashi, N.; Takahashi, K.; Masuda, A.; Yoshimura, K.; Harada, S.; Tamamura, H. Soluble-type small-molecule CD4 mimics as HIV entry inhibitors. *Bioorg. Med. Chem. Lett.,* **2019**, *29*(5), 719-723.

[237] Cushman, M.; Golebiewski, W.M.; Pommier, Y.; Mazumder, A.; Reymen, D.; De Clercq, E.; Graham, L.; Rice, W.G. Cosalane analogues with enhanced potencies as inhibitors of HIV-1 protease and integrase. *J. Med. Chem.,* **1995**, *38*(3), 443-452.

[238] Sechi, M.; Derudas, M.; Dallocchio, R.; Dessi, A.; Bacchi, A.; Sannia, L.; Carta, F.; Palomba, M.; Ragab, O.; Chan, C.; Shoemaker, R.; Sei, S.; Dayam, R.; Neamati, N. Design and synthesis of novel indole beta-diketo acid derivatives as HIV-1 integrase inhibitors. *J. Med. Chem.,* **2004**, *47*(21), 5298-5310.

[239] Ghosh, A.K.; Sean Fyvie, W.; Brindisi, M.; Steffey, M.; Agniswamy, J.; Wang, Y.F.; Aoki, M.; Amano, M.; Weber, I.T.; Mitsuya, H. Design, synthesis, X-ray studies, and biological evaluation of novel macrocyclic HIV-1 protease inhibitors involving the P1'-P2' ligands. *Bioorg. Med. Chem. Lett.,* **2017**, *27*(21), 4925-4931.

[240] Ali, A.; Reddy, G.S.; Nalam, M.N.; Anjum, S.G.; Cao, H.; Schiffer, C.A.; Rana, T.M. Structure-based design, synthesis, and structure-activity relationship studies of HIV-1 protease inhibitors incorporating phenyloxazolidinones. *J. Med. Chem.,* **2010**, *53*(21), 7699-7708.

[241] Ghosh, A.K.; Williams, J.N.; Kovela, S.; Takayama, J.; Simpson, H.M.; Walters, D.E.; Hattori, S.I.; Aoki, M.; Mitsuya, H. Potent HIV-1 protease inhibitors incorporating squaramide-derived P2 ligands: Design, synthesis, and biological evaluation. *Bioorg. Med. Chem. Lett.,* **2019**, *29*(18), 2565-2570.

[242] Makatini, M.M.; Petzold, K.; Alves, C.N.; Arvidsson, P.I.; Honarparvar, B.; Govender, P.; Govender, T.; Kruger, H.G.; Sayed, Y.; JeronimoLameira; Maguire, G.E.; Soliman, M.E. Synthesis, 2D-NMR and molecular modelling studies of pentacycloundecane lactam-peptides and peptoids as potential HIV-1 wild type C-SA protease inhibitors. *J. Enzyme Inhib. Med. Chem.,* **2013**, *28*(1), 78-88.

[243] Ali, A.; Reddy, G.S.; Cao, H.; Anjum, S.G.; Nalam, M.N.; Schiffer, C.A.; Rana, T.M. Discovery of HIV-1 protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel P2 ligands. *J. Med. Chem.,* **2006**, *49*(25), 7342-7356.

[244] Reddy, G.S.; Ali, A.; Nalam, M.N.; Anjum, S.G.; Cao, H.; Nathans, R.S.; Schiffer, C.A.; Rana, T.M. Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres. *J. Med. Chem.,* **2007**, *50*(18), 4316-4328.

[245] Olomola, T.O.; Klein, R.; Mautsa, N.; Sayed, Y.; Kaye, P.T. Synthesis and evaluation of coumarin derivatives as potential dual-action HIV-1 protease and reverse transcriptase inhibitors. *Bioorg. Med. Chem.,* **2013**, *21*(7), 1964-1971.

[246] Scala, A.; Piperno, A.; Micale, N.; Christ, F.; Debyser, Z. Synthesis and Anti-HIV Profile of a Novel Tetrahydroindazolylbenzamide Derivative Obtained by Oxazolone Chemistry. *ACS Med. Chem. Lett.,* **2019**, *10*(4), 398-401.

[247] Makki, M.S.I.; Abdel-Rahman, R.M.; Khan, K.A. Fluorine Substituted 1,2,4-Triazinones as Potential Anti-HIV-1 and CDK2 Inhibitors. *J. Chem.,* **2014**, 1-14.

[248] Thenin-Houssier, S.; de Vera, I.M.; Pedro-Rosa, L.; Brady, A.; Richard, A.; Konnick, B.; Opp, S.; Buffone, C.; Fuhrmann, J.; Kota, S.; Billack, B.; Pietka-Ottlik, M.; Tellinghuisen, T.; Choe, H.; Spicer, T.; Scampavia, L.; Diaz-Griffero, F.; Kojetin, D.J.; Valente, S.T. Ebselen, a Small-Molecule Capsid Inhibitor of HIV-1 Replication. *Antimicrob. Agents Chemother.,* **2016**, *60*(4), 2195-2208.

[249] Leydet, A.; Moullet, C.; Roque, J.P.; Witvrouw, M.; Pannecouque, C.; Andrei, G.; Snoeck, R.; Neyts, J.; Schols, D.; De Clercq, E. Polyanion inhibitors of HIV and other viruses. 7. Polyanionic compounds and polyzwitterionic compounds derived from cyclodextrins as inhibitors of HIV transmission. *J. Med. Chem.,* **1998**, *41*(25), 4927-4932.

[250] Desantis, J.; Massari, S.; Sosic, A.; Manfroni, G.; Cannalire, R.; Felicetti, T.; Pannecouque, C.; Gatto, C.; Tabarrini, O. Design and Synthesis of WM5 Analogues as HIV-1 TAR RNA Binders. *Open J. Med. Chem.,* **2019**, *13*, 16-28.

[251] Tok, T.T.; Tatar, G. Structures and Functions of Coronavirus Proteins: Molecular Modeling of Viral Nucleoprotein. *Int. J. Virol. Infect. Dis.,* **2017**, *2*(1), 001-007.

[252] Hulswit, R.J.G.; Lang, Y.F.; Bakkers, M.J.G.; Li, W.T.; Li, Z.S.; Schouten, A.; Ophorst, B.; van Kuppeveld, F.J.M.; Boons, G.J.; Bosch, B.J.; Huizinga, E.G.; de Groot, R.J. Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proc. Natl. Acad. Sci. U.S.A.,* **2019**, *116*(7), 2681-2690.

[253] Di Gennaro, F.; Pizzol, D.; Marotta, C.; Antunes, M.; Racalbuto, V.; Veronese, N.; Smith, L. Coronavirus Diseases (COVID-19) Current Status and Future Perspectives: A Narrative Review. *Int. J. Environ. Res. Public Health,* **2020**, *17*(8), 2690.

[254] Tortorici, M.A.; Walls, A.C.; Lang, Y.F.; Wang, C.Y.; Li, Z.S.; Koerhuis, D.; Boons, G.J.; Bosch, B.J.; Rey, F.A.; de Groot, R.J.; Veesler, D. Structural basis for human

coronavirus attachment to sialic acid receptors. *Nature Struct. Mol. Biol.,* **2019**, *26*(6), 481-489.
[255] Rabi, F.A.; Al Zoubi, M.S.; Kasasbeh, G.A.; Salameh, D.M.; Al-Nasser, A.D. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. *Pathogens,* **2020**, *9*(3), 231.
[256] Mihindukulasuriya, K.A.; Wu, G.; Leger, J.S.; Nordhausen, R.W.; Wang, D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.,* **2008**, *82*(10), 5084-5088.
[257] Bosch, B.J.; van der Zee, R.; de Haan, C.A.M.; Rottier, P.J.M. The coronavirus spike protein is a class I virus fusion protein: Structural and functional characterization of the fusion core complex. *J. Virol.,* **2003**, *77*(16), 8801-8811.
[258] Lim, Y.X.; Ng, Y.L.; Tam, J.P.; Liu, D.X. Human Coronaviruses: A Review of Virus-Host Interactions. *Diseases,* **2016**, *4*(3), 26.
[259] Li, Y.H.; Hu, C.Y.; Wu, N.P.; Yao, H.P.; Li, L.J. Molecular Characteristics, Functions, and Related Pathogenicity of MERS-CoV Proteins. *Engineering (Beijing),* **2019**, *5*(5), 940-947.
[260] Zhang, H.; Penninger, J.M.; Li, Y.; Zhong, N.; Slutsky, A.S. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive Care Med.,* **2020**, *46*(4), 586-590.
[261] Rabaan, A.A.; Alahmed, S.H.; Bazzi, A.M.; Alhani, H.M. A review of candidate therapies for Middle East respiratory syndrome from a molecular perspective. *J. Med. Microbiol.,* **2017**, *66*(9), 1261-1274.
[262] LabXchange Lifecycle of a Coronavirus. https://www.labxchange.org/library/pathway/lx-pathway:c0ed81ad-dd49-461e-9909-6f70e2762ce8/items/lx-pb:c0ed81ad-dd49-461e-9909-6f70e2762ce8:html:bf37be2c
[263] Song, Z.Q.; Xu, Y.F.; Bao, L.L.; Zhang, L.; Yu, P.; Qu, Y.J.; Zhu, H.; Zhao, W.J.; Han, Y.L.; Qin, C. From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses,* **2019**, *11*(1), 59.
[264] Tahir Ul Qamar, M.; Saleem, S.; Ashfaq, U.A.; Bari, A.; Anwar, F.; Alqahtani, S. Epitope-based peptide vaccine design and target site depiction against Middle East Respiratory Syndrome Coronavirus: an immune-informatics study. *J. Transl. Med.,* **2019**, *17*(1), 362.
[265] Sola, I.; Mateos-Gomez, P.A.; Almazan, F.; Zuniga, S.; Enjuanes, L. RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol.,* **2011**, *8*(2), 237-248.
[266] Vennema, H.; Godeke, G.J.; Rossen, J.W.; Voorhout, W.F.; Horzinek, M.C.; Opstelten, D.J.; Rottier, P.J. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.,* **1996**, *15*(8), 2020-2028.
[267] Kannan, S.; Hemalatha, K.; Masadeh, M.M. Evolving trends in molecular virulence, pathogenesis, symptomatology and epidemiology of recent Middle East respiratory syndrome coronavirus (MERS-CoV). *Infect. Dis. Trop. Med.,* **2016**, *2*(4), e347.
[268] World Health Organization. SARS (Severe Acute Respiratory Syndrome). https://www.who.int/ith/diseases/sars/en/
[269] World Health Organization. Health topics/Coronavirus. https://www.who.int/health-topics/coronavirus#tab=tab_3
[270] Huang, Y.P.; Cho, C.C.; Chang, C.F.; Hsu, C.H. NMR assignments of the macro domain from Middle East respiratory syndrome coronavirus (MERS-CoV). *Biomol. NMR Assign.,* **2016**, *10*(2), 245-248.
[271] Mahajan, M.; Bhattacharjya, S. NMR structures and localization of the potential fusion peptides and the pre-transmembrane region of SARS-CoV: Implications in membrane fusion. *Biochim. Biophys. Acta Biomembr.,* **2015**, *1848*(2), 721-730.
[272] Mahajan, M.; Chatterjee, D.; Bhuvaneswari, K.; Pillay, S.; Bhattacharjya, S. NMR structure and localization of a large fragment of the SARS-CoV fusion protein: Implications in viral cell fusion. *Biochim. Biophys. Acta Biomembr.,* **2018**, *1860*(2), 407-415.
[273] Hakansson-McReynolds, S.; Jiang, S.K.; Rong, L.J.; Caffrey, M. Solution structure of the severe acute respiratory syndrome-coronavirus heptad repeat 2 domain in the prefusion state. *J. Biol. Chem.,* **2006**, *281*(17), 11965-11971.
[274] Surya, W.; Li, Y.; Torres, J. Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.,* **2018**, *1860*(6), 1309-1317.
[275] Ghosh, A.; Bhattacharyya, D.; Bhunia, A. Structural insights of a self-assembling 9-residue peptide from the C-terminal tail of the SARS corona virus E-protein in DPC and SDS micelles: A combined high and low resolution spectroscopic study. *Biochim. Biophys. Acta Biomembr.,* **2018**, *1860*(2), 335-346.
[276] Takeda, M.; Chang, C.K.; Ikeya, T.; Guntert, P.; Chang, Y.H.; Hsu, Y.L.; Huang, T.H.; Kainosho, M. Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.,* **2008**, *380*(4), 608-622.
[277] Chang, C.K.; Sue, S.C.; Yu, T.H.; Hsieh, C.M.; Tsai, C.K.; Chiang, Y.C.; Lee, S.J.; Hsiao, H.H.; Wu, W.J.; Chang, C.F.; Huang, T.H. The dimer interface of the SARS coronavirus nucleocapsid protein adapts a porcine respiratory and reproductive syndrome virus-like structure. *FEBS Lett.,* **2005**, *579*(25), 5663-5668.
[278] Plant, E.P.; Perez-Alvarado, G.C.; Jacobs, J.L.; Mukhopadhyay, B.; Hennig, M.; Dinman, J.D. A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol.,* **2005**, *3*(6), e172.
[279] Ishimaru, D.; Plant, E.P.; Sims, A.C.; Yount, B.L., Jr.; Roth, B.M.; Eldho, N.V.; Perez-Alvarado, G.C.; Armbruster, D.W.; Baric, R.S.; Dinman, J.D.; Taylor, D.R.; Hennig, M. RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.,* **2013**, *41*(4), 2594-2608.
[280] Chatterjee, A.; Johnson, M.A.; Serrano, P.; Pedrini, B.; Joseph, J.S.; Neuman, B.W.; Saikatendu, K.; Buchmeier, M.J.; Kuhn, P.; Wuthrich, K. Nuclear magnetic resonance structure shows that the severe acute respiratory syndrome

coronavirus-unique domain contains a macrodomain fold. *J. Virol.,* **2009**, *83*(4), 1823-1836.
[281] Johnson, M.A.; Jaudzems, K.; Wuthrich, K. NMR Structure of the SARS-CoV Nonstructural Protein 7 in Solution at pH 6.5. *J. Mol. Biol.,* **2010**, *402*(4), 619-628.
[282] Li, Y.; Surya, W.; Claudine, S.; Torres, J. Structure of a conserved Golgi complex-targeting signal in coronavirus envelope proteins. *J. Biol. Chem.,* **2014**, *289*(18), 12535-12549.
[283] Berjanskii, M.V.; Neal, S.; Wishart, D.S. PREDITOR: a web server for predicting protein torsion angle restraints. *Nucleic Acids Res.,* **2006**, *34*(Web Server issue), W63-69.
[284] Hwang, T.L.; van Zijl, P.C.; Mori, S. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *J. Biomol. NMR,* **1998**, *11*(2), 221-226.
[285] Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR,* **2009**, *44*(4), 213-223.
[286] Dominguez, C.; Boelens, R.; Bonvin, A.M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.,* **2003**, *125*(7), 1731-1737.
[287] Kainosho, M.; Torizawa, T.; Iwashita, Y.; Terauchi, T.; Mei Ono, A.; Guntert, P. Optimal isotope labelling for NMR protein structure determinations. *Nature,* **2006**, *440*(7080), 52-57.
[288] Almeida, M.S.; Johnson, M.A.; Wuthrich, K. NMR assignment of the SARS-CoV protein nsp1. *J. Biomol. NMR,* **2006**, *36 Suppl 1*, 46.
[289] Almeida, M.S.; Johnson, M.A.; Herrmann, T.; Geralt, M.; Wuthrich, K. Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J. Virol.,* **2007**, *81*(7), 3151-3161.
[290] Chatterjee, A.; Johnson, M.A.; Serrano, P.; Pedrini, B.; Wuthrich, K. NMR assignment of the domain 513-651 from the SARS-CoV nonstructural protein nsp3. *Biomol. NMR Assign.,* **2007**, *1*(2), 191-194.
[291] Peti, W.; Johnson, M.A.; Herrmann, T.; Neuman, B.W.; Buchmeier, M.J.; Nelson, M.; Joseph, J.; Page, R.; Stevens, R.C.; Kuhn, P.; Wuthrich, K. Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7. *J. Virol.,* **2005**, *79*(20), 12905-12913.
[292] Lee, H.; Ren, J.; Pesavento, R.P.; Ojeda, I.; Rice, A.J.; Lv, H.; Kwon, Y.; Johnson, M.E. Identification and design of novel small molecule inhibitors against MERS-CoV papain-like protease via high-throughput screening and molecular modeling. *Bioorg. Med. Chem.,* **2019**, *27*(10), 1981-1989.
[293] Struck, A.W.; Axmann, M.; Pfefferle, S.; Drosten, C.; Meyer, B. A hexapeptide of the receptor-binding domain of SARS corona virus spike protein blocks viral entry into host cells via the human receptor ACE2. *Antiviral Res.,* **2012**, *94*(3), 288-296.
[294] Reina, J.J.; Diaz, I.; Nieto, P.M.; Campillo, N.E.; Paez, J.A.; Tabarani, G.; Fieschi, F.; Rojo, J. Docking, synthesis, and NMR studies of mannosyl trisaccharide ligands for DC-SIGN lectin. *Org. Biomol. Chem.,* **2008**, *6*(15), 2743-2754.
[295] Zaher, N.H.; Mostafa, M.I.; Altaher, A.Y. Design, synthesis and molecular docking of novel triazole derivatives as potential CoV helicase inhibitors. *Acta Pharm.,* **2020**, *70*(2), 145-159.
[296] Hoever, G.; Baltina, L.; Michaelis, M.; Kondratenko, R.; Baltina, L.; Tolstikov, G.A.; Doerr, H.W.; Cinatl, J., Jr. Antiviral activity of glycyrrhizic acid derivatives against SARS-coronavirus. *J. Med. Chem.,* **2005**, *48*(4), 1256-1259.
[297] Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E.J.; Liu, H.; Hilgenfeld, R. alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment. *J. Med. Chem.,* **2020**, *63*(9), 4562-4578.
[298] Chen, L.R.; Wang, Y.C.; Lin, Y.W.; Chou, S.Y.; Chen, S.F.; Liu, L.T.; Wu, Y.T.; Kuo, C.J.; Chen, T.S.; Juang, S.H. Synthesis and evaluation of isatin derivatives as effective SARS coronavirus 3CL protease inhibitors. *Bioorg. Med. Chem. Lett.,* **2005**, *15*(12), 3058-3062.
[299] Shie, J.J.; Fang, J.M.; Kuo, C.J.; Kuo, T.H.; Liang, P.H.; Huang, H.J.; Yang, W.B.; Lin, C.H.; Chen, J.L.; Wu, Y.T.; Wong, C.H. Discovery of potent anilide inhibitors against the severe acute respiratory syndrome 3CL protease. *J. Med. Chem.,* **2005**, *48*(13), 4469-4473.
[300] Ding, Y.; Zhu, X.; Hao, L.; Zhao, M.; Hua, Q.; An, F. Bioactive Indolyl Diketopiperazines from the Marine Derived Endophytic Aspergillus versicolor DY180635. *Mar. Drugs,* **2020**, *18*(7), E338.
[301] Ozkan, H.; Adem, S. Synthesis, Spectroscopic Characterizations of Novel Norcantharimides, Their ADME Properties and Docking Studies Against COVID-19 M(pr) degrees. *ChemistrySelect,* **2020**, *5*(18), 5422-5428.
[302] Kantsadi, A.L.; Vakonakis, I. Rapid assessment of ligand binding to the SARS-CoV-2 main protease by saturation transfer difference NMR spectroscopy. *bioRxiv,* **2020**.
[303] Wecker, K.; Morellet, N.; Bouaziz, S.; Roques, B.P. NMR structure of the HIV-1 regulatory protein Vpr in H2O/trifluoroethanol. Comparison with the Vpr N-terminal (1-51) and C-terminal (52-96) domains. *Eur. J. Biochem.,* **2002**, *269*(15), 3779-3788.