THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY (PHD)

The roles and interactions of viral- and host proteins in the life cycle of human immunodeficiency virus type 1

by Ferenc Tóth

Supervisor: Prof. Dr. József Tőzsér, PhD, DSc

UNIVERSITY OF DEBRECEN
DOCTORAL SCHOOL OF MOLECULAR CELL AND IMMUNE BIOLOGY

DEBRECEN, 2019

The roles and interactions of viral- and host proteins in the life cycle of human immunodeficiency virus type 1

By Ferenc Tóth, molecular biologist

Supervisor: Prof. Dr. József Tőzsér, PhD, DSc

Doctoral School of Molecular Cell and Immune Biology, University of Debrecen

Examination Committee

Head: Prof. Dr. Gábor Szabó, PhD, DSc

Members: Prof. Dr. János Minárovits, PhD, DSc
Prof. Dr. Péter Bay PhD, DSc

The Examination takes place at the 3.201-3.202 Discussion Room of the Department of Biochemistry and Molecular Biology Faculty of Medicine, University of Debrecen at 11 a.m., 22nd of January, 2020.

Defense Committee:

Head: Prof. Dr. Gábor Szabó, PhD, DSc

Reviewers: Prof Dr. Péter Gál, PhD, DSc
Dr. Endre Kókai, PhD

Members: Prof. Dr. János Minárovits, PhD, DSc
Dr. Péter Bay, PhD, DSc

The PhD Defense takes place at the Lecture Hall of Bldg. A of the Department of Internal Medicine, Faculty of Medicine, University of Debrecen at 1 p.m., 22nd of January, 2020.

1. INTRODUCTION

1.1 Retroviruses:

Retroviruses belong to the Retroviridae virus family, members of which have envelops and diploid RNA genome. Taxonomy of Viruses (ICTV) regrouped Retroviruses into the order Ortervirales. Members of Orterviruses share several features; such as (i) a homologous aspartate protease domain in their polymerase polyprotein, (ii) the use of host tRNA molecules as primers for genome replication, and (iii) the presence of capsid and nucleocapsid proteins in the virions. The Retroviridae family contains the Orthoretrovirinae and Spumaretrovirinae subfamilies. The best-known member of the Orthoretrovirinae subfamily is the human immunodeficiency virus from the lentivirus genus, the causative agent of the acquired immunodeficiency syndrome (AIDS).

Retroviral virions are 80-100 nm in diameter, covered by a lipid envelop (Env) which contains the viral glycoproteins. The viral core is built from matrix (MA) and capsid (CA) proteins, and its shape is characteristic for various genera of the family. Inside the core, we can find the RNA genome associated with the nucleocapsid (NC) protein. Their genome is 7-12 kb long, linear, nonsegmented and of positive polarity. The replication strategy of the family is also characteristic, where DNA intermedier from the viral RNA genome is integrating into the host genome.

Based on organization of the viral genom, we can classify them into simple and complex retroviruses. Every retrovirus contains gene segments coding for 3 polyproteins: the gag which codes structural proteins like matrix (MA), capsid (CA), nucleocapsid; the pol codes the reverse transcriptase (RT), and integrase (IN) enzymes responsible for replication and integration, respectively, and the protease (PR) enzyme, responsible for the processing of polyproteins translated from the aforementioned genes; and env which constitutes the surface (SU) and transmembrane (TM) proteins of the viral envelop. Simple retroviruses contains only these basic genes, while complex retroviruses; such as HIV-1, carries additional genes, which contribute to the control of gene expression (Tat, Rev), successful infection (Nef), survival in the host cell (Vif) and budding of the

virions at the end of the late phase of viral infection (Vpu), primarily through the interaction with the host proteins.

1.2. The retroviral life cycle:

Replication for retroviruses can be divided into two different sections: early- and late phase. The first step of the life cycle of HIV-1 is the viral entry through receptor-mediated endocytosis or direct fusion. In the cytoplasm, viral RNA is transcribed into proviral DNA by the reverse transcriptase enzyme inside the viral core. The preintegration complex (PIC); which is composed of the proviral DNA and the proteins remaining in the viral core, enters the nucleus, where the proviral DNA integrates into the host genome. This is the final step of the early phase of the infection.

The late phase starts with the transcription of viral DNA into messenger RNA by the cellular protein RNA polymerase II enzyme. mRNA coded by the essential (gag, pol, and env) and/or other additional genes is translated into proteins. mRNA transcribed from the gag and pol genes serve as genomic RNA of the newly formed virions or translated to Gag and Gag-Pro-Pol polyproteins. The polyprotein coded by the env gene is translated from a shorter RNA is glycosylated and processed by a cellular proteinase into individual subunits (SU and TM). The primary translation products Gag and Gag-Pro-Pol polyproteins together with the viral genom, are assembled at the cell periphery into a toroid like, immature viral particle, which is enclosed by the viral envelop. Viruses are released from the cell by budding of the cell membrane, which is followed by the proteolytic processing of the polyproteins into smaller functional fragments forming the mature, infectious virions.

1.3. Properties of the HIV-1 capsid protein:

The capsid protein is composed of 231 amino acid residues, and consists of two independent domains: an N-terminal core domain (residues 1–145), and a C-terminal dimerization domain (residues 151–231), that are connected by a short linker region. The C-terminal domain also contains the evolutionarily conserved major homology region (MHR), which is crucial to the maintenance of the structure and interactions with other viral and cellular components. The CA is a product of the *gag* gene, and is cleaved from the Gag polyprotein by the viral PR. It has an important role in the assembly of the mature virion, and the early phase of infection as well. The proteolytic processing of the Gag

polyprotein leads to the appearance of the mature virion containing the conal shaped core structure. In mature virions, approximately 1500 CA monomers form a conical structure, built from approximately 250 hexamers and 12 pentamers.

The proper stability of the monomer CA is crucial to HIV-1 infection, as it influences the decapsidation process in the early phase, where the exact identity of the factors involved is still not fully resolved.

2.4. Properties of HIV-1 protease:

The PR is synthesized as part of the Gag-Pol polyprotein, consisting of 99 residues. It shows characteristic features typical of cellular aspartic proteases; such as inhibition by pepstatin and enzyme inactivation by mutation of the catalytic aspartate residue. However, unlike cellular aspartic proteases, it is built from two identical subunits and works as a dimer. The first and secondary structure are homologous to one domain of the cellular aspartic proteases, it contains many β-sheets and 1 α-helix. The N- and C-terminal domains of the subunits together form a four-stranded antiparallel β-sheet.

The three specific regions of the HIV-1 protease are the active site, the flap region and, the dimerization region. The catalytic triad (Asp-Thr-Gly) forming the active site is found close to the N-terminal end, its residues are connected to each other through a network of hydrogen bonds called the fireman's grip. The flexible flap region bends to the substrate or inhibitor during the formation of the enzyme-substrate complex, stabilizing the association. The third, a conservative region (Gly-Arg-Asn) is found near the C-terminal end, and plays an important role in dimerization.

The pH optimum of the HIV-1 PR is acidic (4,5-6,5), its cleavage sites have been identified using viral polyproteins and N- and C-terminal sequencing of the mature proteins. However, comparison of these identified cleavage sites showed no consensus sequence for cleavage, these sequences are quite hydrophobic. HIV-1 PR requires at least a 7 residues long peptide for recognition and successful cleavage, which has to be bound to the enzyme in an extended beta conformation.

Although the enzyme works as a homodimer, the sequence of the identified cleavage sites does not show symmetrical amino acid residue distribution, and the enzyme has no preference for symmetrical substrates either. Its functioning shows remarkable cleavage site dependency, where the most important features are the size of the amino acid residue, and the presence of the β-branch, at the same time, residues in the P1 and P1' position proved to be critical for cleavage. For this purpose, in our study of CA, we decided to modify these residues of the cleavage sites, and examine the effect on proteolytic cleavage

1.4. The role of PR in the early phase of infection:

In addition to other viral proteins, the protease (PR) is also part of the core and, therefore, it might contribute to the viral infectivity by further processing viral proteins, or by cleaving cellular ones. In the late phase of the viral life cycle, proteolytic cleavage of Gag and Gag-Pol polyproteins is essential to produce the mature, infectious virus. *In vitro* experiments with the cores of another lentivirus; the equine infectious anemia virus, showed that the viral protease can cleave the CA and NC protein into smaller fragments *in vitro*, leading to the suggestion of an early phase role for the PR. Protease inhibition studies provided controversial results, some showing an inhibitory effect on the early phase, while others did not find such an effect. HIV-1 entry occurs in most cases by direct fusion, but the endocytotic route could make a substantial contribution to infectivity depending on the virus isolate and cell type. Although the PR has been shown to be able to process proteins at the neutral pH of the cytoplasm, as the pH optimum of the PR is acidic, the acidic pH of the lysosome is more favorable for its proteolytic action.

The processed CA protein have also been found to be a substrate for the retroviral PR in vitro, and its processing is predicted to be dependent on a pH-induced conformational change. Based on N-terminal sequencing, as well as MS identification of the fragments of recombinant CA separated by SDS/PAGE, two major CA cleavage sites were identified: Ala77/Ala78 and Leu189/Leu190 12. In addition to these sites, another study had identified additional cleavage sites: Ala22/Trp23, Gly116/Trp117, and Ala204/Leu205. All these cleavage sites are located outside the MHR region, which makes them

appropriate to establish mutant CA which has altered proteolytic susceptibility without structural changes. It may also provide an adequate tool for examining the interaction of the two proteins in the early phase of infection.

1.5. The role of capsid interacting proteins in the early phase of infection.

The peptidyl-prolyl isomerase cyclophilin A (CypA) has been shown to bind the Gag protein of HIV-1. It also has been shown to be incorporated into the virions, and its incorporation was found to be essential for viral infectivity. However, later experiments proved that CypA present in cells is more important for infectivity. CypA binds to the CypA-binding loop of the CA, but its precise role in the early phase of infection is still unclear. Over the past decade, genome-wide siRNA screenings and proteomic analyses identified several cellular proteins which are important for HIV-1 infection. Some of these factors like „transportin-3" (TNPO3), a „cellular protein cleavage and polyadenylation specificity factor subunit 6" (CPSF6) or „nuclear pore complex protein 153" (NUP153) and 358 (NUP358 or RunBP2) have been also found to be interaction partners of the CA protein.

Event though TNPO3 can bind to the CA, it acts indirectly through separating CPSF6 protein, affecting the integration of the viral genome.

CPSF6 is also able to bind to the CA, and is persumed to have a role in several steps of the early phase of infection; such as nuclear transport and integration.

NUP153 and NUP358 nucleoporin proteins are essential components of the nuclear pore complex. NUP358 is located at the cytoplasmic side of the pores, near to the point of origin of the cytoplasmic filaments, and has a critical role in the nuclear translocation of the PIC. NUP153 is located at the nuclear side of the pores, it can also bind the CA with another nucleoporin protein NUP93. Its role in HIV-1 replication was corroborated several times. Moreover, a recent study also proved it to be essential for infecting non-dividing cells.

Besides the proteins presented, there are hundreds which were revealed to have a role in HIV-1 replication, however, for most of them, the exact role and connection point to the life cycle remains unclear. The thorough study of the effects of proteins on HIV-1 replication can contribute significantly to the developement of new anti-HIV-1 therapies.

2. OBJECTIVES

The most critical event in the early phase of HIV-1 infection is decapsidation. Acceleration or inhibition of the process or both can terminate the infection. The identification of proteins involved in HIV-1 replication and the study of their roles can contribute significantly to the better knowledge of the viral life cycle, and thereby, to the appearance of new anti-HIV-1 therapies or more effective vectors for gene therapy applications.

Our aims were:

1. Based on previous specificity studies of the viral protease, to introduce mutations to the cleavage sites of HIV-1 capsid protein (W23A, A77P, A78V, L189F, L189I és L189P) which can alter its proteolytic susceptibility. We planned to study the effect of these mutations on the secondary structure of the proteins, their interaction with cyclophilin A, and their proteolytic susceptibility to HIV-1 PR and trypsin. The aim was to identify appropriate candidate mutations that can be used to clarify the potential role of capsid processing at the early phase of viral infection, especially in circumstances where the virus or viral-based vectors are utilizing the receptor-mediated endocytotic route for infection, where the pH and the CA structure is more favorable for proteolytic processing.

2. Hundreds of genes involved in viral transduction were identified recently. A proteomic study extended to several gene products can narrow down the number of targets to be examined. Our goal was to identify host proteins which are possibly involved in the regulation of the early events of pseudovirus transduction, with the utilization of an HIV-1 pseudovirus and label-free proteomics.

3. MATERIALS AND METHODS

3.1. Prediction of effects of mutations

Secondary structure prediction was performed using SOPMA secondary structure prediction server. Effects of the mutations on protein stability were predicted by the Site-Directed Mutator (SDM) server using the crystal structure of the wild-type full-length HIV-1 capsid protein (PDB code: 3NTE) (János Mótyán's work).

3.2. Mutagenesis, expression, and purification of the recombinant HIV-1 capsid protein

The plasmid bearing the CA protein from HIVIIIB isolate with an N-terminal 6-his tag (His_6-HIVCA) was obtained from Dr. Carol Carter (Department of Molecular Genetics and Microbiology, S.U.N.Y. Stony Brook, USA). Mutations were introduced with the QuikChange Site-Directed Mutagenesis Kit.

Protein expression in Escherichia coli BL21(DE3) cells were induced by the addition of 1 mM IPTG to the cultures followed by incubation at 37 °C for 3 h. Bacterial cells were suspended in lysis buffer and then disrupted by sonication. The insoluble fraction was collected by centrifugation and resuspended in denaturation buffer. After sonication, the solution was centrifuged, then the supernatant was filtered and applied to the Ni-nitrilotriacetic acid Superflow affinity resin. The purified protein was dialyzed and concentrated. The concentrated protein was fractionated by gel filtration on Superose 12 10/300 GL column at a flow rate of 1 ml/min at room temperature. Fractions with the proper purity were dialyzed and concentrated again with the same method mentioned above. The purity of samples was verified by SDS/PAGE using 16% polyacrylamide gels.

3.3. Circular dichroism spectroscopy

Circular dichroism spectra were recorded on a Jasco-810 spectropolarimeter at room temperature in 10 mm sodium phosphate buffer (pH 7.5). CD deconvolutions were obtained using the CDSSTR analysis program, kindly provided by Dichroweb.

3.4. Expression and purification of the recombinant HIV-1 protease

The plasmid coding for this protease was a kind gift of Dr. John M. Louis (Laboratory of Chemical Physics, NIDDK, NIH). The PR was expressed in *E. coli* BL21(DE3) cells. Cells were induced for expression with 1 mM IPTG for 3 h, then harvested by centrifugation. After removal of the supernatant, the cell pellets were suspended and lysed in the presence of 100 μg/ml lysozyme by sonication on ice. The lysate was centrifuged, pelleted inclusion bodies were washed and centrifuged again. The final pellet was dissolved in a buffer containing 7.5 M guanidine-HCl, filtered, then applied to a Superose 12 10/300 GL column (GE Healthcare). The separation was performed at a flow rate of 0.5 ml/min at room temperature. Peak fractions were pooled and subjected to reversed-phase HPLC on a POROS 20 R2 column. The purity of selected fractions was assessed by SDS/PAGE using 16% polyacrylamide gels. The protein was folded by two-step dialysis.

3.5. Proteolytic digestion of recombinant CA proteins by HIV-1 protease

Recombinant CA proteins were incubated with PR at PR: CA ratio of 1:20 at pH 5.5. The reaction mixture containing recombinant His6-tagged CA protein (10 μm final concentration) was incubated with recombinant HIV-1 PR (0.5 μm final concentration) in PR reaction buffer. After incubation at 37 °C for 4 h, reactions were stopped by the addition of an equal volume of 2X tricine-SDS sample buffer containing β-mercaptoethanol, and the samples were analyzed using 16 % gradient SDS/PAGE tricine-buffered gels. In the case of samples analyzed later by mass spectrometry, the reactions were stopped by the addition of an equal volume of concentrated formic acid.

3.6. Identification of proteolytic fragments of CA by mass spectrometry

Before the measurements, samples were purified by ZipTip pipette tips loaded by C18 resin. The (MALDI-TOF MS) analysis of the CA fragments was performed by a Voyager-DETM PRO MALDI-TOF mass spectrometer; sinapinic acid was used as matrix. The instrument was used in the linear mode of operation. Results were evaluated by Data Explorer software.

3.7. Examination of the proteolytic susceptibility of recombinant capsid with limited proteolysis by trypsin

Recombinant CA proteins were incubated with trypsin at trypsin: CA ratio of 1:100 at pH 7.5 in a reaction buffer. Reaction mixtures contained recombinant

His-tagged CA protein (10 μm final concentration) and trypsin (100 nm final concentration). Samples were incubated at 37 °C for 2 h, reactions were stopped by the addition of 6X SDS sample buffer containing β-mercaptoethanol. The samples were analyzed using 16% SDS/PAGE gels.

3.8. Construction of pGEX-4T-3-CypA expression vector

The cyclophilin A coding region was amplified from the PPIA (NM_021130) human cDNA ORF Clone using Phusion High-Fidelity DNA polymerase. The forward and reverse primers harbored recognition sites of BamHI and NotI restriction endonucleases, respectively. After the cleavage of the PCR product by BamHI and NotI, it was ligated into a pGEX-4T-3 plasmid.

3.9. Expression and purification of CypA-GST fusion protein

The GST-CypA coding plasmid was transformed into *E. coli* BL21(DE3) cells and the expression was induced by the addition of 0.1 mM IPTG, and the cultures were incubated at 37 °C for 3 h. After harvesting the cells by centrifugation, the pellets were treated as described previously. The clear lysate of cells was then applied to a Bio-ScaleTM Mini Profinity^{TM} GST cartridge. The purified protein was dialyzed overnight at 4 °C, against MQ water, and then concentrated with a lyophilizer. Concentrated proteins were dissolved in MQ water again. The purity of samples was verified by SDS/PAGE using 16 % polyacrylamide gels.

3.10. His_6-HIVCA pull-down assay

After binding of the His_6-HIVCA protein to the Ni-nitrilotriacetic acid magnetic agarose beads (Qiagen) at 4 °C for 1 h using a PBS supplemented with 20 mm imidazole, 0.05 % Tween20, pH 8.0), the mixture was incubated with the purified GST-CypA protein in PBS at 4 °C for 1 h. The non-bound material was washed off and the bound complexes were eluted, followed by SDS/PAGE using 14 % polyacrylamide gels.

3.11. Modification of the pMDLg/pRRE plasmid of the HIV-1 vector system

The original HIV-1 vector system was a kind gift from Dr. Didier Trono, (Department of Genetics and Microbiology, University of Geneva Medical School, Geneva, Switzerland). The capsid coding fragment from the pMDLg/pRRE plasmid was cut out by AvrII and PmlI restriction enzymes and ligated between the corresponding restriction sites of the pT7 Blue-3 plasmid.

The mutations were introduced as written in section 3.2. After verification, the mutated fragments were ligated back into the pMDLg/pRRE plasmid.

3.12. Production of viral particles

Viruses were produced by transient transfection of 293FT cells using polyethyleneimine. The conditioned media containing virus particles were collected after 24, 48 and 72 h, clarified by centrifugation, filtered, concentrated, and stored at −70°C. The amount of the produced pseudovirus were determined from the concentration of the capsid protein (p24) or the activity of the reverse transcriptase enzyme.

3.13. Infectivity assay

For infection, 293T cells were seeded into 24-well plates and transduced by pseudovirus particles containing 10 ng capsid protein at 50 % confluency. On the seventh day, cells were counted by flow cytometer to determine the number of GFP-positive cells. Data were evaluated by FlowJo flow cytometry analysis software.

3.14. Statistical analysis

Statistical analysis was performed using GraphPad QuickCalcs free web calculator.

3.15. Transduction and sample collection for mass spectrometry (MS)

293 T cells in T-25 cell culture flasks were either mock-treated or transduced at 50 % confluency with 5 ng RT equivalent of HIV-based pseudovirions, in the presence of 4 μg/ml polybrene (Sigma-Aldrich), in 1 ml total volume, and incubated at 37 C°. After 0, 4, and 12 h, cells were trypsinized for 10 min, then washed three times with ice-cold PBS to remove non-fused pseudovirion particles. The final pellet was suspended in a lysis buffer, supplemented with cOmplete protease inhibitor cocktail, incubated for 30 min at room temperature, centrifuged, and the supernatant was mixed with cold (− 20 C°) acetone and stored at − 20 C° overnight.

3.16. Mass spectrometry analysis of the protein content of the transduced cells

After acetone precipitation, precipitates were re-dissolved in 25 mM ammonium bicarbonate and digested in-solution with trypsin [50]. The tryptic fragments were used for replicate label-free LC-MS/MS analyses by our

collaboration partner George Tsaprailis (University of Arizona in Tucson, AZ, USA). The protein and peptide identification results were validated with Scaffold v4.4.6. software. Peptide sequences of the identified proteins were validated again by Protein BLAST (Basic Local Alignment Search Tool) program. Protein quantification was done based on spectral counting; the quantitative values were generated by the Scaffold program based on the normalized total spectra.

3.17. Statistical analysis of proteomics data

For both statistical and network analysis, we used in-house developed R-software (work of Miklós Emri) based on STRING, circlize, MASS, lsmeans, matrixStats, reshape2 and ggplot2 packages. Assuming that data from technical repetitions are often characterized by Poisson distribution, and the large variances of biological replicas can be modeled by negative binomial distribution, we used modified general linear models to describe group-level differences in measured protein data in the 4 and 12 h time points.

4. RESULTS AND DISCUSSION

4.1. *In silico* secondary structure analysis of the capsid protein

Effects of mutations on the protein structure were predicted using the SDM server, which calculates a stability score analogous to the free energy difference between wild-type and mutant protein The highest values were predicted for the A77P and L189P mutants, which suggested that these mutations have a highly destabilizing effect on the protein structure and may potentially cause protein malfunction. The L189I mutation was predicted to be neutral, while A78V and L189F mutations were predicted to be slightly destabilizing and cause only a slight deviation of the secondary structural organization. Secondary structure prediction did not predict changes for W23A mutation, furthermore, SDM server predicted only a slightly destabilizing effect for this mutation, as well.

4.2. Cleavage of wild-type and mutant HIV-1 CA by HIV-1 PR

To define the effects of proteolytic cleavage site mutations on *in vitro* CA processing, we prepared mutant capsid proteins in which the P1 (A77P, L189F, L189I, and L189P mutants) or the P1′ (W23A and A78V mutants) amino acids of the cleavage sites were replaced.

As expected, A77P and L189P amino acid changes inhibited the proteolytic processing at these sites, as we had observed a decrease in the product band intensities or disappearance of the fragments related to these modified cleavage sites. On the other hand, the W23A mutation resulted in the accelerated proteolytic cleavage of the protein. The modified cleavage sites became more susceptible to proteolytic cleavage in the case of A78V and L189F mutants, as indicated by the increased amount of the related fragments (78–231 and 1–189, respectively). However, the overall processing of the L189F CA was reduced, and A78V CA did not show significant alteration in its proteolytic susceptibility compared to that of the wild-type. Surprisingly, L189I mutation did not prevent processing at this site, although P1 Ile was not found in HIV-1 cleavage sites. This might be due to a one-residue shift of the cleavage site to TETIL/VQNAN to produce a typical type 2 cleavage site for HIV-1 PR, we, however, were unable to prove this as none of the respective peptides were recovered in MALDI-TOF experiments.

4.3. CD spectroscopy analyses of the secondary structure of wild-type and mutant CA

To predict the effect of cleavage site mutations on the overall CA structure, we examined the secondary structure of the proteins using CD spectroscopy. The molecular ellipticities of the spectra of the proteins were examined in the range of 180–300 nm, then, the spectra of mutant CA proteins were compared to that of the wild-type CA. CD spectra were analyzed by the CDSSTR analysis program. These spectra were measured at pH 7.5 in which the CA protein maintains its structural integrity.

The spectrum of the wild-type CA protein showed positive ellipticity with a maximum at 190 nm, intense negative ellipticity with a distinct minimum at 208 nm, and a shoulder at 220 nm. The crossover point for the spectra was at 200 nm. The distribution of the secondary structural elements is in good agreement with results published previously for the His_6-tagged CA at pH 7.5, and with the predictions made from the amino acid sequence of the wild-type protein. The spectra of A78V and L189F mutants showed similar features to that of the wild-type, indicating that these proteins have the same distribution of secondary structural elements. The spectrum of L189I mutant also exhibited similar features with a higher maximum at 191 nm, suggesting that its structure only slightly differs from that of the wild-type protein in its α-helix ratio. The spectra of the W23A, A77P, and L189P mutants exhibited several different features compared to that of the wild-type, such as highly reduced positive and negative ellipticity and altered crossover points of the curves, indicating the variable distribution of secondary structures.

4.4. Tryptic digestion of wild-type and mutant HIV-1 CA

The accessibility of tryptic cleavage sites was was also studied to assess structural differences between wild-type and mutant CA proteins. An equal amount of proteins was subjected to limited proteolysis with trypsin at the same neutral pH (pH 7.5) that was used for CD spectroscopy, tryptic digestion was then followed by SDS/PAGE analysis.

In the case of the CA proteins bearing the A78V, L189F, or L189P mutations, the efficiency of proteolytic digestion was found to be similar to that of the wild-type, no significant difference was observed. However, in the case of the W23A, A77P, and L189I mutants the percentage of the uncleaved CA

showed significant reduction, suggesting that the W23A, A77P, and L189I mutant proteins are more sensitive to trypsin, compared to the wild-type capsid protein. Therefore, it is expected that these mutant proteins could form aberrant core structures.

4.5. Examination of the interaction of His_6-HIVCA and CypA-GST recombinant proteins by pull-down assay

Binding of CypA to the capsid protein is crucial to the proper decapsidation process in the early phase of infection. We have tested whether mutations of CA protein may affect the CA-CypA interaction and studied the binding efficiency *in vitro*. Purified His_6-HIVCA and GST-CypA proteins were subjected to a His6-HIVCA pull-down assay followed by the analysis of the eluted complexes by SDS/PAGE.

The amount of CypA associated with the W23A, A78V, L189F, L189I, and L189P mutants did not change significantly, while A77P mutant showed a highly reduced CypA binding ability compared to that of the wild-type CA protein, which was an unexpected result as A77P falls outside of the Cyp binding region. Based on these results, the introduced mutations (except for A77P) leading to changes in the secondary structural organization did not affect the CypA-binding ability of the CA protein.

4.6. Examination of the infectivity of pseudovirions carrying wild-type and L189F mutant CA

293T cells were transfected by plasmids coding for the wild-type and L189F capsid in different proportions (1:0, 1:1 and 0:1). 293T cells were transduced with pseudoviruses containing an equal amount of p24 protein and after 1 week, the number of GFP positive cells was determined by flow cytometry. For the control transduction, we used pseudoviruses that were assembled using salmon sperm DNA beside the plasmid coding for the wild type CA. Transduction efficiency of the pseudovirions was 32 % as measured by GFP fluorescence. Increasing the amount of the plasmid coding for the L189F mutant protein reduced the number of the infected cells to 28% when used in a 1:1 ratio. When we solely used the L189F mutant coding plasmid, this number was reduced to 2 %.

As our examinations showed, the L189F mutation did not affect the structure of CA monomers at a pH similar to the cytoplasm, hence, we can

assume that the reduction of infectivity of the pseudoviral particles containing mutant CA is caused by the altered proteolytic susceptibility. It should be noted that these are only preliminary results, to verify this hypothesis and the role of the HIV-1 PR in the early phase of infection, further *in vitro* experiments are necessary.

4.7. Mass spectrometric study of protein level changes caused by transduction with a lentiviral vector

In this study we examined changes in the protein content of host cells as a result of transduction by pseudovirions in the early phase of infection. We examined the proteome changes with the use of label-free mass spectrometry.

The list of proteins identified from analysis of the mass spectrometry data were manually curated, and in the case of non-human or non-viral identifications, the sequences were verified. The origin of mistakenly designated non-human proteins was corrected, and proteins not matched to any of the human or viral proteins were omitted from further analyses.

4.8. Statistical analysis of mass spectrometry data

After the qualitative analysis, only those proteins were considered for statistical analysis which could be quantified in at least 2 out of 4 replicates, and were not quantified in other conditions. With this method, we identified 25 proteins in the transduced cells, amount of which differed from the control in any time point used for sample collection.

HIST1H1E, HNRNPL, PRRC2A, and TRIM28 were quantified only 4 hours after transduction. CSDA, EEF1A1, EEF1D, HN1, NPM1, PGAM1, and SRSF6 increased significantly, while the amount of HIST1H1D and HSPA5 significantly decreased compared to the control and 0 time point.

Some of the proteins were quantified in all time points except after 12 hours of transduction. These include ALYREF, CCDC86, CSDA, COX5A, HN1, MYL6, PPIF, SEPT2, SRSF6, TCOF1, and TPM3.

The disappearance of these proteins suggests that the pseudovirus particles controlled protein expression of the cell. The amount of some protein (CSDA, HN1, and SRSF6) showed increased expression after 4 hours of transduction, but 8 hours later, we were not able to quantify them in the samples. 12 hours after transduction, the amount of COX6B1 and PDIA3 increased, while that of EEF2 and GAPDH decreased significantly.

When the function of proteins showing statistically significant changes was examined, we observed an increase in the number of proteins implicated in RNA binding 4 hours after transduction, and an overall decrease in their amount 12 hours after transduction. Beside the change of gene expression, the stabilizing effect of the interaction between viral- and host proteins, as well as the reduction of the rate of protein turnover, may cause a temporary increase in the amount of proteins.

4.9. The interactions of host proteins with HIV-1 proteins during early events of its life cycle

Replication of HIV-1 highly depends on the proteins of the host cell machinery, in our study, we aimed to identify host proteins which are possibly involved in the regulation of the early events of pseudovirus transduction, with the utilization of an HIV-1 pseudovirus and label-free proteomics. In the literature, all identified proteins were connected to the life cycle HIV-1, with only one exception (HN1), however, in many cases their exact role is still unclear.

EEF1A1 and EEF1D proteins are parts of the EEF1 complex, which facilitates reverse transcription in eukaryotic cells. Their role in the life cycle confirmed elsewhere. However, the binding of EEF1D is not yet verified, as part of the EEF1 complex, it probably affects the process through the cooperation with EEF1A1.

EEF2 binds to Gag, and this connection can be stabilized by CypA, their complex is able to inhibit the assembly of stress granules which is an important factor in defense against infections. The lack of EEF2 reduces the number of produced virions and decreases their infectivity.

The role of MYL6 in the life cycle of HIV-1 was already known, it is also able to bind to Gag, but the relevance of this interaction is not yet understood.

NPM1 and ALYREF can bind to the viral Rev protein. Supposedly, NPM1 supports the nuclear import of Rev, while ALYREF bound to Rev is unable to bind RNA, explaining how Rev possibly favours the export of viral RNA to cellular RNA.

The roles of COX5A, COX6B1, and PDIA3 in the replication of HIV-1 and their interaction with viral envelope proteins is already known, but as the

viral envelope protein is not present in the pseudovirus, their connections to the virus and their role in HIV-1 life cycle remains to be elucidated.

SRSF6 protein is possibly inhibits viral gene expression by the splicing of the Tat coding mRNA. Its interaction with other viral protein coding mRNA is not yet known.

TRIM28 is also a well-known factor in HIV-1 life cycle, a present study suggests that it has a crucial role in the inhibition of viral gene expression. It is probably an important factor in the cell's defense, which is supported by its emergence after 4 hours of transduction and their disappearance 8 hours later.

HIST1H1E, HNRPL, CCDC86, CSDA, PPIF, SEPT2, TCOF1, TPM3, HIST1H1D, PGAM1, HSDA5 and GAPDH are also identified as factors involved in the regulation of HIV-1 life cycle, nonetheless, their exact role is still unclear.

The role of HN1 in the HIV-1 life cycle is not yet recognized, though its interacting proteins mainly involved in transcription and construction of microtubule structure.

Many of these proteins were identified earlier by several research groups as host factors necessary for viral replication, which supports the validity of our method, as well as their role in the life cycle of HIV-1. In our opinion, thorough examination of these proteins and their interactions with other proteins can contribute to the identification of signaling pathways involved in HIV-1 replication. These data can provide additional targets for antiviral therapies and for more successful gene therapy strategies.

6. SUMMARY

Throughout the work summarized in my dissertation, I had the possibility to plan, establish, and purify mutant (W23A, A77P, A78V, L189F, L189I, and L189P) capsid proteins of human immunodeficiency virus type 1. In this work, we have predicted the changes generated by the introduced mutations with *in silico* analysis. The results of the predictions were verified by CD-spectroscopy analysis and the proteolytic susceptibility of the proteins was determined by the analysis of the proteolytic fragments. We concluded, that mutations of the capsid cleavage sites not only affect the proteolytic susceptibility to the HIV-1 protease, but may also substantially alter the secondary structure, tryptic susceptibility and cyclophilin A binding ability of the monomeric capsid protein. Although the W23A, A77P, and L189P mutations were able to modify the cleavage of the capsid by HIV-1 protease, due to the observed structural alterations, they were not suitable for our goal. Our experiments also showed that only the L189F mutant protein remains structurally unchanged, hence, it could be the best candidate for use in experiments regarding the role of HIV-1 protease in the early phase of HIV-1 infection.

With the utilization of an HIV-1 pseudovirus and label-free proteomics we have identified 25 host proteins, which showed significant alterations in their expression level 4 and/or 12 hours after transduction. Review of the literature regarding the interaction and role of these proteins to the HIV-1 life cycle showed that with only one exception, these proteins were already verified as possible factors in the HIV-1 replication, however, in most cases, their exact role is still not clarified. Based on these data, we concluded that label-free proteomics is a capable method for the investigation of proteomic changes during HIV-1 replication. Moreover, we think that the thorough examination of the identified proteins presented here would give us more information regarding the host players and signaling pathways of HIV replication, eventually leading to the possibility of the development of new and more effective antiviral therapies.

ACKNOWLEDGMENTS

I would like to thank my supervisor Prof. Dr. József Tőzsér, full professor, Head of Department for his professional directions and support he had given mee during my work.

I would like to thank Prof. Dr. László Fésüs academic, full professor, former Head of Department for enabling me to carry out my studies at the Department of Biochemistry and Molecular Biology.

I would like to thank Dr. János Kádas, Dr. János Mótyán, Dr. Éva Csősz and Prof. Dr. Tibor Kurtán for their professional help, and for Dr. Krisztina Joóné Matúz, Dr. Mohamed Mahdi, Mária Golda, and Zsófia Szojka my colleagues for that, I could count on them any time.

I would like to thank Dr. George Tsaprailis for performing the label-free quantification experiments, and for Dr. Miklós Emri for performing the statistical analysis of the data of mass spectrometry analysis.

I am grateful for Szilvia Janics-Pető assistant for her countless practical help.

I would like to acknowledge the assistance of the employees of the Department of Biochemistry and Molecular Biology and all those who contributed to the preparation of my Ph.D. thesis.

I would like to thank my mother† for her continuous support without whom all these could not have been achieved.

This work was supported in part by the TÁMOP 4.2.2.A-11/1/KONV-2012-0023 and TÁMOP-4.2.2.D-15/1/KONV-2015-0016 projects implemented through the New Széchenyi Plan co-financed by the European Social Fund. This work was also sponsored in part by the Hungarian Scientific Research Fund (OTKA K-101591).

This work was supported by the Hungarian Scientific Research Fund (NKFI-6, 125238) to József Tőzsér and by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Biotechnology thematic program of the University of Debrecen.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/333/2019.PL
Subject: PhD Publikációs Lista

Candidate: Ferenc Tóth
Neptun ID: VZNIWD
Doctoral School: Doctoral School of Molecular Cellular and Immune Biology

List of publications related to the dissertation

1. Csősz, É., **Tóth, F.**, Mahdi, M., Tsaprailis, G., Emri, M., Tőzsér, J.: Analysis of networks of host proteins in the early time points following HIV transduction.
   BMC Bioinformatics. 20, 1-18, 2019.
   DOI: https://doi.org/10.1186/s12859-019-2990-3
   IF: 2.511 (2018)
2. **Tóth, F.**, Kádas, J., Mótyán, J. A., Tőzsér, J.: Effect of internal cleavage site mutations in human immunodeficiency virus type 1 capsid protein on its structure and function.
   FEBS Open Bio. 6 (8), 847-859, 2016.
   DOI: http://dx.doi.org/10.1002/2211-5463.12094
   IF: 2.143

List of other publications

3. Bozóki, B., Gazda, L., **Tóth, F.**, Miczi, M., Mótyán, J. A., Tőzsér, J.: A recombinant fusion protein-based, fluorescent protease assay for high throughput-compatible substrate screening.
   Anal. Biochem. 540-541, 52-63, 2018.
   DOI: http://dx.doi.org/10.1016/j.ab.2017.11.001
   IF: 2.507
4. **Tóth, F.**, Tőzsér, J., Hegedűs, C.: Indukálható bone morphogenetic protein 7 (BMP-7) termelő fogbél eredetű össejtvonal létrehozása és vizsgálata.
   Fogorv. Szle. 111 (2), 38-43, 2018.
5. Mahdi, M., Matúz, K., **Tóth, F.**, Tőzsér, J.: A Modular System to Evaluate the Efficacy of Protease Inhibitors against HIV-2.
   PLoS One. 9 (11), 1-13, 2014.
   DOI: http://dx.doi.org/10.1371/journal.pone.0113221
   IF: 3.234

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

6. Mótyán, J. A., **Tóth, F.**, Tőzsér, J.: Research Applications of Proteolytic Enzymes in Molecular Biology.
Biomolecules. 3 (4), 923-942, 2013.
DOI: http://dx.doi.org/10.3390/biom3040923

7. Csősz, É., Boross, P., Csutak, A., Berta, A., **Tóth, F.**, Póliska, S., Török, Z., Tőzsér, J.: Quantitative analysis of proteins in the tear fluid of patients with diabetic retinopathy.
Journal of Proteomics. 75 (7), 2196-2204, 2012.
DOI: http://dx.doi.org/10.1016/j.jprot.2012.01.019
IF: 4.088

Total IF of journals (all publications): 14,483
Total IF of journals (publications related to the dissertation): 4,654

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

01 October, 2019