Scuola Normale Superiore
Pisa

Structure Function Studies of

Rotavirus NSP5

Thesis submitted for the Degree of Doctor Philosophiae

(Perfezionamento in Genetica Molecolare e Biotecnologie)

Academic Year 2006-2007

Candidate
Bartosz Muszyński

Supervisor
Dr. Oscar R. Burrone

Mojej Córce Martynie,
Pamięci Mego Ojca,
Cierpliwości Mojej Matki…

CONTENTS

ABSTRACT

Rotaviruses, causative agents of gastroenteritis in young animals and humans, are large icosahedral viruses with a complex architecture. The double-stranded RNA (dsRNA) genome composed of 11 segments, that codes for 6 structural and 6 non-structural proteins, is enclosed within three concentric capsid layers.

NSP5, a non structural protein, is encoded by segment 11. It is produced early in infection and localizes in 'viroplasms', cytoplasmic inclusion bodies in which viral RNA replication and packaging take place. NSP5 is essential for the replicative cycle of the virus since, in its absence, viroplasms are not formed and viral RNA replication and transcription do not occur.

NSP5 is known to undergo two different types of posttranslational modifications, a cytoplasmic O-glycosylation and phosphorylation, which lead to the formation of proteins differing in electrophoretic mobility. Although the hyperphosphorylation process of NSP5 seems to be very complex, its role in the replicative cycle of rotavirus is unknown.

We demonstrated that NSP5 operates as an auto-regulator of its own phosphorylation as a consequence of two distinct activities of the protein: substrate and activator. In the first part of the thesis we have shown, that phosphorylation of Ser-67 within the SDSAS motif (amino acids 63-67) was required to trigger hyperphosphorylation by promoting the activation function. The evidence coming from *iv vitro* experiments, including kinase assay with recombinant casein kinase 1α from zebrafish, proved that this enzyme is responsible for a key phosphorylation step that initiates the whole hyperphosphorylation cascade of NSP5.

In the second part of the dissertation, using MALDI TOF/TOF spectroscopy, we added new data to the information about the posttranslational modifications of NSP5. We confirmed that the region of the protein encompassing Ser-67 is phosphorylated *in vivo*. Additionally we managed to map which parts of NSP5 sequence carries N-acetyloglucosamine and which regions bear phosphorylated serines or threonines.

There is no evidence about structure of NSP5 so far. In the last chapter we focused on investigating the structural organization of this crucial viral protein. To achieve this, in addition to the full length protein, one point mutation and two different truncation mutants were constructed, expressed, purified and refolded. The secondary structure of the different proteins was analyzed by circular dichroism spectroscopy and general information about protein conformation was provided. Our findings, together with an analysis of NSP5 sequence indicate that NSP5 can be an intrinsically unfolded/disordered protein.

LIST OF ABBREVIATIONS

3'CS	3'consensus sequence
5'CS	5'consensus sequence
aa	amino acids
Ala	alanine
Asp	aspartic acid
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
CD	circular dichroism
CID	collision-induced dissociation
CK1	casein kinase 1
CK2	casein kinase 2
C-terminal	carboxy-terminal
DLP	double layered particles
DMEM	Dulbecco's modified Eagle's medium
dsRNA	double stranded RNA
DTT	dithiothreitol
EDTA	diaminoethanetetraacetic acid
EGFP	eukaryotic green fluorescent protein
eIF4GI	eukaryotic initiation factor 4GI
ER	endoplasmic reticulum
ESI	electrospray ionization
FPLC	fast protein liquid chromatography
GST	glutathione-S-transferase
GTP	guanosine triphosphate
HIT	histidine triad
HPLC	high performance liquid chromotography
HRP	horse radish peroxidase
IEC	ion exchange chromatography
IMAC	immobilized metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
Km	Michaelis constant
MALDI	matrix-assisted laser desorption/ionization
min	minute
m.o.i.	multiplicity of infection
MS	mass spectrometry
N-terminal	amino-terminal
NSP	viral nonstructural protein
ORF	open reading frame
PABP	polyA binding protein
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffer saline
PFU	plaque forming units
p.i.	post infection
PKC	protein kinase C
RPHPLC	reverse phase HPLC
SDS	sodium dodecyl sulfate
Ser	serine
SLP	single layered particle
ssRNA	single stranded RNA
TLP	three layered particle
TOF	time-of-flight
UTR	untranslated region
VLP	virus like particle
VLS	viroplasm like structure
VP	viral structural protein
wt	wild type

1. INTRODUCTION

Until the 1970s, the etiological agents of diarrhea were not specified. Bacterial, viral, or parasitic agents could be detected in only 10% to 30% of children with diarrhea. In 1973, Bishop et al. noticed a 70 nm virus particle while using electron microscopy to detect infection in the duodenal epithelium of children hospitalized for treatment of acute diarrhea [11]. It soon became clear that the 70 nm particle, subsequently called rotavirus for its wheel-like appearance (Fig.1) (Latin, rota=wheel), was a causing agent of acute infantile gastroenteritis. Within 5 years after that discovery, rotavirus was recognized as the most common cause of diarrhea in infants and young children worldwide, responsible for approximately 600 000 infant deaths annually [12].
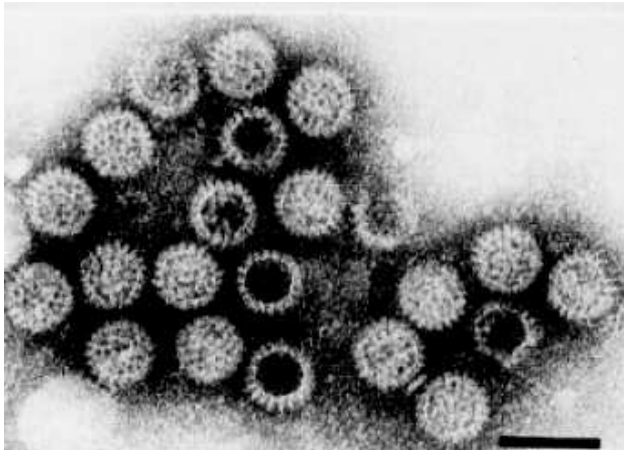
Fig.1. Rotavirus particles visualized by immune electron microscopy in stool filtrate from child with acute gastroenteritis. 70-nm particles possess distinctive double-shelled outer capsid. Bar = 100 nm [13].

1.1. Rotavirus classification

The dsRNA viruses are classified into six major families that are distinguished by they genome organization, strategies for protein coding, virion structures, host range and other differences. While the simplest members of the Totiviridae family posses single dsRNA segment, more complex viruses like members of the Reoviridae family contain 10-12 segments of dsRNA. Rotaviruses (RV) are classified as a genus within this family. Common morphological, and biochemical features of Rotaviruses are listed in table 1.

Structure

65- to 75-nm icosahedral particles
Triple-layered protein capsid
Nonenveloped (resistant to lipid solvents)
Capsid contain all enzymes for mRNA production

Genome

11 segments of dsRNA
Purified RNA segments are not infectious
Each RNA segment codes for at least one protein
RNA segments from different viruses reassort at high frequency during dual infections of cells

Replication

Cultivation facilitated by proteases
Cytoplasmic replication
Inclusion body formation
Unique morphogenesis involves transient enveloped particles
Virus released by cell lysis or by nonclassic vesicular transport in polarized epithelial cells

---

Tab.1. General characteristics of Rotaviruses [1].

Rotavirus strains are classified by three antigenic specificities- group, subgroup, and serotype/genotype. There are seven RV groups (A-G), determined by serological specificity to the structural protein of the inner capsid, VP6, predominant group antigen. The human pathogens belong to groups A, B, and C, the most important of which are those in group A.

Group B viruses have been associated with outbreaks of severe adult diarrhea in China, those in group C cause rare and sporadic outbreaks of disease in children in many countries [14] [14]. RVs in groups D-G infect only animals.

Subgroup is also determined by epitopes of VP6, and is divided into two categories, marked as I and II. Further classification of RVs involves a binary system based on two proteins, VP7 and VP4. The outermost capsid glycoprotein, VP7, determines the G (G=glycoprotein) type of a RV strain. G type refers to either VP7 serotype or genotype as determined by antibody neutralization or nucleic acid sequence, respectively. To date, all identified G serotypes and genotypes are concordant. Thus, the G type of RV is expressed as a single digit, for example, G2. There are 15 G types found in humans but most abundant are those from G1 to G4.

The RV spike protein, VP4, determines the P type (protease-sensitive) of a strain. P types also consist of serotype and genotype, though each is denoted separately in current nomenclature. P serotype is listed first followed by P genotype in brackets, e.g. P1[8]. Eleven P serotypes have been identified while twenty one P genotypes are known. Among them seven are found in humans [15].

RV field isolates are monitored worldwide and have provided useful information about RV genetic variation. Prevalent strains can differ between regions in the same country, as well as from year to year in the same region. Additionally, no correlation exists between disease severity and serotype [15]. Over the past 30 years, the majority of RV disease in North America, Europe, and Australia has been caused by P[8]G1 strains [14].

Summarizing, classification of rotaviruses is based on scheme: group, designed by roman capital letters (A-G), subgroup, represented by Roman numerals and serotype (G or P), designed by Arabic number. However emergence of an increasing number of unusual strains that cannot be classified in either subgroup or are classified to both subgroups [16] have been described. This highlights the importance of RV strain surveillance in vaccine development.

1.2. Structure of rotavirus

1.2.1. Rotavirus architecture

Tree- dimensional structures of different members of Reoviridae family, as bluetongue virus (BTV) of Orbivirus genus and several strains of Reovirus genus were obtained using electron cryomicroscopy (Cryo-EM) and computer image reconstruction techniques [17]. The overall organization is similar among these viruses.

Rotavirus is the best characterized virus of Reoviridae. The first Cryo-EM reconstruction (made up to a resolution of 40Å) of rotavirus was demonstrated by Prasad et al. in 1988. These studies were performed on the simian strain SA11. Two years later reconstruction (at higher resolution) of another rotavirus strain, rhesus RVV, was showed [4]. The analysis of these two rotavirus strains revealed very similar structural features [1].

The virion of rotavirus is, as all virions of dsRNA viruses except bacteriophages from Cystoviridae family [18], non-enveloped. It is relatively large, has a diameter of around 1000Å and a left handed T=13 icosahedral symmetry, characterized by 132 aqueous channels and 60 surface spikes [19]. Capsid is composed of three concentric protein layers. The complete rotavirus virions are called TLPs (triple-layered particles), particles that lacks the outer layer are named DLPs (double-layered particles) and, in contrast to TLPs, are non-infectious. Particles that lack two the outer shells are called SLPs (single-layered particles) or cores. The outer layer is made of two proteins- spike-protein VP4 and VP7. The intermediate one is built of VP6 trimers and surrounds the inner shell, core, composed of the structural protein VP2. The layer of VP2 protects two structural viral proteins involved in virus replication and transcription (VP1 and VP3) and eleven dsRNA segments, the rotavirus genome, that encode six structural and six nonstructural proteins (Fig. 2).

1.2.2. The outer capsid layer

The outer capsid layer plays the crucial role in first phase of virus infection as is implicated in host cell attachment, membrane penetration and cell entry [20]. In rotavirus, the outermost shell is made up of VP4 and VP7 proteins. Cryo-EM studies indicate that the 780

molecules of VP7, arranged as 260 trimers and located at the local and strict three-fold axes of a T = 13 (left-handed) icosahedral lattice, are uniformly distributed and form a smooth surface. From the surface protrude 60 spikes, bilobed at the distal end and 100- 120Å long. The spikes are composed of dimeric VP4 protein, so in the whole virion there are 120 copies of VP4 [19] [21]. Subsequent cryo-EM studies discovered that VP4 has a large globular domain that is buried inside the inner layer, making the total length of the spike about 200Å .The buried part interacts with the intermediate layer protein VP6 [22] [23].
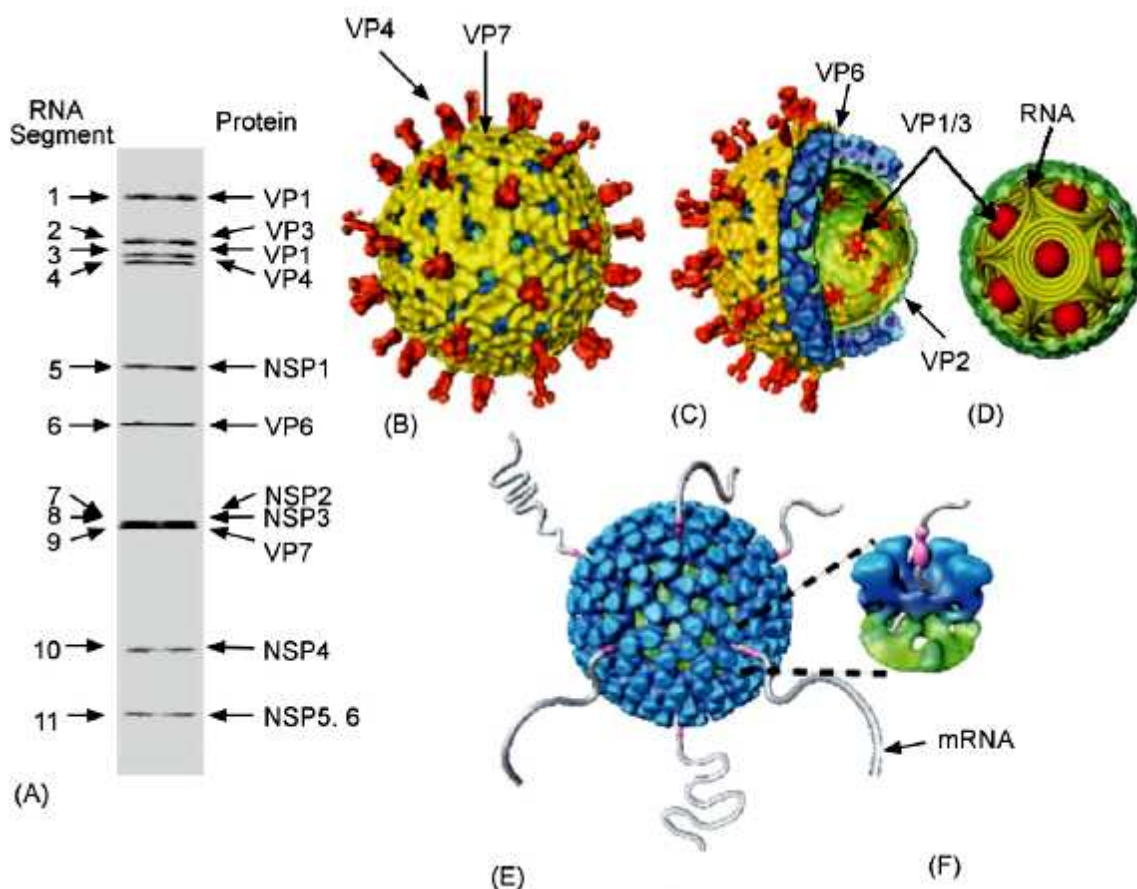
Fig 2. Architectural features of rotavirus. (A) PAGE gel showing 11 dsRNA segments comprising the rotavirus genome. The gene segments are numbered on the left and the proteins they encode are indicated on the right. (B) Cryo-EM reconstruction of the rotavirus triple-layered particle. The spike proteins VP4 is colored in orange and the outermost VP7 layer in yellow. (C) A cutaway view of the rotavirus TLP showing the inner VP6 (blue) and VP2 (green) layers and the transcriptional enzymes (shown in red) anchored to the VP2 layer at the five-fold axes. (D) Schematic depiction of genome organization in rotavirus. The genome segments are represented as inverted conical spirals surrounding the transcription enzymes (shown as red balls) inside the VP2 layer in green.(E and F) Model from Cryo-EM reconstruction of transcribing DLPs. The endogenous transcription results in the simultaneous release of the transcribed mRNA from channels located at the five-fold vertex of the icosahedral DLP [20].

The characteristic feature of rotavirus is the presence on the surface of not only VP4 spikes but also 132 aqueous channels (Fig.2F, 3 and 4) that link the outer surface with the inner core [17]. The channels, about 140 Å deep, are localized at all the five- and six-coordinated positions of the T = 13 lattice. There are three types of channels that can be distinguished according to their position and size (12 type I, 60 type II and 60 type III) [1]. Type II and III are about 55Å wide while type I have a narrower and more circular opening around 40Å in diameter and they are defined only by VP6.
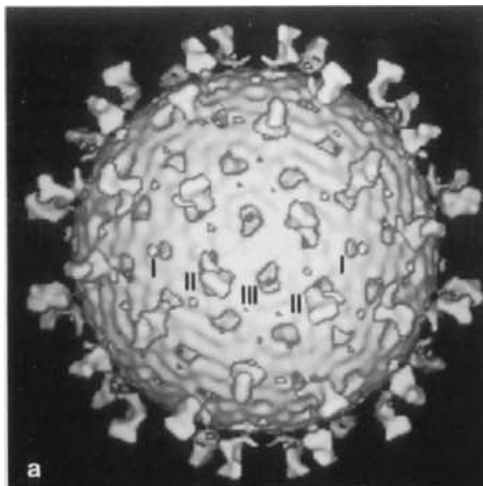
Fig. 3. Surface view of the 37Å rhesus rotavirus reconstruction viewed along two fold symmetry axes display T = 13 icosahedral lattice symmetry. The characteristic features of the outer surface are the 60 prominent spikes extending over 100 Å from the virion surface and a relatively smooth, spherical outer capsid perforated by 132 holes of three types: 12 type I holes at the icosahedral vertices; 60 type II holes at the peripentonal positions; and 60 type III holes encircling the icosahedral threefold axes of symmetry [4].
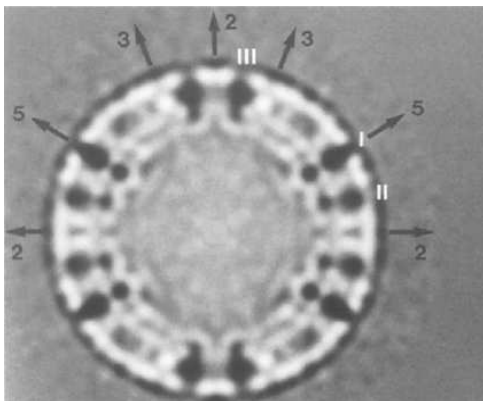
Fig. 4. Aqueous channels.
Central (equatorial) sections of the 37Å, reconstructions viewed along a twofold direction are displayed with reverse contrast (bright regions correspond to high mass density). The three types of holes identified in Fig.4 are labeled. Icosahedral symmetry axes that lie within the equatorial plane are also shown [4].

The base of the type I channels is closed in the center by the VP2 pavement [24] [25]. Type I channels are involved in importing metabolites required for RNA transcription and exporting the nascent mRNA (Fig 2E, F) [24].

1.2.3. The intermediate layer

Intermediate layer of rotavirus virion is formed by 260 trimers of structural protein VP6, the most abundant protein of the virion. The distribution of the protein mass in the second layer of the virion is not uniform and has a bristly appearance in contrast to the smooth VP7-VP4 surface. The trimers are assembled, like the outer shell, on T=13 lattice, and interact with outer shell proteins in a way that the channels in these shells are in register [19] [4]. All the surfaces of VP6 that are implicated in interactions with VP4, VP7 and VP2 contain the most conserved residues. Extensive lateral interactions between trimers and outer shell involve charged residues, whereas contacts with VP2 are mainly hydrophobic [26].

1.2.4. The inner layer, subcore and genome structure

The most inner shell of rotavirus is composed of 120 molecules of VP2 protein (60 dimers), possess T=1 symmetry and is quite smooth on the exterior [27] [1].The layer is interrupted by small pores that connect the core interior with environment. Whereas type I and type II channels are terminated at this layer, the type III channels continue beyond this shell [1]. This surface provides a structural platform for assembly of VP6 trimers, preventing aggregation of the core (VP1 and VP3 are highly hydrophobic proteins). This is the only rotavirus structural protein that has ability to self-assemble into a native-like icosahedral structure on the physiological conditions when expressed in insect cells. This suggests, that VP2 can be the scaffold for three layered particle formation [28] [29]. The two other proteins that are part of the core, VP1, the RNA-dependent RNA polymerase [30], and VP3, the guanylyl and methyl transferase (mRNA capping) [31] [32], are ten times less abundant than VP2 [33], and provide enzymatic functions required for producing the capped mRNA transcripts and genome replication. Biochemical and structural studies have shown that VP1 and VP3 are located in close proximity to each other and that form a heterodimer [34] [25]. VP2 is known to bind RNA through its N-terminal residues, what was observed in cryo-EM structure of DLS. The N-terminal region of VP2 protrudes inward at the fivefold axis to form a pentagonal shape and is also implicated in anchoring VP1/VP3 complex. This position is

consistent with releasing of nascent mRNA though the type I channel [35]. It is still not completely clear, but the N-terminus residues of one VP2 subunit appears to be occupied in transcription enzyme holding, while the N-terminal residues of the other subunit may be involved in interaction with the underlying genomic RNA [20].

The precise structural organization of rotavirus genome is not completely clear. It was however demonstrated that the genome forms concentric layers, separated by a distance of 28-30Å [36], where each segment of RNA is spooled around transcription enzymes complex, located at the icosahedral vertices. Possible model suggested by Goeyt et. al. for the bluetongue virus [37] seems to be very realistic also for the rotavirus [5] (Fig. 2D, 5).
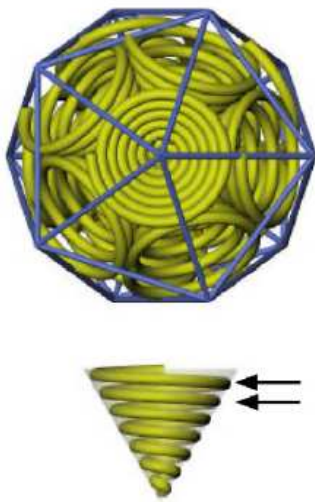
Fig. 5. Model of structural organization of rotavirus genome. Each dsRNA segment is coiled in a cone shape at the fivefold vertex [5].

This model allows 12 independent transcription complexes, each attached to an individual dsRNA segment. It is concordant with the fact that till now no dsRNA virus with more than 12 segments is known, however most of the members of this family have not 12 but 10 or 11 segments [20].

1.3. Organization of rotavirus genome segments

The total length of rotavirus genome is about 18 550 kbp. It consists 11 segment of dsRNA with a size range of 0.6 to 3.3 kbp [38]. Deproteinated rotavirus dsRNA segments are not infectious, confirming the need for presence of RNA-dependent RNA polymerase to transcribe the individual RNA segments into active mRNA. Purified dsRNA segments can be resolved by polyacrylamide gel electrophoresis and, according to the order of migration (length) of segments, are numbered from 1(the slowest) to eleven (the fastest) (Fig. 2A).

The sequence of different genome segments, also derived from distinct rotavirus strains, show common characteristics (Fig. 6.) [1]:

- sequences are A+U rich (58 to 67%)
- segments are base-paired end to end
- the positive RNA strands contain cap structures at the 5' end, while lacking 3' terminal poly (A) tails (distinctly to the most of cellular mRNAs) [39]
- each (+) RNA segment starts at 5'-prime end with guanidine, followed by a set of conserved nucleotides that are part of 5'-noncoding sequences (called untranslated 5' region, 5'UTR), then followed by usually unique open reading frame (ORF) for the protein product, that ends with the stop codon, and then finished by set of noncoding conserved 3' sequences (3'UTR) with two terminal cytosines at the very end
- all the sequenced genes possess at least one long ORF initiated by a strong initiation codon based on Kozak's rules, although some of the genes contain additional in-phase ORF (segments 7, 9 and 10) or out of phase alternative ORF (segment 11)
- all genes are monocistronic, except gene 11, that codes for two different proteins.

The roles of the UTRs are not completely elucidated but they are probably targets for rotavirus RNA-binding proteins that participate in RNA synthesis, regulation of gene expression and packaging. In addition to primary sequences, secondary structures present in the UTRs may serve as component to the recognition signals for the RNA-binding proteins [40] [41].
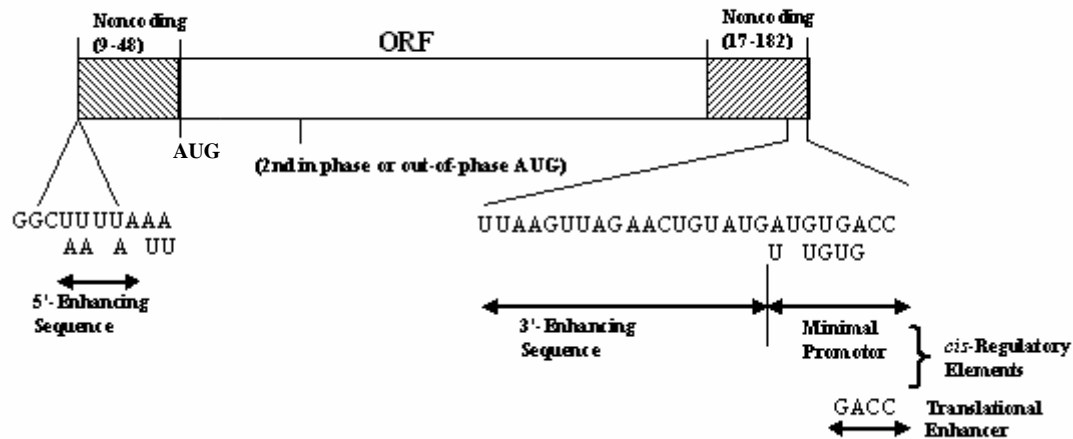
Fig. 6. Major features of rotavirus gene structure. Schematic shows the overall structure of rotavirus genes from the published sequences of genes 1 to 11. All eleven rotavirus genes lack a polyadenylation signal, they are A+U rich, and they contain conserved consensus sequences (UTRs) at their 5' and 3' ends [1].

As all the 11 mRNAs are replicated by the same VP1 polymerase, they must share common cis-acting signals and these signals are likely to be formed by secondary structures rather than by primary sequences. Some of the cis-acting signals for rotavirus RNA replication and translation have been identified but packaging rules still have to be clarified. Each mRNA has to contain a unique signal because 11 mRNAs must be distinguished from one another. Possible explanation can depend on secondary structure of positive RNA strands because secondary panhandle-like, and especially, different in each (+)RNA, stem-loop structures, have already been predicted [42].

1.4. Structural proteins

1.4.1. VP1

The structural protein VP1 is codified by the gene 1 of rotavirus. It is a basic protein and for the bovine rotavirus (RF strain) has an apparent molecular mass of 125 kDa [43]. It is present in the cores only in a few copies and together with VP3 forms a heterodimer. VP1 has two recognized functions:

- acts as a transcriptase - using as a template minus strand of dsRNA synthesizes primary viral transcripts which are extruded into the cell's cytoplasm through the I-class channels

- functions as a replicase synthesizing minus strand on the template of (+)RNA strand.

These functions were found on the basis of several observations. It reveals non specific affinity to RNA [44] and it is the only protein of rotavirus that shows sequence-specific recognition of viral RNA (binds to 3'end of gene 8 mRNA) [45]. Moreover, it shares the four common motifs conserved among sequences of all RNA- dependent RNA polymerases of other RNA viruses [43]. VP1 binds nucleotides and was demonstrated that crosslinking of photoreactable nucleotide azido-adenosine triphosphate (azido-ATP) to VP1 inhibits transcription [30].

Additionally its temperature sensitive mutant of do not synthesize ssRNA at the non permissive temperature [46]. Reconstitution experiments with baculovirus-expressed protein have shown that VP1 requires VP2 for replicase activity and presence of VP2 stimulates VP1 replicase activity several fold. VP1 is able to bind viral mRNA in the absence of any other viral proteins but its replicase activity requires previous VP2 interaction with RNA [45] [47].

Moreover it was demonstrated several times that VP1 is able to interact with NSP2 [48, 49] and can be chemically crosslinked in living cells with NSP5 [49]. However, the experiments performed recently in our laboratory showed, that the interaction of NSP5 with VP1 was found to be stronger than the interaction of NSP5 with NSP2[50].

1.4.2. VP2

VP2 is the most abundant protein of the rotavirus core (around 90%). It is encoded by gene 2 and yields the 882aa protein with a molecular mass of 102.5 kDa. Between amino acid 536 and 686 are localized two leucine zippers that may play a role in VP2 oligomeryzation into the core where it is present in the form of 60 dimers [1] .VP2 interacts with VP6, genomic dsRNA and with VP1/VP3 replication-transcription complex. Contact with the VP6 layer is created on the base of hydrophobic interactions [26]. VP2 is the only structural protein that has the ability to self-assembly into a native like icosahedral structure, when expressed in insect cells. This observation strongly indicates that it may play as structural scaffold for the proper assembly of other viral proteins. Co-expression in baculovirus of VP2 together with VP6 leads to formation of VP2/6 double layered complex called Virus Like Particles (VLPs) [29].Same results were obtained expressing these two proteins in vaccinia system [51] [52].

It is known that VP2 interact with VP1 and VP3 by its N-terminal region. That was deduced from biochemical studies on recombinant VLPs, containing VP2 with amino-terminal deletions, co-expressed with VP6, VP1 and VP3. Lack of 25 residues on the N-terminus completely prevents incorporating of VP1/VP3 replication/transcription complex [53] [20].

Further evidence suggests that the amino terminus of VP2 plays a major role in organizing the major components of the endogenous transcription apparatus, the genomic dsRNA and the VP1-VP3 enzyme complexes, within the core of the virion. It was found that the N-terminal 132aa domain of VP2 is implicated in RNA binding. Through that region VP2 is able to bind single and double stranded RNA [54]. Subsequent experiments have shown that removal of as few as the first 25 amino acids from the amino terminus of VP2 completely abolishes RNA binding activity, suggesting that the integrity of the complete amino-terminal domain is critical for the conformation of the RNA binding site. The predicted amino acid sequence of this domain contains several motifs which could potentially form the site of interaction between VP2 and RNA [55] [56]. In support of these observations, the structure of the native double-layered particle (DLP) determined by electron cryomicroscopy shows significant interactions between the inner surface of VP2 and the genomic dsRNA near the icosahedral fivefold axes and minor interactions along the icosahedral twofold axes [25]. These interactions result in nearly 25% of the genomic dsRNA adopting a highly ordered

conformation. Clearly the VP2 capsid layer is highly instrumental in organizing the genomic dsRNA within the core.

1.4.3. VP3

VP3 is a 98kDa (835aa) basic protein coded by genome segment 3. It is the minor component of the core that together with VP1 forms a homodimer- transcription/replication complex. Encapsidation of VP1 and VP3 in the inner core depends on interaction with the N-terminal of VP2 [53]. VP3 expressed in rotavirus covalently bound GTP what suggest that VP3 alone is a guanyltransferase [32]. Subsequent studies showed that VP3 interacts with GTP [57] and forms covalent VP3-GMP adducts [58] that confirmed the previous finding. It was also noticed that VP3 has ability to bind ssRNA but does not bind dsRNA. The ssRNA-binding is sequence independent, but higher interaction affinity was found for uncapped than for capped RNA [47] [44]. The guanylyltransferase activity of VP3 is nonspecific as it can cap plus- and minus-strand viral RNAs, non viral RNAs, and RNAs initiating with G and A residues, but not dsRNA [32]. It was also demonstrated by VP3 binding to S-adenosyl-l-methionine (SAM), a substrate necessary for cap methylation of RNA, that open rotavirus cores possess active methylotransferase activity.

All the data indicate that VP3 is a multifunctional capping enzyme as it shares characteristics with capping enzymes of other members of the Reoviridae, including the VP4 protein of bluetongue virus (BTV) and the λ2 protein of reovirus [44].

1.4.4. VP4

VP4 is a structural non glycosilated protein, encoded by genome segment 4. It is known to be the main component of rotavirus cell entry apparatus therefore plays essential role in the virus life cycle. Indeed antibodies against VP4 neutralize the virus and block cell entry [59].

It is a very distinctive element of rotavirus architecture as it forms characteristic 60 spikes on the surface of the virion. Recent studies discovered that VP4 is a very flexible

protein and during virus cell entry process performs series of molecular rearrangements that change its conformation. Initially, prior to trypsin cleavage, VP4 has a flexible conformation. Trypsin digestion rigidifies VP4 molecules and prime rotavirus to infect the cell. In this state VP4 is organized in a dimer that forms a rigid spike, third VP4 molecule remain flexible (Fig. 9A). This is a crucial moment in rotavirus infection as without proteolytic activation virus can not enter the cell. During digestion, VP4 (88 kDa) is cleaved into VP8* (28 kDa, aa 1–247) and VP5* (60 kDa, aa 248–776), the cleavage products remain associated in the virion [60]. The VP8* fragment contains a globular domain, the VP8* core, which forms the 'head' of the spike. Portions of both VP8* and VP5* make up the 'body' that is linked by an asymmetric 'stalk' to a 'foot,' which is buried beneath the VP7 shell (Fig.7.) [23] [22].
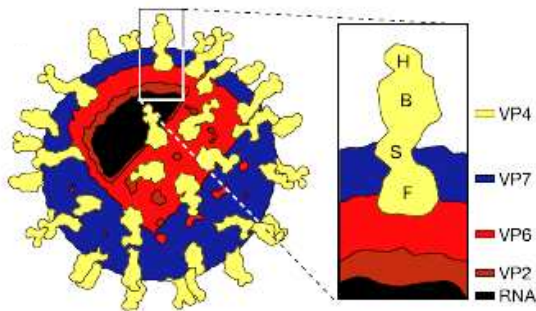
Fig. 7. VP4 domains scheme .Head (H), body (B), stalk (S) and foot (F) regions of the VP4 spikes are indicated. The drawing is based on an electron cryomicroscopy image reconstruction from [4]. Copied from [9].

VP8* masks the hydrophobic apex of VP5* on primed spikes [9]. It contains a hemagglutination motif and sialic acid (SA) binding site in the strains that use this receptor to enter the cell. These strains are able to agglutinate red blood cells by binding SA on the surface of erythrocytes [60]. Involvement of SA during rotavirus infections is not an essential step for all rotavirus strains. In many rotavirus strains, including human rotaviruses, cell entry is SA independent [61]. In these viruses, the majority of neutralizing monoclonal antibodies that recognize VP4 select mutations in VP5* [62], suggesting that cell entry is mediated mainly by the VP5*. The VP8* core is an important target of neutralizing antibodies against rotavirus and major determinant of P serotype for both SA-dependent and SA-independent strains [63]. VP8* can also play intracellular role in virus replication as it has been shown to activate cell signaling pathway upon binding the tumor necrosis factor receptor-associated factors (TRAFs) [64].

VP8* has a β-sandwich fold of the galactines- a family of sugar binging proteins. The crystal structure reveals that the core is organized in a single domain composed from two β-sheets, formed by five-stranded and six-stranded β-strands. β-strands are flanked by two α-helixes, a shorter one, laying in the inter-sheet loop and C-terminal, bridging parallel β-structure and extended β-ribbon. The central part of a structure is composed by the cleft situated between two β- sheets. The tight fold of β-sandwich with short loops between strands and dense hydrophobic cores between the major structural elements suggest a rigid structure that is unable to undergo major rearrangements during cell entry. The compactness favors protease resistance [65]. Although crystal structure of the equivalent VP8* domains from the sialic acid-independent rotavirus strains differ functionally, share the same galectin-like fold. Differences in the groove region that corresponds to the SA binding site make it unlikely that SA-independent rotavirus binds an alternative carbohydrate ligand in this location [66].

Another, not yet elucidated event triggers VP4 to change the conformation for a second time (Fig. 8 B.). Two VP5* subunits fold-back and join a third subunit to form a tightly associated trimer. VP8* domain dissociates (interactions with VP5* are due to hydrophobic interactions) and expose a hydrophobic apex of VP5* that is a potential membrane interaction region [7].
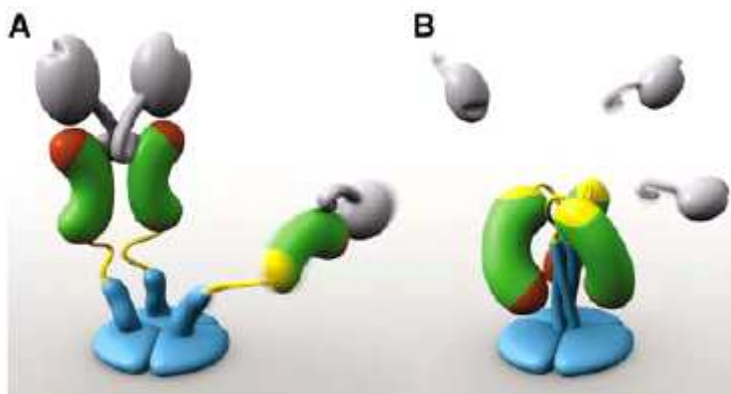
Fig. 8. Models of two VP4 conformations. (A) The trypsin primed state. Two rigid subunits form the spike visible in electron cryomicroscopy, a third subunit is flexible. VP8* is gray, with an N-terminal tether and a globular head creased by the sialoside-binding site. The VP5* antigen domain is green bean-shape, with a red membrane interaction region and a yellow GH loop. An additional b-strand C-terminal to the antigen domain is also yellow. The spike body includes the VP5* antigen domain, part of the VP8* tether, and the GH loop. The foot is blue, as is a protruding region that rearranges into the coiled-coil. (B) The putative post-membrane penetration state. VP8* has dissociated; the yellow parts of each subunit have joined in a b-annulus; the a-helical triple coiled-coil has zipped up; and the VP5* antigen domain has folded back [7].

The structure of VP5* fragment, VP5CT, was solved by crystallization. It is the strongest evidence confirming trimeric, folded-back state of VP5*. VP5CT is a protease cleaved fragment of rhesus rotavirus (RVV) that consists of residues A248 to L525 (or 528); the missing C-terminal is a part of the 'foot' that is buried in the VP7 layer and interacts with VP6. VP5CT is a well-ordered homotrimer that resembles a folded umbrella (Fig. 9A). The 'post' of the umbrella is a C-terminal, α-helical, triple coiled-coil. Each of the three panels comprising the 'shade' of the umbrella represents an N-terminal globular domain. The core of each globular domain is an eight-stranded anti-parallel β-sandwich. The flexible tip of one of the β-hairpins is exposed to the solvent and contain a sequence motif (DGE), probably implicated in rotavirus binding to α2α1 integrin (Fig.9A). Another interesting aspect of this structure is that the tips of loops projecting from the bottom edge of globular domains possess a hydrophobic region that may function in membrane penetration. The sequence of these loops shares a sequence similarity with alphavirus loops implicated in membrane disruption but the loop and domain structures are different. Thus, the sequence similarity probably reflects selection for hydrophobicity and flexibility in both loops, but not common ancestry [9].
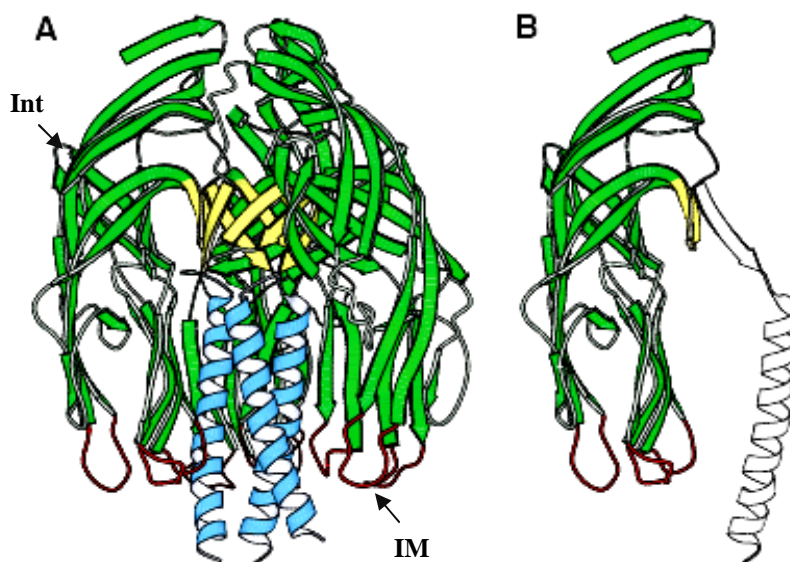
Fig. 9. VP5CT and the VP5* antigen domain. (A) Ribbon diagram of the VP5CT trimer, colored to match Figure 1. VP5CT does not include the foot region. Int- integrin binding motif; IM- potential membrane interaction loop (B) Ribbon diagram of a single VP5CT subunit. The part that forms the VP5* antigen domain is green, yellow, and red. Modified from [9] [7].

23

Apart from cell entry VP4 may be implicated in other interactions inside of infected cell. It was demonstrated that VP4 interacts with VP7 and NSP4 [67]. Newly synthesized VP4 particles are present in the space between periphery viroplasms and outside of ER [68]. It was found at the plasma membrane and it colocalizes with the cytoskeleton of infected cells. Recently it was shown that it is strongly associated with lipid raft microdomains and binds to two cellular proteins, GTPase Rab5, and to prenylated Rab acceptor (PRA1) that regulate the vesicular traffic and the motility of early endosomes along microtubules. These results suggest that Rab5 and PRA1 may be involved in the localization and trafficking of VP4 in infected cells [69]. The precise goal and mechanism of all these interactions has to be already entirely clarified.

1.4.5. VP6

VP6 is encoded by segment 6 of the genome. It is a highly immunogenic, very hydrophobic 45 kDa protein that is conserved among most of rotavirus strains. It is the most abundant structural component of the virus. It forms, organized in 260 very stable trimers, the intermediate layer of virion structure and the most outer shell in DLP in the way so that the aqueous channels in the two T = 13 layers are in register. In the overall organization of the rotavirus, VP6 appears to integrate the two principal functions of the virus, cell entry and endogenous transcription, through its interactions with the outer layer proteins VP7 and VP4, and the inner layer protein VP2. Biochemical studies indicated that none of the component of DLP alone is able to transcribe the dsRNA but VP6, despite of lack of enzymatic function, is essential for endogenous transcription of the genome [70] [71].

The structure of VP6 was revealed by crystallography [26]. Each VP6 monomer has two distinct domains, defined as B and H. Domain B is present at the base of the molecule and is composed of a bundle of eight α-helices that derive from two different segments of polypeptide chain. The first segment comes from N-terminal part of VP6 while remaining three helices are formed from C-terminus end of the sequence. In domain B the only β-sheet element consists of β-hairpin that connects two α-helices. That β-hairpin motif of contains a highly conserved sequence that protrudes into the channel I and plausibly may play a functional role in the translocation of the nascent mRNA transcripts during endogenous transcription. The electrostatic repulsion between the negatively charged channel surface and the nucleic acid is likely to facilitate the extrusion of the transcript by increasing its fluidity.

Domain H making up the top of the molecule, folds into β-sandwich with the jelly-roll topology. Hydrophobic interactions in this domain are responsible for trimerization of VP6, as mutant lacking that domain can not form the trimer while deletion mutants that lack the domain B are capable to trimerize [72]. The characteristic feature of the VP6 trimer structure is a Zn^{2+} ion that essentially contributes to the trimer stabilization.

VP6 trimers interact specifically with itself and with all three structural proteins in the virion. Interactions with VP2 are mainly hydrophobic whereas contacts between VP4 at the sides of domain H and VP7 at the top of domain B utilize charged residues. The lateral interactions between the trimers that involve charged residues are not sufficient to form the closed shell. Therefore it is likely that a proper scaffold for the assembly of VP6 into T=13 icosahedral organization is provided by VP2 layer. All the interaction surfaces relay on the most conserved amino acids of polypeptide chain. The experiments like detailed mutational analysis based on pseudo-atomic model of VP6 [71] and cryo-EM structural studies on DLP-anti (VP6) MAb complexes [34] [73] clearly indicate that not only the observed proper assembly of VP6 trimers on VP2 but also intact dynamics of interaction between these two proteins are absolutely required for correct endogenous transcription.

1.4.6. VP7

VP7 (37 kDa) is encoded by genome segment 9 in SA11 strain. It is the second, after VP6, most abundant structural protein of rotavirus. Together with VP4 it is responsible for forming the most outer layer of the virion implicated in cell attachment and entry.

VP7 is a glycoprotein that contains three N-glycosylation sites, from which only two are apparently used. Biochemical studies indicate that modifications contain N-linked adding of high mannose oligosaccharides, which are processed by trimming [74] [75] [76]. $Man_8GlcNAc_2$ and $Man_6GlcNAc_2$ oligosaccharide residues are found on intracellular VP7 while $Man_6GlcNAc_2$ is found on mature virus particles [77] [74]. VP7 is cotranslationally glycosylated as it is inserted into the membrane of the ER, and insertion is directed by a cleavable signal sequence localized at the amino-terminus of the protein [78] [79] [74]. The ORF of 326 amino acids begins with an initiation codon with a weak consensus sequence. A second, in-frame initiation codon precedes two regions of hydrophobic amino acids (H1 and H2), which can act as the signal sequence to direct VP7 to the ER, although the second is thought to be the major species used in cells. The site of cleavage of signal peptide is

glutamine 51 [80]. VP7 lacks the characteristic KDEL sequence important in conferring ER [81] retention but contains a conserved region with hydrophobic amino acids (consensus peptide LPXTG) which acts as the signal sequence in directing VP7 to ER [82].

Although these residues are critical for retention, the method by which VP7 remains in the ER is unresolved. After its insertion into membranes, VP7 is resistant to digestion with proteolytic enzymes, suggesting it is not a membrane-spanning protein [74] [76]. It remains membrane associated after high salt treatment and release of microsomal contents at alkaline pH what suggest that it is an integral membrane protein with a luminal orientation 120 [74].

VP7 in infected cells forms oligomers with VP4 and NSP4 [67].These interactions appear to be important for the assembly of VP7 into outer capsid. Mature VP7 lacks all amino acids proximal to Gln 51 but many conserved amino acids are present before the second hydrophobic domain. Proper folding of VP7, in which are involved highly conserved cysteins, requires ATP [83].It was found that this process requires also cellular factors like ER-associated chaperone calnexin, interacting also with NSP4 [84].

The precise role of VP7 during early interactions of the virus with the cell is not clear, but it has been postulated that VP7 may modulate the function of VP4 during the attachment and entry process [85]. It may interact with cell surface molecules after the interaction is initiated by VP4 [86]. It was found that VP4 contain sequences that bind integrins. It has two sequences motifs that may be important in this process, CNP that interact with integrin $\alpha v\beta 3$ [87] and GPRP that is bound by integrin $\alpha x\beta 2$ [88]. VP7 in its sequence containes pralines that binds calcium and the sensitivity of virions to low calcium concentrations is strain-dependent [89] [90]. Several studies also have suggested calcium-driven conformational changes in VP7 [91]. Studies on baculovirus-expressed recombinant VP7 have shown a requirement for calcium in the formation of VP7 trimers, which crystallize into hexagonal plates mimicking the arrangement of VP7 on the capsid [92]. Thus, while appropriate levels of calcium help maintain the structural integrity of the VP7 layer, low calcium concentrations, similar to those in the cytoplasm, trigger the disassociation of VP7 trimers leading to uncoating of the VP7 layer and releasing transcriptionally competent form of rotavirus -DLP.

1.5. Nonstructural proteins

1.5.1. NSP1

Rotavirus NSP1 (55kDa) is the product of the genome segment 5. It is known to be the least abundant protein of rotavirus. Although in infected cells it localizes throughout the cytoplasm [93] in contrast to most other rotavirus proteins that concentrate in viroplasms, it is able to interact with NSP3, NSP5 and NSP6 [94]. Moreover NSP1 was found associated with the cytoskeleton when analyzed by subcellular fractionation [93].

NSP1 is also the least conserved protein among all rotavirus nonstructural proteins. The only conserved part is the N-terminus. In its sequence there are present eight cysteins and two histidines that form the part of zinc finger domain. That domain has an affinity to 5' region of all viral mRNAs [93] [95].

NSP1 apparently is not required for rotavirus replication. Strains isolated from animals and from both immune-deficient and immune-competent children containing rearrangements in gene 5 that result in the synthesis of truncated NSP1 have showed to replicate in cell culture to titers close to those of their wild-type counterparts [96]. Also knockdown experiments using RNA interference have confirmed that the protein is not needed for virus replication [97]. Although NSP1 is not crucial for replication, production of other viral proteins and formation of viroplasms, it seems to be essential for proper rotavirus cell-to-cell spread [98] [99].

Analysis of interaction with cellular proteins showed that NSP1 interact with interferon regulatory factor 3 (IRF-3) [96] [100]. This protein during normal cell state accumulates as an inactive monomer in the cytoplasm. Events associated with virus infection like production of double-stranded RNA and expression of viral proteins trigger the innate immune response mechanisms that include the phosphorylation of IRF3 by cellular kinases. That modification initiates structural changes in IRF3 that results in its dimerization [101]. The dimer, an active form of IRF3, is translocated to the nucleus, where it interacts with specific promoter elements, stimulating expression of IFN-α and IFN-β. The secreted IFNs promote the production and activation of antiviral proteins in neighboring cells blocking the virus cell-to-cell spread [102]. It was demonstrated that NSP1 mediates IRF3 degradation through a

proteasome-dependent pathway. The role of this interaction in rotavirus infection could be diminishing the cellular interferon response [96].

The fact that NSP1 protein is much more conserved among rotaviruses infecting the same host may explain in part why rotavirus strains causing severe gastroenteritis in a homologous animal model are usually less infectious and asymptomatic in a heterologous animal model [103].

1.5.2. NSP2

NSP2, encoded by segment 8, is a highly basic protein of relative molecular mass 35kDa. In vivo studies have shown that NSP2 and NSP5 together with VP1, the RNA polymerase, are co-localized in the viroplasms of infected cells and are the main constituents of the replication intermediates [104] [105]. Indeed several facts demonstrate that NSP2 is essential for viral replication. Temperature-sensitive mutants of NSP2 fail to replicate the genome and produce mostly empty particles, which implicates NSP2 in genome replication and packaging [106] [107]. This experiment is concordant with the observation that silencing NSP2 synthesis by RNA interference causes complete blocking of viroplasms formation, production of viral proteins and rotavirus replication [97]. Additionally functional and structural similarities were described among NSP2, bluetongue protein NS2 and reovirus σNS, suggests that they are functional homologs [108].

NSP2 selfassembles into stable doughnut-shaped octamers, formed by the tail-to-tail interaction of two tetramers [10]. The overall architecture of such octamers is highly conserved even among distantly related groups of rotaviruses [109].

Many biochemical studies on recombinant NSP2 have specified its roles in genome replication and packaging. The octamers have single-stranded sequence-independent ssRNA-binding activity thanks to which are capable of destabilizing RNA–RNA duplexes by an ATP and Mg2+ independent mechanism [110] [111]. In addition, the octamers have a Mg2+-dependent nucleosidetriphosphate phosphohydrolase (NTPase) activity that cleaves the γ-β phosphoanhydride bond of any nucleoside triphosphate (NTP), yielding the products NDP and Pi. Following cleavage, the γP is transferred to NSP2, generating a short-lived phosphorylated form of the protein [110] [112]. It was demonstrated that hydrolytic activity

of NSP2 is essential for genome replication [109]. Moreover NSP2 octamers in the presence of NTPs undergo a conformational transition, shifting from a relaxed to a more condensed state as is typical of molecular motors [113].

All together, these properties have led to the suggestion that the NSP2 octamer may facilitate genome packaging and replication by relaxing secondary structures in viral template RNAs that impede polymerase function and by assisting in the translocation of viral RNAs into pre-virion cores in genome replication and packaging [114].

The understanding of these mechanisms has been much facilitated by solving the structure of NSP2 by X-ray crystallography to a resolution of 2.6A° [10]. NSP2 crystallizes as an octamer using crystallographic 4-2-2 symmetry, resulting in one monomer per asymmetric unit. The NSP2 monomer has two distinct domains (Fig. 10), an N-terminal domain from residue 1 to 141 and a C-terminal domain from 151 to 313. A striking feature of the monomer is a presence of electropositive 25 A° deep cleft that lies between the two domains.

The N-terminal domain has a novel fold. Although predominantly α -helical, has two pairs of anti-parallel β -strands towards the N terminus. In this domain can be noticed two sub-domains that are connected by a 24-residue-long predominantly basic loop (Fig. 10, arrow). There are three highly conserved prolines (53–55) at the start of this loop, which may be essential for its conformation. This loop lines the prominent grooves that run diagonally across the 2-fold axes of the doughnut-shaped octamer (Fig 11), and contributes to their basic character.

The C-terminal domain has an α /β fold. The important feature of this domain is the twisted anti-parallel β-sheet flanked by helices, formed by residues from 150 to 245, in a β3αβ2α configuration. The anti-parallel β -strands, between residues 186 and 191 (β7), and between 226 and 230 (β9), along with the loop, between residues 221 and 226, constitute the base of the cleft. The interdomain loop and the helix α5 from the N-terminal domain form a major part of one side of the cleft, whereas the other side of the cleft is made of helix α 11 and a loop between residues 245 and 260. The remainder of the C-terminal region, from residues 260 and 311, consists predominantly of α -helices (α 12– α16). That domain has a significant, limited only to the tertiary structure level, homology with the protein kinase C interacting protein (PKCI), a member of the Histidine Triad (HIT) family of nucleotide-binding proteins [115] [10].

On that base the location of the NTP-binding site was initially proposed. Now it is known that it is localized in 25 Å cleft, between the two domains of the monomer. Although lacking a precise signature of HIT motif, mutagenesis studies have indicated that conserved

basic residues in the cleft form a HIT-like motif, responsible for the binding and hydrolysis of NTPs [112]. More recently, data obtained by co-crystallization of NSP2 with nucleotide analogs has shown that His225 is the catalytic residue of the motif [116].

Extending diagonally across the NSP2 octamer surface (Fig. 11.) are four highly basic grooves, 30 Å wide and 25 Å deep, which function as ssRNA-binding sites. [10] [117]. Each groove is lined by two 24-residue electropositive loops, originating between the two subdomains of the N terminus of each monomer (Fig. 10, arrow). The location of these loops is such that they position the electropositive residues at the entrance of the clefts containing the HIT-like motif.
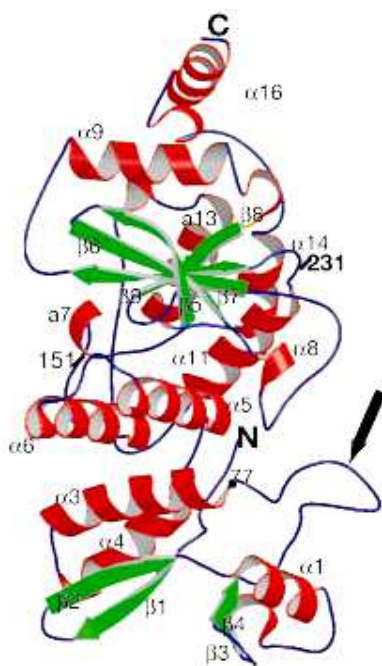
Fig. 10. The ribbon representation the monomeric subunit of NSP2. The secondary structural elements in the monomer are coloured: α-helices in red, β-strands in green and loops in blue. The basic loop in the groove between the subdomains in the N-terminal domain is shown by an arrow [10].
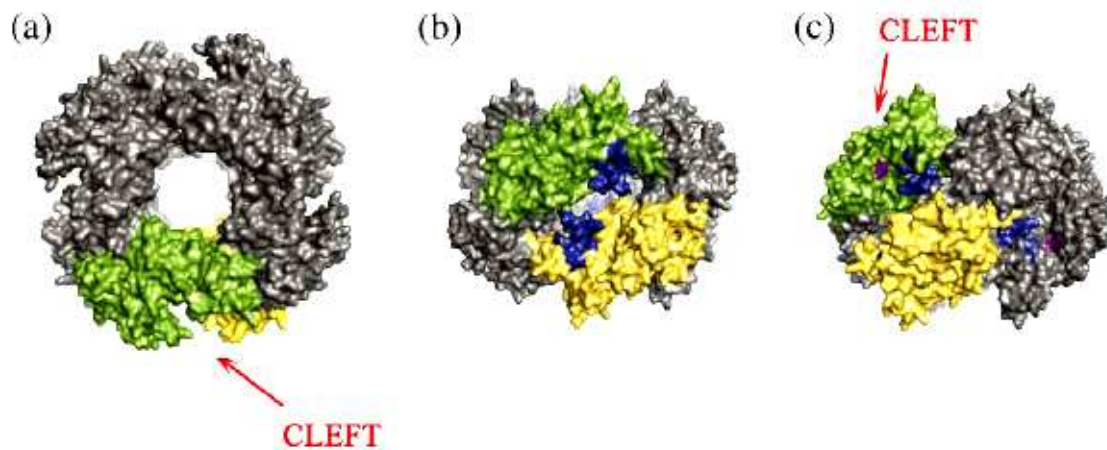
Fig.11. Structural proximity of the NTP-binding cleft and the RNA-binding grooves. Surface representation of the NSP2 octamer viewed along the 4-fold (a) and 2-fold axes ((b) and (c)). Two monomers have been identified in the octamer (green and yellow). The image in (b) represents a 90° rotation along the x-axis with respect to the image in (a). Residues 53 to 76 that form part of the electropositive loops, which line the RNA-binding grooves, are denoted in blue. (c) The image in (b) was rotated 45° clockwise along the y-axis to allow visualization into the cleft. Residues within the cleft that make up the active site for NTP hydrolysis are shown in purple [114].

Recently it was demonstrated that that the NSP2 octamer has also RTPase activity and that the RTPase and NTPase activities of the octamer utilize the same HIT-like motif and generate indistinguishable phosphorylated intermediates. It was also shown that ssRNA is preferred as a substrate over NTPs in hydrolysis reactions, likely due to the higher affinity of the octamer for ssRNA. It was discovered that NSP2 uses a HIT-like motif to direct both RTPase activity and NTPase activity [114].

The donut-shaped NSP2 octamer also displays a central hole with a diameter of 35 Å. One interesting possibility is that the central hole of the NSP2 octamer could be used as a protective environment for newly synthesized dsRNA emerging from VP1 or to serve as a passive conduit for its packaging during the assembly of VP2 capsid layer [117]. It looks as the octameric structure of NSP2 forms a tempting platform or a scaffold around which the replication complex is organized. It is possible that the hydrophobic side of the octamer, around the four-fold axis, may bind to the VP1. Furthermore the molecular partner of NSP2, NSP5, is an acidic protein and the basic grooves of the NSP2 octamer may be the sites for its binding. In fact recent observation utilizing cryo-electron microscopy has indicated that a truncated species of NSP5 binds along the same electropositive grooves involved in RNA

binding, showing that both NSP5 and RNA share the same binding site on the NSP2 octamer [117].

1.5.3. NSP3

NSP3, encoded by gene seven is a basic nonstructural 36 kDa protein that localizes in cytoplasm of infected cells [118]. The experimental evidence shows that NSP3 is responsible for shutting off cellular protein synthesis, while ensuring at the same time the translation of viral mRNAs. Rotaviruses rely on the host translation machinery to produce the proteins encoded by their genome. Rotaviral transcripts are capped at their 5' end by the action of the structural protein VP3 during transcription but their 3' ends are not poly-adenylated like most of cellular mRNAs. In the host cells only poly-adenylated and capped messages are known to be efficiently translated. This is brought about by recognition of the 5' cap by eIF4E (cap binding protein) and the poly-A tail by PABP (poly-A binding protein) which then interacts with a cellular factor eIF4GI. This factor is a multipurpose adaptor protein that is responsible for assembling the cellular translational complex composed by the proteins eIE4E, eIF4A (helicase), PABP, eIF3 as well as bound and circularized by the complex mRNA. The complex then is delivered to the ribosome where the translation begins. Rotaviruses overcome the lack of a poly-A tail by possessing on 3' ends of theirs mRNAs a consensus sequence (UGACC) which is specifically bound by NSP3 [119] [120] [121] [122]. While the N-terminal domain of NSP3 binds this consensus sequence, the C-terminal half interacts with eIF4GI with an affinity greater than cellular PABP to give translation of the rotavirus messages a selective boost following infection [120].

Resolving the structure of NSP3 by X-ray crystallography has helped to understand better mechanisms of these actions. NSP3 consists of two readily separable domains divided by a dimerization domain [94] [121]. N-terminal domain forms a heart-shaped, asymmetric homodimer with a medial basic deep cleft on the surface that creates a basic dead end tunnel that binds the 3' terminal consensus nucleotides of rotavirus mRNA. This tight interaction not only promotes translation of the rotavirus mRNA but also prevents degradation of the rotavirus message by cellular nucleases and excludes the possibility of cellular sequences recognition. Biophysical studies demonstrate a high affinity binding leading to increased thermal stability and slow dissociation kinetics, consistent with the NSP3 functions [123, 124]. On the other hand C-terminal domain is an α-helical symmetric homodimer. It

recognizes a fragment of eIF4GI and binds it with an affinity greater than PABP [125] in pockets at the dimmer interface. Site-directed mutagenesis and isothermal titration calorimetry documented that NSP3 and PABP use analogous eIF4GI recognition strategies, despite marked differences in tertiary structure[126, 127].

Recent studies [128] seem to revise mentioned model questioning how NSP3 blocks cellular protein synthesis and ensures at the same time the translation of viral mRNAs. In this work, the expression of NSP3 in infected cells was knocked down using RNA interference. Unexpectedly, under these conditions the synthesis of viral proteins was not decreased, while the cellular protein synthesis was restored. Also, the yield of viral progeny increased, which correlated with an increased synthesis of viral RNA. Silencing the expression of eIF4GI further confirmed that the interaction between eIF4GI and NSP3 is not required for viral protein synthesis. These results indicate that NSP3 is neither required for the translation of viral mRNAs nor essential for virus replication in cell culture.

These newly found discoveries can be elucidated in different ways. Even though the interactions of NSP3 with eIF4GI and the 3' end of viral mRNAs have been clearly established, there is no direct evidence that NSP3 engages simultaneously in these two interactions to promote the circularization of viral mRNAs, and although it is generally accepted, there is also no evidence that these interactions favor the translation of viral mRNAs.

Moreover interaction of NSP3 with eIF4GI is not necessary for viral translation. Furthermore, an increased level of viral RNA synthesis (both single stranded and double stranded) was detected in cells where NSP3 was silenced, suggesting that rather than promoting the translation of viral mRNAs, the interaction of NSP3 with the 3' end of viral mRNAs might prevent them from being selected for replication.

The explanation why eIFG4I silencing decrease total cellular protein production only slightly can be explained by the fact that eIF4GII can functionally complement eIF4GI [129]. Thus, the small reduction of total protein synthesis when the expression of eIF4GI was silenced could result from complementation of eIF4GII under these conditions. The fact that in standard rotavirus-infected cells (where NSP3 is expressed at normal levels) a more severe shutdown of cellular protein synthesis is observed suggests that NSP3 binds to both eIF4GI and eIF4GII. Indeed, the region of eIF4GI that interacts with NSP3 is very similar, if not identical, in eIF4GII [127]. Thus, although not formally proven, it might be expected that NSP3 could bind both factors, displacing PABP from both eIF4GI and eIF4GII, resulting in the severe shutoff of cell protein synthesis. It has not been already ruled out, but rotaviruses

might use more than one mechanism to control the translation machinery of the cell. It is possible that virus needs such a sophisticated mechanism to shut off the synthesis of a particular set of cellular proteins that could interfere with the replication cycle and/or propagation of the virus in vivo. The inhibition of protein synthesis could also be required to impair the structural integrity of the cell, facilitating cell lysis and the release of progeny viruses [128].

1.5.4. NSP4

NSP4 encoded by gene segment 10 is a multifunctional protein with roles in viral assembly and pathogenesis of infection. The 28 kDa glycoprotein has three hydrophobic domains on N-terminus and a C-terminal coiled-coil domain. NSP4 is inserted into the ER bilayer by its N-terminus. Its firs hydrophobic domain that is exposed to the ER lumen contains two N-linked high-mannose oligosaccharide residues. The transmembrane domain anchor the protein in ER bilayer (aa 22-44) [130]. The basis of ER retention of the protein is unknown, as NSP4 contains no characterized retention signals [131]. The coiled-coil domain is responsible for oligomerization of the protein that form dimers and tetramers stabilized by Ca^{2+} [132] what is interesting because of the fact that calcium mobilization by NSP4 is considered to be one of the mechanisms by which that protein fulfils its enterotoxic function. The cytoplasmic C-terminus (aa 45-175) interacts also with viral and cellular proteins. Residues 161-175 bind VP6 on the surface of DLPs [133]. During rotavirus maturation, NSP4 acts as a receptor for these subviral particles budding into the ER lumen, where the outer coat and spike protein assemble [134]. Furthermore residues 112-148 of the C terminus bind VP4 and VP7 during outer layer assembly [135]. Another studies have shown that last 54 residues of NSP4 act as a microtubule binding domain that despite of lack of homology with other microtubule associated proteins, seems to play a role in microtubules binding and preventing the traffic between ER and Golgi [136]. Recent studies that used RNA interference indicated that silencing of NSP4 in rotavirus-infected cells disrupts assembly, evidenced by a 75% reduction in progeny and the alteration in synthesis and redistribution of other rotaviral proteins. While there was a strong production decrease of VP2, VP4, VP7, NSP2 and NSP5

the synthesis level of NSP3 was twofold increased. Also the redistribution of some proteins in the cell was changed [137].

NSP4 displays enterotoxigenic activities pertinent to rotavirus pathogenesis. Purified NSP4 or even peptide 114-135 is sufficient to induce Cl⁻ secretion and diarrhea in neonatal mice by a Ca^{2+} mediated mechanism [138] [139].

NSP4 induces intracellular Ca2+ mobilization by at least two mechanisms. When added exogenously to uninfected Sf9 and HT-29 cells, as well as to isolated mouse intestinal villi and crypt epithelia, the enterotoxin induces Ca^{2+} release via a receptor-mediated phospholipase C cascade with phosphatidylinositol 4,5- bisphosphate cleavage to release inositol trisphosphate [140]. Endogenously expressed NSP4 induces a PLC-independent mechanism for mobilizing Ca^{2+} [141]. NSP4 also promotes plasma membrane Ca^{2+} permeability, resulting in influx of Ca^{2+} from extracellular sources [140]. Thus, NSP4 stimulates increased intracellular Ca^{2+} by multiple mechanisms and is postulated to do so for varying purposes. Proposed model [142] in which intracellular NSP4 induces Ca^{2+} permeability of the plasma membrane and Ca^{2+} ions mobilization early in rotavirus infection to promote an ionic environment desirable for virus maturation [143]. It is well documented that high calcium concentration is required for structural stability of the outer capsid [142] [144]. During subsequent cycles of replication, NSP4 is proposed to be secreted from infected cells to exert enterotoxic effects on neighboring uninfected cells via PLC-dependent Ca^{2+} ions mobilization with accompanying Cl⁻ secretion. Other findings suggest NSP4 forms a pore in the ER membrane to elicit Ca^{2+} release. This model arose from crystallographic analysis of the NSP4 oligomerization domain (aa 95- 137) which reveled a core metal Ca^{2+}-binding site in homotetramers [132]. Clearly, NSP4 contributes to rotavirus pathogenesis through a complex interplay with Ca2+ that may initiate secretory diarrhea.

1.5.5. NSP5

NSP5, formerly called NS26, is a nonstructural acidic protein encoded by genome segment 11. Its sequence has 196-198aa, depending on a strain, and is characterized by high serine (24%) and threonine content (4.5%) [145, 146]. NSP5 is produced early in infection and was originally characterized to have the mass of 26kDa on SDS-PAGE but further studies demonstrated that this form of the protein is a precursor for subsequent modifications.

NSP5 is subjected to two different types of posttranslational modifications, O-glycosylation and phosphorylation. Glycosylation of NSP5 is based on the addition of O-linked N-acetyloglucosamine (O-GlcNAc) monosaccharide residues to serines and threonines [146] performed by the cytoplasmic O-GlcNAc transferase [147]. This modification is present in many proteins localized to cytoplasm and nucleoplasmic compartments of the cell. Although the function of O-glicosylation in NSP5 is not known, its importance might be apparent, since many of known GlcNAc-modified proteins are also phosphoproteins [148]. In many cases, including also NSP5, the sites of phosphorylation and GlcNAcylation are localized to the same or neighbouring residues (serines or threonines).

SDS-PAGE of immunoprecipitated NSP5 derived from rotavirus infected cells revealed presence of not only two main bands with apparent molecular mass of 26 and 28 kDa but also of higher isoforms that span from 32 to 34kDa (Fig. 12). From previous studies it is known that 26 and 28kDa forms correspond to phosphorylation and glycosylation while the mobility shift to the highest bands takes place due to the phosphorylation.
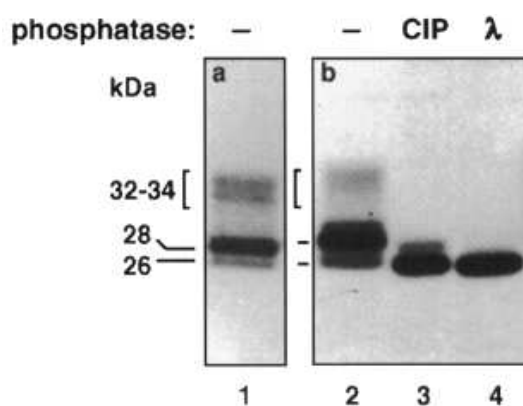
Fig. 12. SDS-PAGE analysis of NSP5. Lane 1: immunoblot analysis of extracts of MA104 cells infected with rotavirus SA11(4 h post-infection) and reacted with anti-NSP5 serum. Lanes 2-4: immunoprecipitation of NSP5 from virus infected cells labeled in vivo with [35S]methionine and treated with CIP and λ-Ppase as indicated. From [3].

Experiments with rotavirus infected cells labeled with [1,6-H^3] glucosamine revealed that the presence of glycosylation is localized mainly in 26 and 28kDa bands and suggested that hyperphosphorylated forms do not contain or contain very little of O-glycosilated residues. Treatment of hyperphosphorylated NSP5 with phosphatases caused the disappearance of higher molecular weight forms with a complete reduction of the NSP5 migration to only 26 kDa band. However the 26kDa band still contains phosphorylated sites proving that some of them are resistant to phosphatases activity.

Additionally it is known from the partial acid hydrolysis followed by the two-dimential thin-layer electrophoresis that in NSP5 phosphorylated residues are limited to serines and threonines [3].

Phosphorylation of NSP5 seems to be a complex process that gives rise to a number of isoforms. Although NSP5 immunoprecipitated from infected cells is hyperphosphorylated, western blotting analysis of the protein derived from transfected cells shows the presence of only basic 26 and 28kDa bands and only traces of the higher forms. This has suggested that NSP5 hyperphosphorylation may depend on the presence of other viral proteins. In deed, similar hyperphosphorylation effects to those obtained in virus infected cells were observed in cells cotransfected with NSP5 and NSP2. That data indicated that NSP2 alone was sufficient for up-regulating NSP5 hyperphosphorylation process, probably on the way of interaction between these two proteins (Fig.13).
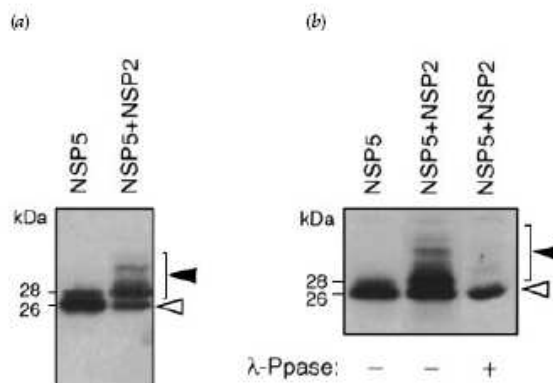
Fig. 13. Anti-NSP5 western immunoblot of cellular extracts of MA104 cells transfected with pT7v-NSP5 or co-transfected with pT7v-NSP5 and pT7v-NSP2, as indicated. λ-Ppase treatment of the extract was performed before PAGE. Open and solid arrowheads indicate the NSP5 26 kDa precursor and phosphorylated forms respectively [Afrikanova, 1998 #103].

To facilitate the studies on this phenomenon the sequence of NSP5 was arbitrarily divided into five regions (from 1 to 4 and a tail). Several experiments with NSP5 deletion mutants have been done and, interestingly, some of them have suggested that NSP5 hyperphosphorylation is the consequence of a complex autoregulatory mechanism in which

the protein is, at the same time, substrate and activator of the process that leads to its hyperphosphorylation [149].

Some deletion mutants showed to behave as good substrates, while the others played as good activators in that process. Deletion mutants lacking either region 1, region 3 or both (Δ1, Δ3 or Δ1/ Δ3) efficiently induced SDS-PAGE mobility shift in an *in vitro* translated deletion mutants such as Δ1 or Δ2. Similarly, *in vivo* coexpression experiments have shown that the same deletion mutants, Δ3 or Δ1/ Δ3, were able to cause the phosphorylation of the mutant Δ2 (fig. 14.). This observation suggested that region 1 and 3 have an inhibitory effect on the phosphorylation of NSP5. The fact that purified recombinant histidines tagged deletion mutant deriving from transfected cells did not show the phosphorylation activity of a total cellular extract containing that mutant, suggested that NSP5 does not posses the kinase activity by itself but may have a capability of promoting the activation of cellular kinase(s) [149].
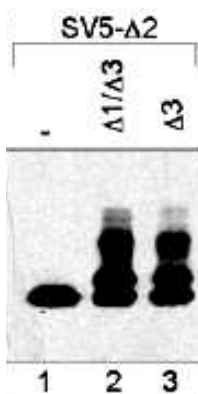
Fig. 14. Coexpression of SV5- Δ2 with or without activators Δ1/ Δ3 or Δ3 as indicated. Δ2 was indicated with anti-SV5 sera [6].

The interaction with NSP2 was described *in vitro* with purified proteins [117] and *in vivo* for both infected and cotransfected cells [49] [150] [151]. The co-expression of both proteins has a dramatic effect on their localization in the cell, leading to the formation of discrete structures that were called VLS (viroplasms-like structures). Analysis by confocal microscopy demonstrated a precise co-localization of NSP2 and NSP5 (Fig. 15) [152].
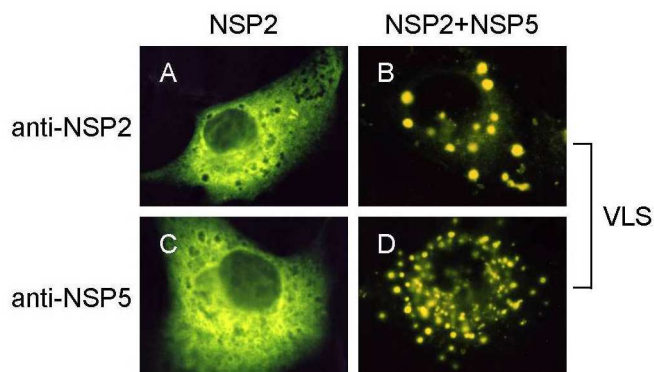
Fig. 15. Immunofluorescence of VLS: panel. Localization of rotavirus NSP2 and NSP5. The two proteins were analysed by immunofluorescence microscopy using specific antibodies for NSP2 or NSP5 as indicated, in cells transfected with either NSP2(A) or NSP5 (C) or co-transfected with NSP2 and NSP5 (B,D) [Fabbretti, 1999 #102].

Experiments focused on viroplasm like structures formation with different NSP5 deletion mutants showed that localization to the VLS requires N and C terminal parts of NSP5. The same parts of the protein were also found to be necessary for its hyperphosphorylation ability (Fig. 16) [Fabbretti, 1999 #102].
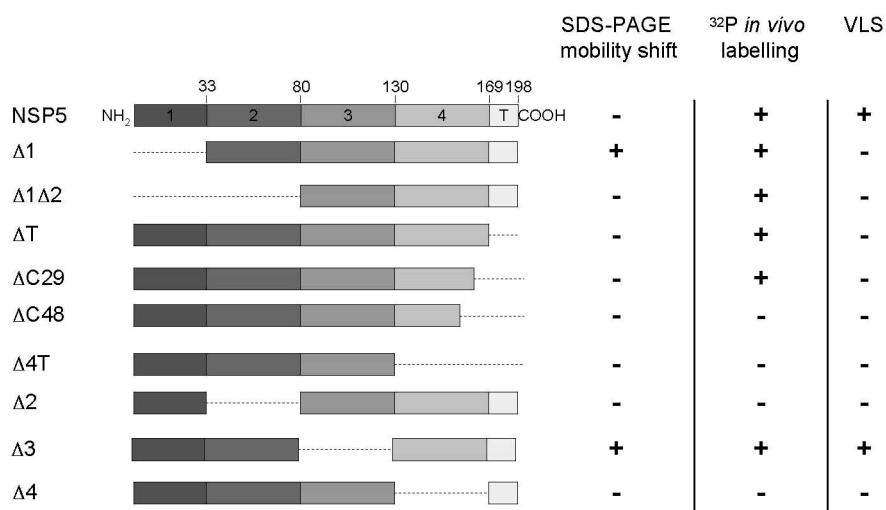
Fig. 16. Schematic representation of the NSP5 mutants constructed. The ability of mutant to produce mobility shift, to be phosphorylated *in vivo* and to form VLS is indicated [Fabbretti, 1999 #102].

The work of several research groups has demonstrated another important characteristics of these two proteins. Separately from the involvement of NSP2 in the phosphorylation of NSP5 and co-localisation in VLS it was found that NSP2 and NSP5 co-localise in infected cells in structures called viroplasms. Viroplasms, formed early in infection, are specialized, large cytoplasmic, electron-dense, nonmembrane bound structures, where the replication and packaging of the viral genome into the capsids takes place. Apart from NSP2 and NSP5 they are rich in viral RNA and contain the viral structural proteins VP1, VP2, VP3, and VP6 [153] [33] and the nonstructural protein NSP6 [154]. Several studies, including neutralization of NSP5 with anti-NSP5 specific intrabodies [155] and recently also by RNA interference [156] [97] have demonstrated that both NSP2 and NSP5 are critical for the nucleation of viroplasms and for virus replication. The analysis of viroplasm formation in infected cells indicated that NSP5 is localized in the outer part while NSP2 has more internal localisation (Fig. 17) [104].

Since NSP5 localizes to viroplasms, its capability of ss and dsRNA binding has been investigated. These studies have shown that NSP5 has unspecific affinity for both ss and dsRNA suggesting that the protein could have a role in recognizing secondary and tertiary structures of the mRNAs that have to be packaged into the forming core and together with NSP2, that has NTPase and ssRNA binding activity, unwind these structures [157]. Recent studies [117] showed that both NSP5 and RNA share the same binding site on the NSP2 octamer, and in the consequence, pointing that NSP5 may compete with ssRNA binding site on NSP2. This can suggest that one of its functions could be the regulation of NSP2-RNA interactions during genome replication.
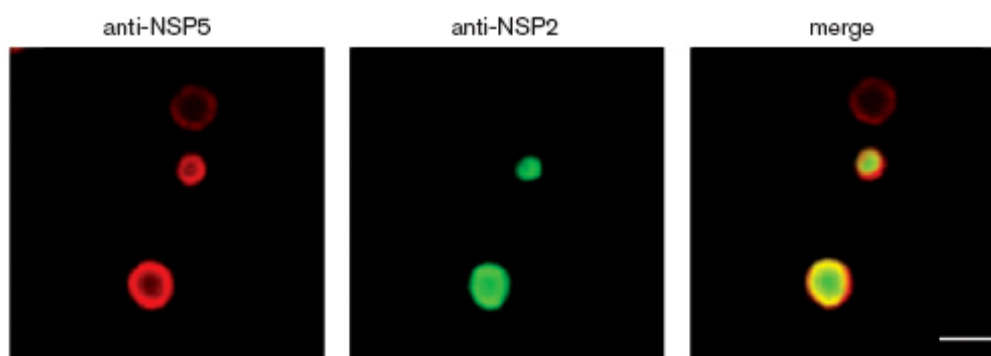
Fig. 17. Amplified images of viroplasms in rotavirus-infected cells. Double immunofluorescence in non-transfected infected cells visualized with anti-NSP5 (red) and anti-NSP2 (green) sera. Images were obtained by confocal microscopy. Bar, 2 mm. Modified from [150].

Various studies have shown that NSP5 interacts not only with NSP2 but also with the viral polymerase VP1 [49] and VP2 [158]. Recent studies using deletion mutants of NSP5 or different fragments of NSP5 fused to enhanced green fluorescent protein have proved that the 48 C-terminal amino acids are the essential for the interaction with VP1. These studies have also shown that NSP5 interaction with VP1 is stronger in the absence of other viral proteins or viral RNA, that the one with NSP2 [50], suggesting that NSP5 may act at the crossroad of important processes of virus morphogenesis, such as replication, carried out by VP1 in conjunction with the scaffold protein VP2 [47, 159, 160] and packaging, for which NSP2 is a putative molecular motor[110].

Many different information is known about NSP5, encompassing recent data about its ability (together with NSP2) to interact with β-tubulin in microtubules (Cabral-Romero & Padilla-Noriega, 2006), but no clear function has been discovered so far. In addition no homology between NSP5 and any other protein was found and there is no data about its structure, making the understanding of its function in viral infection even more incomprehensible.

1.5.6. NSP6

NSP6 is the smallest rotavirus protein. It is encoded in only some strains by segment 11 that possess an alternative reading frame. In simian strain SA11 it encompasses 92 aa that yield 11kDa protein. Until now very few reports have found the role of NSP6, probably also due to its low level of expression. It was shown that NSP6 can interact with NSP5 and seems that this interaction is due to the 35 C-terminal amino acids of NSP5, overlapping the multimerization domain of the protein, suggesting that NSP6 might have a regulatory role in the self-association of NSP5 [161] [94]. The absence of NSP6 protein in several strains (group A and C strains) [161] suggests that this protein is not essential for virus life cycle.

1.6. The virus replication cycle

1.6.1. Cell binding and entry

A crucial step in the productive infection of a virus is its successful entry into a host cell. It involves a specific attachment of the viral particle to a receptor(s) present on the cell surface. In case of many viruses, including rotaviruses, a cell entry is a complex process that requires multiple steps and multiple receptors interaction.

Rotavirus strains can be divided, if it goes to their cell attachment capability, to neuroaminidase-resistant or -sensitive strains. Resistant strains do not bind sialic acid residues on the cellular membrane and their interaction with the target cell consists contact with the other membrane component, whereas in case of the first type of strains interaction of VP4 with sialic acid takes place and treatment of cells with neuraminidase (NA) greatly diminishes virus attachment. Many animal rotavirus strains and, interestingly, most if not all human strains are NA-resistant [162].

Although earlier studies implicated VP7 in the rotavirus cell entry [163], subsequent studies have increasingly indicated that VP4 is a major player in this process. VP4, the characteristic 'spike' protein, is involved not only in a cell attachment and penetration but also in hamagglutination, neutralization, virulence and protease-enhanced infectivity of the virus [164] [165]. The last phenomenon is particularly relevant, considering that rotavirus replicates in the mature enterocytes of the small intestine, an environment rich in proteases. As previously mentioned, proteolytic cleavage of VP4 causes forming of VP5* and VP8* (that remain associated with the virion) [60] that results in transition from disordered to ordered state of the 'spike'. It was demonstrated that VP5* is able to interact with the integrin α2β1 (through the sequence DGE) that is considered to be the first receptor for NA-resistant rotaviruses and the second in case of NA-sensitive strains, suggesting that α2β1 functions as binder or as a second interactor [166, 167]. The subsequent events are less clear but it was shown that the interaction of VP7 protein with the integrin αxβ2 (by GRP motif) and αVβ3 facilitates the virus cell entry and infection [88].

The internalization mechanism of rotavirus is not fully elucidated but it can be mediated by integrins [167] as well as by the cognate protein 70 (Hsc70) that interacts with VP5*[168] [169].

The cell entry mechanism of nonenveloped viruses possesses the complication that large, hydrophilic particles must traverse a lipid membrane without having the resource of fusing two lipid membranes. Up to now there is no clear evidence about the rotavirus internalization pathway and it is possible that more than one is utilized. The ultrastructural studies have suggested that the virus can penetrate the cell membrane both by endocytosis and direct cell membrane penetration [1, 153]. However more recent studies showed that rotavirus entry can be specified as caveola- and raft-dependent endocytosis, defined by its clathrin independence, its dependence on dynamin, and its sensitivity to cholesterol depletion [170].

1.6.2. Virus uncoating, transcription and translation of proteins

One of the first events after rotavirus cell entry is the loss by infective TLP of two proteins of the outer layer, VP4 and VP7, that results in exposition of the transcriptionally active DLP to the cytoplasm.

Biochemical studies indicated that the *in vivo* decapsidation can be mimicked by treating TLPs with calcium chelators like EDTA [171]. It was proved that VP7 binds calcium and the sensitivity of virions to low calcium concentrations is strain-dependent [171] [89] [90]. Several studies have also suggested calcium-driven conformational changes in VP7 [91]. Another studies on baculovirus-expressed recombinant VP7 have shown a requirement for calcium in the formation of VP7 trimers, which crystallize into hexagonal plates, mimicking the arrangement of VP7 on the capsid [92]. Thus, while appropriate levels of calcium help maintain the structural integrity of the VP7 layer, low calcium concentrations, similar to those in the cytoplasm, trigger the disassociation of VP7 trimers, leading to uncoating of the VP7 layer [20]. This is an important event in virus replication cycle because it induces the transcriptional activity of the virus. Shortly after the virus uncoating the production of RNA transcripts starts and last until 9 to12 hours from the virus infection, after when the synthesis begin to decrease [172].

The transcription is conservative and leads to the formation of mRNAs that are identical to the plus strand of the viral dsRNAs. All the transcripts are synthetized *de novo* [39]. The synthesis is mediated by the action of endogenous viral polymerase VP1, that

behaves as a transcriptase, and by VP3. The transcription machinery to function properly requires also participation of VP2 and VP6 proteins what was described earlier. Since the RNA template is double stranded, the necessity of a helicase activity to unwind dsRNA has been proposed, but up to now it is not clear which protein can play this role and VP2 or VP3 involvement is suggested [173] [64].

The newly produced mRNAs are capped by the addition of ^{m7}GpppG^m at the 5' end, but do not contain a poli-A tail at the 3' end. Although this process is mostly performed by VP3 that was characterized as a viral guanylyl and methyl transferase, it is still not fully elucidated, because phosphohydrolase activity needed for complete reaction has not been attributed to any viral protein so far [32] [31] [44].

The primary viral transcripts extrude into the cell's cytoplasm through the 12 I-class channels located at the icosahedral five-fold vertices of the DLP particle (Fig. 2EF, 3, 4) [174].

The RNA transcripts direct the synthesis of six structural and six non-structural viral proteins (i.e., function as mRNAs) and also serve as RNA templates (RNA(+)) for the synthesis of the RNA negative strands (RNA(−)), to form the dsRNA genome segments. The production of viral proteins starts 2-3 hours after infection with a concomitant inhibition of cellular polypeptides synthesis [76].

Translation is performed by the host translation machinery and is facilitated by the action of NSP3 protein that recognizes the four-nucleotides consensus sequence on the 3' on the viral mRNAs.

1.6.3. Replication and RNA packaging

The least understood stages in viral morphogenesis are genome replication and packaging mechanism. The replication process starts around three hours after viral infection and increases in time, reaching a maximal level around 9-12 hours [172]. Once a critical mass of viral proteins is accumulated, specialized, cytoplasmic compartments (inclusions bodies) known as viroplasms are formed. It was suggested that these structures are the site where viral replication, packaging and assembly of the virus take place [105, 175]. This is in agreement with the fact that free dsRNA was never found in infected cells. The analysis of the proteins found in viroplasms revealed that they contain the proteins found in cores and DLPs (i.e. VP1,

VP2, VP3 and VP6) but additionally also two non structural proteins, NSP2 and NSP5. Interestingly, *in vitro* studies have shown that NSP2 and NSP5, although found in viroplasms, do not seem to be necessary for replication, whereas VP1 and VP2 do [176] [160]. However additional experiments, including studies on NSP2 temperature sensitive mutant, have proved that NSP2 (alone or in association with NSP5) is essential for proper packaging process [107] [20] [106] [154].

A significant uncertainness in understanding of the rotavirus biology was the origin of the plus-strand RNAs that serve as templates for dsRNA synthesis. An interesting and plausible model for events related to plus-strand RNA synthesis was suggested by Silvestri et all. Up to this model, after virion entry and loss of the outer capsid shell, DLPs initiate transcription within the cytoplasm, leading to the synthesis of plus-strand RNAs that are incorporated into polysomes and translated to produce viral proteins. At least some of the DLPs then serve as focal points for the accumulation of newly made proteins into viroplasms and are thereby incorporated into the inclusions. Transcripts produced by DLPs in the viroplasms are captured by one or more of viral RNA-binding proteins that accrue in these inclusions (e.g., NSP2, NSP5, VP1, and VP2). The captured plus-strand RNAs serve as sources of templates necessary for genome replication and assembly of progeny cores and DLPs. When the level of transcripts in the viroplasm exceeds the binding capacity of the RNA-binding proteins, the transcripts escape and can become incorporated into polysomes. Thus, viral gene expression is enhanced, and the additional proteins necessary for increasing the sizes and numbers of viroplasms are produced. In this scenario, viroplasm development becomes an autoregulatory process controlled by the capacity of the RNA-binding proteins to capture transcripts made by the DLPs within the inclusion [97].

Moreover, studies based on RNA interference have suggested that the viral mRNAs involved in protein production were not the same ones as these used for viral replication and that the second ones were protected from the action of siRNA, probably because of their localization into viroplasms [97].

One of the most intriguing aspects of rotaviruses and all segmented dsRNA viruses relates to how these viruses co-ordinately replicate and package the 11 viral mRNAs. Although the packaging signals in viral mRNAs have not been described, the hypothesis that mRNA packaging precedes dsRNA synthesis implies that it is the mRNAs which contain cis-acting packaging signals as opposed to the double-stranded products of replication. The 11 mRNAs must share common cis-acting signals because they are replicated by the same polymerase, and these signals are likely to be formed by secondary structures rather than the

primary sequences. In addition, each mRNA must also contain a signal that is unique to it alone because the 11 mRNAs must be distinguished from one another during packaging. [42].

The identification of the RNA sequences necessary for replication was possible thanks to experiments with open cores [176]. Three regions have been found to play an important role. The first is represented by 3' consensus sequence (3'CS). This is the most important cis-acting signal, that consists of very conserved sequence. Although the first experiments with open cores had shown that the last 26 nucleotides of viral positive RNA strand were important for replication [176], later studies have demonstrated that only last seven nucleotides are crucial [177], and last two nucleotides are essential for initiation of replication [178]. Additionally, it is necessary that 3'CS is single stranded to function efficiently as replication signal [179]. A second important region is located upstream from the 3'CS at the 3' end of viral mRNAs. It was shown for the genome segment 8 that several important sequences were recognized in this region by the viral polymerase [180]. A third region, situated on the 5' terminus, has been shown to enhance replication. This region lacks conserved sequence and its role in replication is not elucidated. It was proposed that secondary structure of this region may be responsible for the formation of a panhandle structure together with the single strand of 3'CS [181] [45] [177]. Moreover, the fact that predicted secondary structures of viral mRNAs contain more stem-loop structures localized on 3' and 5' ends of mRNAs and that these structures differ for the various mRNAs, can be correlated with packaging and assortment phenomenon [42].

All these processes lead to the production of transcriptionally active, dsRNA-containing double-layered replication intermediate (RI) particles. These particles are responsible for an enhanced second round of transcription, which results in a second wave of assembly of double-layered RI particles, which then bud through the membrane of the ER (Fig.18).
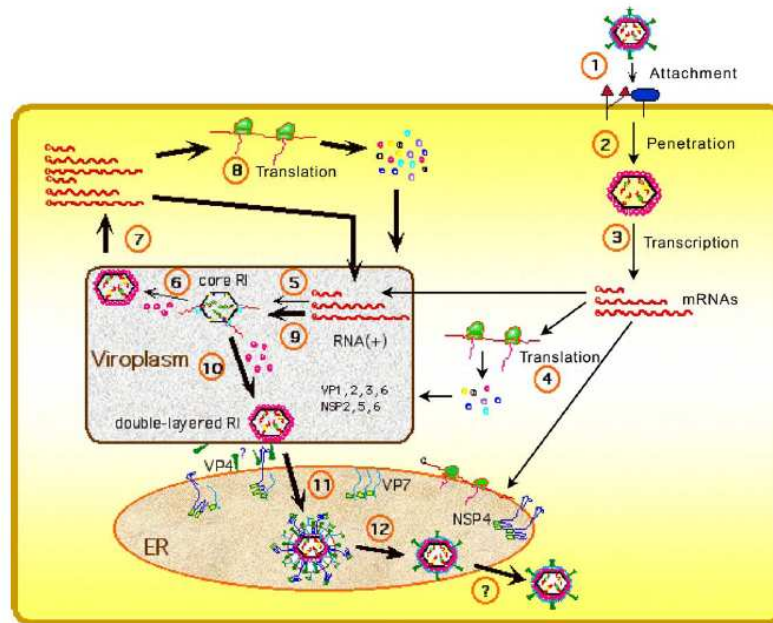
Fig. 18. Replication cycle of rotaviruses. The different steps in the replication cycle of the virus are indicated by numbers—1: attachment of the virion to the cell surface; 2: penetration and uncoating of the virus particle to yield DLPs; 3: primary transcription of the genomic dsRNA; 4: synthesis of viral proteins; 5: assembly of core RIs and negative strand RNA synthesis; 6: assembly of double-layered RIs; 7: secondary transcription from double-layered RIs; 8: secondary, enhanced synthesis of viral proteins; 9: secondary, increased assembly of core RIs and negative strand RNA synthesis; 10: secondary, increased assembly of double-layered RIs; 11: budding of double-layered RIs through the membrane of the endoplasmic reticulum (ER), and acquisition of a membrane envelope; 12: loss of the membrane envelope and generation of mature triple-layered virions [182].

1.6.4. Virus release

Rotavirus undergoes a rather unique morphogenesis that still requires further studies to become fully understood. The first step of maturation occurs inside viroplasms, where immature subviral particles assemble. The cores contain structural proteins VP1, VP3 and 11 genomic segments surrounded by the inner layer protein VP2. After the core particle is assembled it acquires the intermediate layer made of VP6 protein. VP6 localizes in the cytoplasm between the viroplasms and reticulum [68] and the assembly of intermediate layer is believed to occur concomitantly with the exit of particles from viroplasms into ER [1].

During this budding, which is mediated by the interaction of the double-layered RIs with the ER membrane-associated rotavirus protein NSP4, the particles acquire a transient

membrane envelope [183] [68] [1]. The studies with siRNA against NSP4 have shown that this protein is necessary for the formation of the intermediate layer, because in its absence the formation of DLPs and TLPs was affected, suggesting that VP6 interaction with last 20aa of a cytoplasmic tail of NSP4 may be crucial for this process. On the contrary, interaction of NSP4 with VP7 and VP4 do not have a role in the viral budding in ER [137]. The last step of rotaviral morphogenesis occur in the ER. A transient envelope and the outer layer of the virus is formed. The transiently enveloped particles where the VP7 localizes in the interior possibly relocating across the membrane, contains in addition to NSP4, the virus surface proteins VP4 and VP7, as well as minor amounts of other non-structural proteins [184]. The lipid envelope is then removed [185] by a largely unknown mechanism, related to two aspects- NSP4 glycosilation and the difference of Ca^{2+} concentration between cytoplasm and ER [186] [142].

Many interesting information about the virus release come from the use of Caco-2 cells. The human polarized intestinal epithelial CaCo-2 cells, established from a human colon adenocarcinoma, have been shown to spontaneously display many of the morphological and biochemical properties of mature enterocytes. In these cells rotavirus was released almost exclusively at the apical pole before any cell lysis was detected. It was demonstrated that the virus is able to associate with rafts (that act as platform for assembly of a virus) through a direct interaction with VP4. Following these studies it was proved that rotavirus does not leave the cell in the way of classical exocytosis but is released from the apical surface of intestinal cells through nonconventional vesicular transport that bypasses the Golgi apparatus [187].

1.7. Epidemiology, clinical aspects and vaccines

Diarrhea remains one of the most common illnesses of children worldwide [188] [188]. Although more than twenty different microorganisms (bacteria, parasites, and viruses) cause diarrhea [12], one pathogen, rotavirus, is the most frequent causative agent for the most severe disease in children younger than 5 years worldwide. Rotavirus causes 25–55% of all hospital admissions for diarrhea and more than 0,6 million deaths every year [189] [12]. The greatest mortality is in developing countries of south Asia and sub-Saharan Africa. The seriousness of the problem is emphasized by the fact that only in India alone occur more than

100 000 deaths every year. About 1 in 200 children born in these regions will die of rotavirus [190] [12] [16] [191] (Fig. 19).
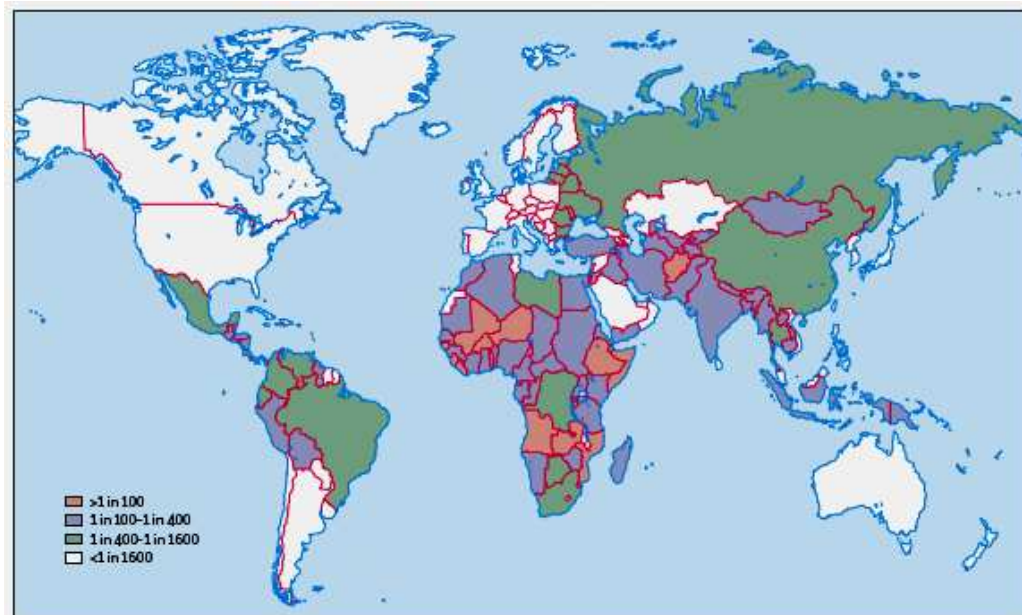
Fig. 19. Map of the world with estimated mortality from rotavirus diarrhea. Shading represents risk of death per child by 5 years of age [8]

Rotavirus is transmitted by the faecal-oral route and a small infectious dose (<100 virus particles) facilitates spread from person to person or possibly via airborne droplets [192]. Once ingested, virus not neutralized by stomach acid attaches to the proximal small intestine. During the incubation period of 18–36 h, the virus enters epithelial cells where it first elaborates a potent enterotoxin, NSP4, that can cause diarrhea, and then goes on after 18–36 h to destroy the epithelial surface leading to blunted villi, extensive damage, and shedding of massive quantities of virus ($>10^{12}$ particles per gram) in stools [1] [138]. The outcome is a profuse watery diarrhea with loss of fluid and electrolytes that can last 2–7 days and might lead to severe or fatal dehydration. Aggressive rehydration with oral or intravenous fluids can correct these imbalances and sustain a child until the diarrhea stops.

The large burden of rotavirus disease has led many international agencies, including WHO and the Global Alliance for Vaccines and Immunizations, to select rotavirus as a high priority target for accelerated development of global, efficient and safe oral vaccine.

The first efficient vaccine was a tetravalent rhesus rotavirus vaccine, RotaShield (Wyeth-Lederle, Pearl River, NY, USA), developed by Kapikian at the National Institutes of Health [193]. The variable efficacy of previous monovalent vaccines based on animal strains

was improved by development of reassortant strains that had the attenuation properties of the animal strains and individual genes encoding the outer capsid proteins of the common human strains. Rotashield was tested in a series of clinical trials in the USA, Finland, and Venezuela, and has shown to be initially safe and highly effective. In August 1998 the vaccine was recommended for routine immunization of children in the USA, and was administrated over 600 000 infants in the first 9 months of the program [194]. Unfortunately, in July 1999, intussusception, potentially fatal bowel obstruction was reported to occur in the first 2 weeks after administration of the first dose. Because of this event the vaccine was removed from the market. This decision till now remains controversial as the exact intussusception risk was judged to be about one case of in 10 000 vaccine recipients. The mechanism of this association still remains unclear, especially that intussusseption has never been associated with natural rotavirus infection. Subsequent analysis indicated that the level of risk was age-related. The vaccine was offered to children at the time of their routine immunizations at 2, 4, and 6 months of age, but catch-up immunization was provided any time up to 7 months of age. Most of these cases of intussusception happened in children who were older than 90 days old at the time of immunization [195].

The Rotashield vaccine has been acquired by a small biotech company called BioVirx, which plans to manufacture the vaccine for use in developing countries. Re-analysis of the safety profile of this vaccine suggested that the risk of intussusception could be substantially reduced if the first dose of the vaccine was administered only to infants aged younger than ninety days. Furthermore, despite the complication of intussusception, the vaccine might be more acceptable in developing countries where the risk of a child dying from rotavirus would be more than 100-fold greater than the risk of this rare complication [8].

Recently two new vaccines have been licensed in Europe and the USA and in several other countries. These vaccines are based on slightly different principles to achieve broad immunity against the diverse strains of rotavirus in circulation.

The GlaxoSmithKline vaccine, Rotarix, was developed from a human rotavirus strain (89-12) isolated from a patient in Cincinnati by Ward and Bernstein. Strain 89-12, a representative of the most common serotype in humans, (P1A[8], G1), was first attenuated by passaging 43 times and was then cloned and further passaged in *Vero* cells and renamed RIX 4414 [196]. Infants aged younger than 3 months who received the vaccine did not develop diarrhea, vomiting, or fever. To minimize the risk the possible of intussusception, GlaxoSmithKline pursued an unusual clinical development plan. First the vaccine was tested in Finland, then in middle-income countries of Latin America (i.e., Mexico, Venezuela, and

Brazil), followed by the trial in Asia (Singapore). The vaccine showed high efficacy against any rotavirus diarrhea and was safe. Rates of minor side-effects did not differ between recipients of the vaccine versus placebo. Additional results from the trial in Latin America showed significant efficacy against non-G1 serotypes, that were predominantly G9 strains. Based on these results, GlaxoSmithKline undertook a large trial to confirm that the vaccine did not cause intussusception .The trial in which 63 000 infants from 12 countries of Latin America and from Finland were enrolled showed that more intussusception events were found in recipients of the placebo [197]. The regulatory approach to license this vaccine has also been innovative. New drugs are generally licensed first in their country of origin, most often in the USA or Europe. This vaccine, however, was first licensed in Mexico and the Dominican Republic in 2004, and has been licensed in more than 35 countries and the European Union. Licensure in Brazil and the European Union which have national regulatory authorities meeting WHO standards, will facilitate licensure elsewhere in Latin America. WHO and UNICEF require that a vaccine tendered for global use be licensed in the country of origin. As a result of the European licensure, the vaccine will now be eligible for international procurement by UNICEF for developing countries or by the Pan American Health Organizations revolving fund for Latin American countries. In addition, WHO has recommended that the effectiveness of any new live oral vaccine be shown in poor populations in Africa and Asia before a global recommendation can be made. These studies are just starting and results could take 2–3 years [8].

The second vaccine, RotaTeq, was developed by Merck. It was prepared from a bovine strain of rotavirus, WC3, isolated from a calf in Pennsylvania by Clark and Offit at the University of Pennsylvania. The original monovalent bovine strain that was naturally attenuated for human beings showed variable effectiveness in several clinical trials done in the USA, China, and Africa. To improve the effectiveness of the vaccine against the diversity of common serotypes, reassortant strains were prepared. They contain ten genes from the parent bovine strain and an individual capsid gene from the most common human serotypes. Five single gene reassortants, every one containing a gene for a capsid protein for human serotypes G1, G2, G3, G4, and PIA, were combined in a pentavalent vaccine. Preliminary clinical trials showed immunogenicity and effectiveness of different formulations, and a final formulation containing 12×10^7 infectious units per dose was selected. This formulation was assessed for safety against intussusception in a field trial done in more than 70 000 infants, mainly in the USA and Finland but also in several countries of Central and South America, Europe, and Asia. Intussusception was monitored intensively for 42 days after receipt of any

dose and 11 cases of the disorder were identified, six in vaccinated patients versus five in recipients of the placebo. Within 14 days of receipt of any dose of vaccine, one case was identified in each of the vaccine and placebo groups. In a detailed follow-up study, the vaccine showed an efficacy of 86% against doctor visits, 93% against emergency department visits, and 96% against admissions for rotavirus diarrhea. It also protected against the range of rotavirus serotypes in circulation. The vaccine was licensed by the US Food and Drug Administration in February, 2006 and was recommended by the Centers for Disease Control's Advisory Committee on Immunization Practices for the routine immunization of children in the USA two weeks later.

1.8. Introduction to the spectroscopic techniques used in this work

1.8.1. Circular dichroism spectroscopy

Circular dichroism (CD) is an optical phenomenon that is based on the asymmetric nature of biological macromolecules which are able to rotate the plane of light polarization due to the presence of centers of chirality (carbon atoms). CD is a measure of the differential absorbance between left and right circularly polarized light. Circularly polarized light and linearly or plane-polarized light are readily interconvertible. A plane polarized light beams consists of right and left circularly polarized beams of equal intensity and, conversely, a circularly polarized light beam consists of two orthogonal plane-polarized beams 90° out of phase. A linearly polarized light beam becomes elliptically polarized after passing through an optically active medium (Fig.20). Since the conformation of the macromolecules affects its optical activity, these properties can be utilized to obtain information about structural organization of such molecules in solution. Theoretical calculations and several experimental measurements have demonstrated high sensitivity of CD spectra towards protein and nucleic acids structure investigation [2].
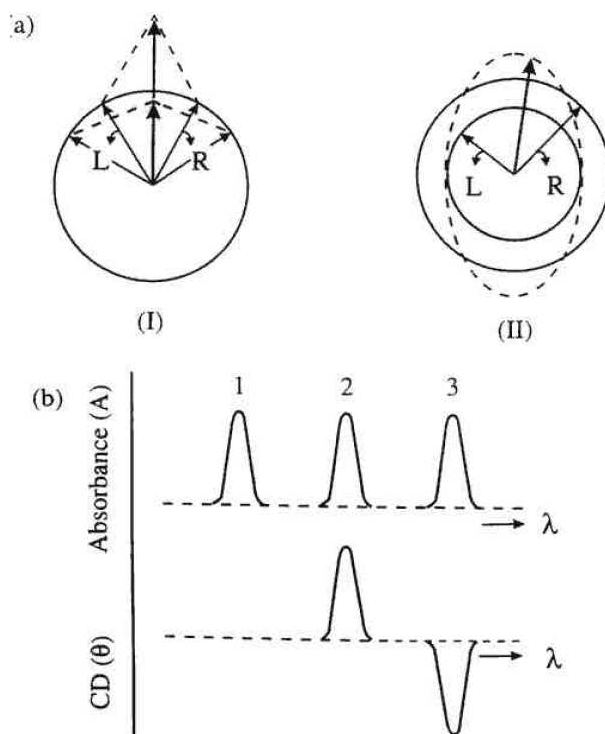
Fig. 20. Origin of the CD effect. (a) The left (L) and right (R) circularly polarised components of plane polarised radiation: (I) the two components have the same amplitude and when combined generate plane polarised radiation; (II) the components are of different magnitude and the resultant (dashed line) is elliptically polarised. (b) The relationship between absorption and CD spectra. Band 1 is not chiral; band 2 has a positive CD spectrum with L absorbed more than R; band 3 has a negative CD spectrum [2].

The CD signal of proteins can be measured in two spectral regions. The spectrum in the near UV region (320-260 nm) reflects the environments of the aromatic amino acid side chains and thus gives information about the tertiary structure of the protein. In the far UV region (240-180 nm) or amide region, which corresponds to peptide bond absorption, the CD spectrum can be used to estimate the content of regular secondary structural components such as α-helix and β-strand. All alpha-proteins show a strong double minimum at 222 and 208-210nm and a stronger maximum at 191-193nm. All beta-proteins usually have a single, negative band (between 210 and 225nm) and a single positive CD band between 190 and 200nm with intensities much lower than alpha-helical proteins [198]. For proteins, however, the major objective is to deduce average secondary structure from CD data.

Over the last three decades, various methods have been developed for the analysis of protein CD spectra, based upon the spectral characteristics of protein secondary structures [199] [200] [201] [202] [203] [204] [205] [206]. Methods use various algorithms and calculate the fraction that each component structure contributes to the net experimental

spectrum. The information contained in spectra can be treated as a sum of the characteristic individual spectra arising from each type of secondary structure present in a protein sample. Typically, empirical analysis methods utilize a reference database comprised of spectra of proteins whose crystal structures (and therefore their secondary structures) are known.

1.8.2. Mass spectrometry

Importance of mass spectrometry in biological research

The biological applications of mass spectrometry (MS) currently encompass very diverse areas of biology and medicine but is strongly linked to the term "proteomics". Proteomics could be defined as the effort to establish the identities, quantities, structures, and biochemical and cellular functions of all proteins in an organism, organ, or organelle, and how these properties vary in space, time, and physiological state (by National Academy of Science) [207]. The current importance of MS to biological research has been highlighted by the 2002 Nobel Prize in Chemistry, which was awarded to John Fenn and Koichi Tanaka for "the development of soft desorption ionization methods for mass spectrometric analysis of biological macromolecules". For a number of years, mass spectrometers were restricted to measuring of volatile compounds and small molecules. Ionization of larger molecules like proteins was not possible until the time when the fast atom bombardment ionization method was introduced [208]. The ability to ionize larger molecules was further improved with the elaboration of Electrospray Ionization (ESI) by Fenn and co-workers in 1988 [209]. The electrospray ion source was easily connected to on-line liquid chromatography (LC), which made possible the analysis of complex mixtures. Time-of-flight (TOF) mass analysis had been developed and commercialized by 1956, but relatively poor mass resolution was a problem until improvements were made in the early 1970s [210]. An ionization technique for biological molecules that could be used with TOF analysis was introduced in 1991. This ion source, Matrix-Assisted Laser Desorption/Ionization (MALDI), was the result of work in Germany by Hillenkamp, Karas, and coworkers [211] and in Japan by Tanaka and coworkers [212]. Like ESI, the MALDI ion source was capable of ionizing and vaporizing as large molecules as proteins. These methods solved the problem of generating ions from large, non-

volatile analytes, such proteins or peptides, without significant analyte fragmentation. The ability to do that conferred to these two methods the attribute of "soft" ionization methods.

Main applications of MS in biological sciences

Recent advancement increased the role of mass spectrometry-based proteomics as an indispensable tool for molecular and cellular biology and for the emerging field of systems biology. These include the study of protein–protein interactions via affinity-based isolations on a small and proteome-wide scale, the mapping of numerous organelles, and the generation of quantitative protein profiles from diverse species. The ability of mass spectrometry to identify and, increasingly, to precisely quantify thousands of proteins from complex samples can be expected to impact broadly on biology and medicine [213].

One of the most important applications of MS is an identification of proteins. There are two different approaches: first, the masses of peptides obtained from a digested protein are measured and serve as a fingerprint to identify the protein. Then these experimental masses are compared with the theoretical values calculated by applying known cleavage rules of the used protease (e.g. cleavage after K or R for trypsin) to every sequence in the database. If the experimental data matches the theoretical data of a protein sufficiently well, this protein is supposed to be present in the sample.

Another important application is quantification of proteins, since diseases often produce abnormal expression levels for certain proteins. The expression level of the corresponding mRNA molecules does not necessarily correlate with the protein level [214], especially for proteins with low numbers of copies (some 10,000 copies per cell), which can have an important regulating role in a biological pathway. Therefore, the mRNA expression levels found by hybridization experiments convey only limited information and should be complemented by the levels of the proteins they encode. Peak intensities in MS spectra cannot easily be correlated with the amount of protein present in the sample, since these intensities depend on experimental conditions and on chemical properties of the protein. However, if the intensities are normalized with an internal standard of known concentration and similar chemical properties, these normalized intensities correlate linearly with protein concentration [215].

Posttranslational modifications play an important role in biological pathways. Protein modifications are determined by examining the measured mass and fragmentation spectra via manual or computer-assisted interpretation. Phosphorylation for example is a chemical modification of certain amino acids that can switch off and on the activity of a protein. Therefore, the amount of phosphorylation for a certain protein can provide an important information about the state of a cell. Since this modification always adds the same mass (80Da) to a protein, it is easily detectable by MS. Many modifications are detectable this way and software that helps identifying modifications of peptides is available [216].

The analysis of protein complexes is another area where MS-based proteomics has a significant impact. Most proteins exert their function by way of protein-protein interactions and enzymes are often held in tightly controlled regions of the cell by such interactions. Thus, one of the first questions usually asked about a new protein is to which proteins does it bind? To study this question by MS, the protein itself is used as an affinity reagent to isolate its binding partners. Compared with two-hybrid and chip-based approaches, this strategy has the advantages that the fully processed and modified protein can serve as the bait, that the interactions take place in the native environment and cellular location, and that multi-component complexes can be isolated and analyzed in a single operation [217]. However, because many biologically relevant interactions are of low affinity, transient and generally dependent on the specific cellular environment in which they occur, MS-based methods in a straightforward affinity experiment will detect only a subset of the protein interactions that actually occur. Bioinformatics methods, correlation of MS data with those obtained by other methods, or iterative MS measurements possibly in conjunction with chemical crosslinking [218] can often help to further elucidate direct interactions and overall topology of multi-protein complexes.

Principles of the method and instrumentation

Mass spectrometric measurements are carried out in the gas phase on ionized samples. By definition, a mass spectrometer consists of three devices, an ion source, a mass analyser that measures the mass-to-charge ratio (m/z) of the ionized analytes, and a detector that registers the number of ions at each m/z value.

The most common ionization techniques used in biological applications to volatize and ionize peptides and proteins or peptides for mass spectrometric analysis are electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) [209] [211].

ESI ionizes the analytes out of a solution and is therefore readily coupled to liquid-based (for example, chromatographic and electrophoretic) separation tools. In this technique, sample solution enters through a narrow steel needle. Strong electric fields around the tip of the needle charge the surface of the emerging solvent and electrostatic repulsion of these charges disperses the solvent into a very fine spray of charged droplets, all at the same polarity. The process of droplets shrinking is followed by 'Coulumbic explosion' and is repeated until individually charged 'naked' ions of analyte are formed and released as gas-phase ions. The completely dissolved ions are formed at the atmospheric pressure and then guided into the high-vacuum region of the mass spectrometer through a capillary or by series of differentially pumped skimmers of the mass analyzer. For proteins a large number of different charge states *z* occur and the same protein is detected at various *m/z* values. Multiple charged ions are mathematically transformed into a simple mass spectrum that reveals the molecular weights of the fragments [219]. A number of variants of the electrospray technique are available; some operate without gas but with a heated capillary. Very sensitive detection can be obtained with thinner needles and nL/min flow rates [220]. Since ions are extracted continuously, ESI works well with triple quadrupole instruments or ion traps but it can also be used with most other mass analyzers [221].

While integrated liquid-chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples, MALDI-MS is normally used to analyze relatively simple peptide mixtures. MALDI sublimates and ionizes the samples via laser pulses. Peptides are suspended or dissolved in the matrix that is small, organic, UV-absorbing molecules, normally an aromatic organic weak acid (e.g. derivatives of benzoic or cinnamic acid) which strongly absorbs energy at the wavelength of the laser. N2-lasers (337 nm) and Nd-YAG (355 nm) have been successfully applied. The analyte is mixed with the matrix in a solution and the mixture is allowed to dry as a crystalline coating on a metal target plate. The matrix plays an important role in several ways. It absorbs energy and protects the analyte from excessive energy i.e. analyte decomposition and enhances the ion formation of the analyte by photoexcitation or photoionization of matrix molecules, followed by proton transfer to the analyte molecule. Moreover sample dilution into the matrix prevents association of analyte molecules. The resulting crystals are irradiated with laser pulses at the wavelength at which the matrix has high spectral absorption. This process desorbs the mixture and photoexcites the

matrix. The excited matrix then ionizes the sample thought proton transfer, causing the subsequent passage of matrix and analyte ions into the gas phase. The principal ion detected using a threshold laser intensity for MALDI is [M + H+], although signals for multiply charged ions and oligomeric forms of the analyte may be seen, especially for large proteins. At the time that the laser is pulsed and a voltage is applied to the target plate, the ionized sample is accelerated by the electrostatic field and expelled into a flight tube.

Once a sample has been ionized, it must be mass analyzed. The beam of ions is then focused and directed into a mass analyzer which separates the ions in respect to their *m/z* value. In the context of proteomics, the key parameters of a mass analyzer are sensitivity, resolution and mass accuracy. There are four basic types of mass analyser currently used in proteomics research. These are the ion trap, time-of-flight (TOF), quadrupole and Fourier transform ion cyclotron (FT-MS) analysers. They are very different in design and performance, each with its own strength and weakness. These analysers can be stand alone or, in some cases, put together in tandem to take advantage of the strengths of each [222].

MALDI is usually coupled to 'time of fligh' (TOF) analysers that measure the mass of intact peptides, whereas ESI has mostly been coupled to ion traps and triple quadrupole instruments and used to generate fragmented ion spectra (collision-induced (CID) spectra) of selected precursor ions [221].

 TOF analyzer is conceptually the simplest spectrometer. Applying a fixed voltage at the source (ca. 20kV) the ionised gas-phase sample is accelerated to a fixed kinetic energy (Ek) and then guided into a high-vacuum field-free flight tube, through which they reach the detector. Because the flight tube is a field-free region, all the Ek of the ions results from their initial acceleration, therefore a group of ions accelerated with the same constant voltage, at a fixed point and initial time, and allowed to drift in a field-free region, will traverse this region and separate in a time which depends upon their *m/z* ratios. Measuring that time-of-flight at the detector it is possible to determine the *m/z* value.

Earlier MALDI-TOF spectra were characterized by poor mass resolution due to the distribution of kinetic energy among ions that vary in initial desorption velocity [223] [224] and in energy deficits from collisions of desorbed matrix during ion acceleration. In continuous ion extraction linear mode, there was no compensation for ions with the same m/z value but different initial velocity.

To overcome this problem two improvements have been implemented in present TOF instruments configurations. First, the events of desorption/ionization and acceleration has been separated by applying the acceleration field with a slight delay, up to hundreds of

nanoseconds, relative to the laser pulse enabling the ions to be focused [225]. In the delayed extraction conditions (DE-mode), in fact, a variable voltage grid in front of the source allows MALDI-generated ion cloud to expand in a transient field free region (grid has the same voltage of source). Then a fast high-voltage pulse (grid has slightly lower voltage than source) creates a potential gradient in the ionization region that accelerates slow ions more then fast ones. This pulse voltage and time delay corrects the initial velocity differences in such a way that identical *m/z* ions arrive simultaneously at the plane of the detector, enhancing resolution of the mass spectrum.

The second improvement was the introduction of a reflectron device [226] at the axial-end of the field-free drift region. Even applying DE-mode, after ions are accelerated and velocity-focused (ions of the same mass align in time), they can defocus flying down the tube causing broadening of signal detector. The reflectron tries to correct this drifting apart and tries to focus the ions on the detector. It is an electrical mirror, with a voltage potential applied across the sides, which creates a retarding field at a voltage slightly higher than the accelerating one. The ions are sequentially slowed down through the reflector until they stop, turn around and re-accelerate back in a second drift region to a second detector. Ions species with a kinetic energy lower than the accelerating voltage will penetrate less in the reflector and will turn back sooner (as ions species with higher kinetic energy will penetrate deeper and turn back later) allowing ions of the same *m/z* value and slightly difference in flight times to be packed and focused in space and time to the detector.

To allow the fragmentation of MALDI-generated precursor ions, MALDI sources have been recently coupled to quadrupole ion-trap mass spectrometers [227]and two types of TOF instruments: the quadrupole TOF hybrid instrument (Fig.21) [228] and the MALDI-TOF-TOF (Fig.20) [229]. A collision cell is inserted between the two analyzer. Ions of a particular m/z value are selected in the quadrupole mass filter or in the first TOF, fragmented in the collision cell and the fragments ions masses are read out by a TOF analyzer (Fig.22). These instruments have very high sensitivity, resolution, mass accuracy and high speed data acquisition. During MS/MS analysis, MALDI-TOF-TOF instrument offers the possibility to finely control fragmentation conditions and simultaneously provides both low-energy and high-energy collision induced spectra. The degree of fragmentation and structural information obtained are related to adjustment in laser intensity, collision gas pressure and collision energy. A collision energy of 1-2 keV in the CID-cell results in the production of low mass internal fragments, ions from amino acid side chain fragmentations, ions specific of particular

amino acids increasing the confidence in the peptide sequence interpretation and identification of multiple proteins from complex mixtures.
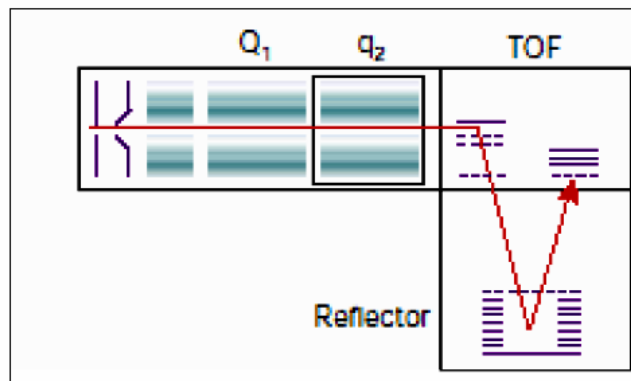
Fig. 21. The quadrupole TOF instrument combines the front part of a triple quadruple instrument with a reflector TOF section for measuring the mass of the ions [228].
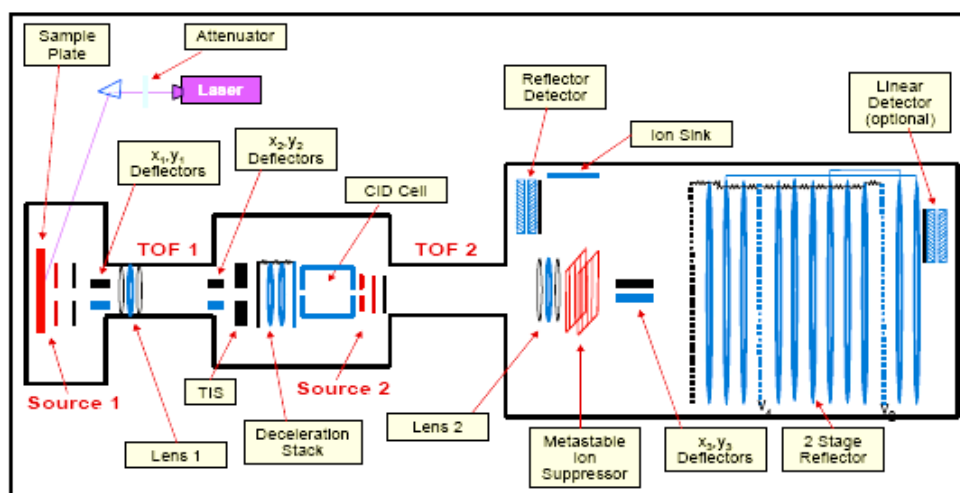
Fig. 22. MALDI TOF-TOF AB 4700 TOF/TOF™ Ion Optics [229].

2. MATERIALS AND METHODS

2.1. *Nsp5 wt* **and mutants constructs preparation and production of anti NSP5 serum**

Plasmid construction was performed using the *E. coli* DH5α strain (Life Technologies Inc.). *Nsp5 wt*, derived from porcine OSU rotavirus strain (Genbank accession number: D00474) was amplified by PCR from cDNA using specific primers on 5' and 3' ends (already characterized by [49]). The PCR fragment was cloned to pGemTEasy (Promega) vector, sequenced and then subcloned to pcDNA3-His, that was described earlier by Eichwald [149]. Final cloning to pET23d (+) (Novagen) was done using NcoI and BamHI sites. NSP5S67D was subcloned from pCDNA3 construct as mentioned in [6] into pcDNA3-His [149] and finally cloned to pET23d (+). Delta C18 mutant was amplified on the matrix of pET23d (+)NSP5 wt using subsequent primers:

forward: 5'GATTCTAGAGTCGACCATGGGTCATCACCATCAC3'

reverse: 5'GTGGGATCCTTACTTTTTCTTATATT3'

and then cloned to pET23d(+) where it was sequenced (MWG-Biotech, Ebersberg, Germany). Mutant NSP5Δ32 was prepared on the basis of already existing construct characterized by [49]. NSP5 truncated gene was cloned using NcoI/BamHI sites into pET23d (+) vector.

The following constructs were obtained as described in [6].pT7v- Δ3a, pT7v-Δ3b, pT7vΔ3/S63A, pT7v-Δ3/S65A, pT7v-Δ3/S67A, and pT7v-Δ 3(S63,65A_S67D) were obtained by double-step PCR by using internal oligonucleotides to amplified regions 1 and 2 and cloned KpnI_ClaI in pT7v-(4TclaI5'). pT7v-(4TclaI5') was obtained by PCR of the regions 4 and T, inserting the restriction sites KpnI and ClaI at the 5' and BamHI at the 3' The fragment was cloned as KpnI/BamHI in pcDNA3. pT7v-SV5 Δ 2 was obtained by inserting at the N-terminus the SV5 tag with oligonucleotides, 5'-AGCTTGTACCATGGGCAAACCAATCCCAAACCCACTGCTGGGTCTGGATGGTAC-3' and 5'-CATCCAGACCCAGCAGTGGGTTTGGGATTGGTTTGCCCATGGTACA-3', and into HindIII/KpnI. pT7v-NSP5a, pT7v-NSP5/S57A, and pT7v-NSP5(S63,65A/S67D) were obtained by PCR of the respective pT7v-Δ3 point mutation by using specific primers to

incorporate KpnI and BstBI restriction sites at the 5'and 3' ends of the 1+2 region, respectively. The fragments were cloned KpnI/ BstBI into pT7v-Δ1Δ2(KpnI/BstBI). This construct was obtained by PCR with specific primer to insert KpnI/BstBI and BamHI restriction sites at the 5'and 3'ends of the Δ1Δ 2 region, respectively.

pQENSP5(S63,65A/S67D), and pQE-NSP5a were obtained by digestion KpnI/EcoRV of above described pcDNA3 construct and cloned KpnI/SmaI in pQE-30 (Qiagen, Valencia, CA).

Anti-NSP5 sera were obtained through protein immunization using purified His-NSPwt produced in E. coli M15 [Rep4] strain as described before [6]. Guinea pigs were injected intraperitoneally with 100μg of purified protein for each boost. The boosts were repeated three times every 15 days and the animals were bleeded one week later. Antibodies were tested by Western Blot analysis on extracts of rotavirus infected and non infected cells and dilution 1:7000 was experimentally established as sufficient.

2.2. Tissue culture

MA104 cells were routinely cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal calf serum (FCS), 2mM L-glutamine (Gibco BRL) and gentamycin (100 μg/ml). Cells were propagated using trypsin (500 μg/ml), inhibited with complete medium, centrifugated at 1100 rpm for 4 min. and resuspended in complete medium.

2.3. Transient transfection of MA104 cells

To prepare cellular extracts cells were transfected essentially as described by Africanova, [49]. Briefly, $5x\ 10^5$ cells growing on a 35 mm diameter Petri dish, were infected with T7-recombinant vaccinia virus (strain vTF7.3) [230] and 1 h later transfected with 5 μl of Transfectam reagent (Promega) mixed with 2 μg of plasmid DNA and incubated for 16 h.

At 16 h post-transfection cells were washed twice with cold PBS, kept on ice and lysed in 80 μl of cold TNN lysis buffer (100 mM Tris-HCl pH 8.0, 250 mM NaCl, 0.5% NP40 and 1mM PMSF) for 10 min at 4ºC and centrifuged at 10000 x g for 5 min.

2.4. Virus infection and propagation

The porcine rotavirus strain OSU, bovine strain RF and simian SA11 were propagated and grown in MA104 cells as characterized by Estes, [231]. Briefly, the inoculum was activated in presence of 10 μg/ml trypsin until cythopatic effect was visualized. The infective medium was frozen and thawed three times, centrifuged at 500 rmp for 10 min to eliminate cellular debris and stored at – 80ºC. Virus was titrated using plaque assay as described in [232].

The recombinant vaccinia virus VT7/LacOI/NSP2 was propagated and grown in BSC-40 cells as described by Ward et al, [233].

2.5. Western blotting

Samples were loaded and ran in a SDS-PAGE according to Laemmli method [234]. After, samples were transfered to a PDVF membrane (Immobilion-P, Millipore) for 2 hours at 200mA or O/N at 50 mA. The membrane was blocked in PBS-milk 5% for 45 minutes and incubated for 1, 5 hour with the primary antibody in PBS-milk 5%. After, three washes in PBS-milk 5%, the secondary antibody conjugated to peroxidase was added and incubated for 1 hour. Finally, the membrane was washed 3 times in PBS-milk 5% for 5 minutes and at last once in PBS. The reaction was developed using the ECL kit (Pierce).

2.6. Sample preparation for MALDI TOF/TOF analysis

MA104 cells were cultivated on 15 cm diameter Petri dishes (Costar) until they reached the state of confluence. Cells then were mock-infected (control samples preparation) or infected with rotavirus (OSU, RF SA11 or strain) at a multiplicity of infection (MOI) of 5 PFU/cell. After 12 hours cells were washed 3 times with cold PBS, scraped on ice and lysed with cold TNN buffer containing 20mM NaF (phosphatases inhibitor). For every 1 ml of cell lysate 30 µl of protein A-sepharose beads (Repligene) were used. Beads were resuspended and washed 3 times with lysis buffer (TNN). Prebinding mix was prepared- 1 µl of anti-NSP5 sera (2.1.) was used for every 4 µl of protein-A beads. Mix was incubated on the rotator in the 4°C for 45 min. Afterwards beads were washed twice with cold TNN and lysates of rotavirus and mock-infected cells were added. Reactions were incubated on the rotator in 4°C for 2 h. After incubation, protein-A beads were washed 3 times and resuspended in 1 ml of cold TNN. To increase purity of immunoprecipitation a quick glycerol step gradient was performed. To the tubes containing beads resuspended in 1 ml of TNN 500 µl of TNN-glycerol 20% was gently overlayed. Samples were centrifuged 30 seconds to pellet the beads and supernatant was removed from the top to down. Beads were washed 3 times with TNN. From this point samples for MALDI TOF/TOF analysis were prepared with two different methods in purpose to increase the possibility for obtaining more reliable results. Schematic representation of this processing is illustrated in figure 23.
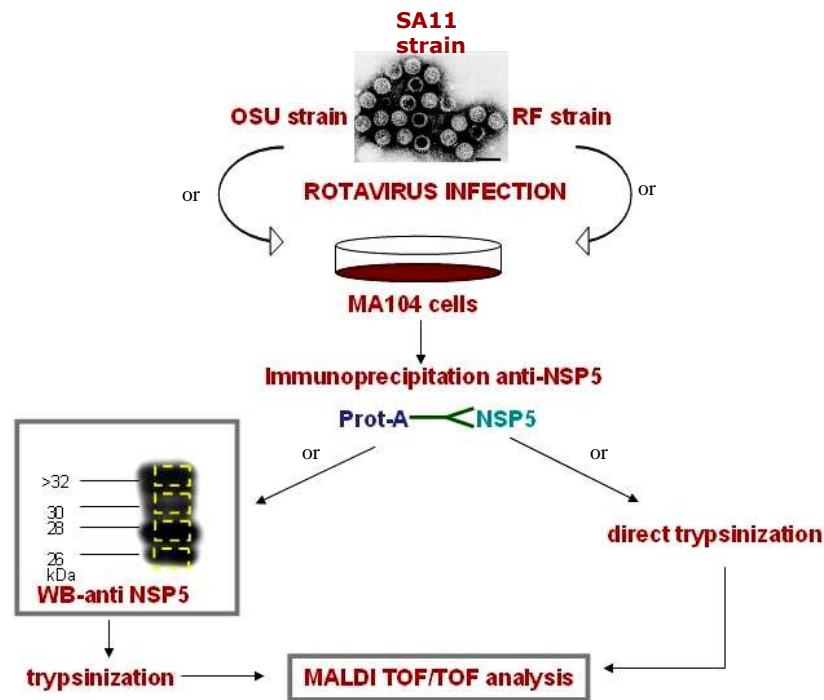
Fig. 23. Scheme of sample preparation for MALDI TOF/TOF analysis. Cells were infected with OSU, RF or SA11 rotavirus strain. NSP5 was immunoprecipitated and subsequently subjected to direct trypsinization or immunoprecipitated samples were separated by SDS-PAGE, then NSP5 was fractionated, excited from the gel and trypsin-digested.

2.6.1. Protein extraction from SDS-PAGE gel

Protein-A beads with bound NSP5 and immunoprecipitated in the same way control samples (mock-infected lysate) were placed in microcentrifuge columns. Attached proteins were eluted from protein A-beads with 0.2 M glycin HCl pH 2,8. Immediately after elution law pH was neutralized with 1M Tris-HCl pH 8,0. To decrease the volume of samples, proteins were acetone precipitated. Pellets were dried by centrifugation in a SpeedVac, resuspended in 1x SDS-PAGE loading buffer and boiled. Samples were loaded on 15% polyacrylamide gel in the way that control samples and samples containing immunoprecipitated NSP5 were separated by lines of prestained marker (Precision Plus Protein Standards, Bio-Rad). One tenth of each sample was loaded in the same order on the second half of the gel. After protein separation, the gel was cut into two halves, and on the part containing less material, Western blotting was performed. The second half was kept

during this time in a SDS-PAGE transfer buffer. Result of WB, the film, was superimposed on the first half of the gel. NSP5 bands were excised with a clean scalpel, cut in sample pieces (~ 1 mm^3) and put in a 1.5 ml tube. The lowest band represented the least modified protein, the highest contained the fraction of the protein that underwent more advanced posttranslational processing. On the basis of marker bands location control samples were precisely isolated.

Subsequently samples were prepared using a modification of the method proposed by Shevchenko et al. 300 μl of 50 mM ammonium bicarbonate (pH 8,0) was added to each tube and sample were vortex for 15 min. The supernatant was discarded and volume sufficient to immerse the gel pieces of 50 mM ammonium bicarbonate (pH 8,0)/50% acetonitrile was added. The supernatant was removed and 100 μl of acetonitrile was added to dehydrate the gel pieces for 5 min. The supernatant was removed, the gel pieces were dried in a SpeedVac for 5 min. Pieces were covered with 10 mM DTT solution and reduced for 1 hr at 56°C. Samples were cooled to room temperature and DTT solution was replaced with the same volume of 55 mM iodoacetamide solution. Tubs were incubated for 45 min at room temperature in the dark with occasional vortexing. Then washing with 300 μl of 50 mM ammonium bicarbonate, pH 8,0 for 15 min was performed by vortexing, and 100 μL of 50 mM ammonium bicarbonate/50% acetonitrile was added. After 15 min. the supernatant was discarded and 100 μl of acetonitrile was added to dehydrate the gel pieces for 5 min. Liquid phase was removed and the gel pieces were dried in a vacuum centrifuge for 5 min. The gel pieces were reswollen with a minimum volume of sequencing grade porcine trypsin (0.1 μg/μl) and after 10 min 100 μl of 50 mM ammonium bicarbonate, pH, 8,0, was added. Samples were incubated overnight at 37 °C. The tryptic peptides were extracted, once with 100 μl water, twice with extraction buffer (50 mM ammonium bicarbonate, 60% CH3CN and 1 % TFA) and at last with 100 μl of CH3CN and finally collected in 1.5 ml tube. The extracts of peptide solutions were dried in a SpeedVac evaporator.

2.6.2. Direct trypsinization

After washings with TNN, beads were washed three times with 50mM ammonium bicarbonate and then ammonium bicarbonate buffer in volume sufficient to cover the beads was added. The buffer contained 0.4μg of trypsin (1ng/μl). Control samples and those

containing NSP5 were incubated for 6 hours in 37°C. To stop the reactions acetic acid to final concentration of 0.1% was added. Peptides were eluted from the beads by centrifugation and frozen in -80°C.

2.7. MALDI TOF/TOF analyses

Analyses were performed using 4800 MALDI TOF/TOF™ tandem TOF technology analyzer from Applied Biosystems/MDS SCIEX. Peptide masses present also in control reactions were considered to be cellular contaminants and were manually eliminated from the spectra. The protein was identified with high score by Mascot tool [235] (http://www.matrixscience.com/) as the database search engine.

The measured masses were manually verified using Peptide Mass program (www.expasy.com) in case of not modified peptides or using Glycomod tool (www.expasy.org) in case of phosphorylated/glycosilated peptides. The peptides that masses difference did not exceed 0.2 Da between theoretical calculations and experimental data and that were found in more than one experiment were assigned to NSP5.

2.8. Expression of NSP5 and its mutants

E. coli BL21 [DE3] cells (Novagene) carrying the plasmid that encoded for a target gene under T7 promoter were grown in Terrific Broth medium, containing ampicillin (100ug/ml) in 37°C until the culture reached an optical density 0.7-0.8 (absorbance measured at 600 nm, GeneQuant Pro, Amersham Pharmacia). Then bacteria were induced with IPTG until the final concentration of 1mM and grown for subsequent 5 hours. After that time culture was centrifuge for 30' /6000rpm/+4 ^{o} and pellet was frozen in -80°C. Expression level was checked on SDS-PAGE Coomassie stained gels.

2.9. Proteins purification and refolding

2.9.1. Inclusion bodies preparation

The *E.coli* cells pellet was resuspended in buffer 1 (Table 2), complemented with protease inhibitor cocktail (Sigma Aldrich). Resuspended bacteria were disrupt by sonication on ice, 4x15''/15 units and centrifuged 10'/+4C ^{o}/10.000 rmp. The supernatant was discarded and cell paste was resuspended in cold buffer 2. Sonication was repeated as above and samples were centrifugeted 10'/10.000 rmp/4C^{o}. A second washing with buffer 2 was performed and centrifugation was repeated. The pellet was washed with buffer 3, centrifuged and resuspended in buffer 4. The samples were stirred for 30-60 min at room temperature to soulubilize inclusion bodies, centrifuged for 30'/20.000/4C ^{o} and passed through the 0, 45 filter (Millipore).

Inclusion bodies purification	
Buffer 1	20mM Tris-HCl, 0,5 M NaCl, 2% Triton X-100, 2mM DTT, pH 8
Buffer 2	20Mm Tris-HCl, 0,5M NaCl, 5mM imidazole, 6M guanidinium hydrochloride and 2mM DTT, pH 8
Buffer 3	20mM Tris-HCl, 0,5 M NaCl, 2% Triton X-100, 2mM DTT, pH 8
Buffer 4	20mM Tris-HCl, 0,5 M NaCl, 5mM imidazole, 6M guanidinium hydrochloride, 2mM DTT, pH 8
Affinity chromatography	
Buffer A	20Mm Tris-HCl, 0,5M NaCl, 5mM imidazole, 6M guanidinium hydrochloride, 2mM DTT, pH 8
Buffer B	20Mm Tris-HCl, 0,5M NaCl, 20mM imidazole, 6M urea, 2mM DTT, pH 8
Buffer C	20Mm Tris-HCl, 0,5M NaCl, 500mM imidazole, 6M urea, 2mM DTT, pH 8
Anion exchange chromatography	
Buffer D	Tris-HCl 20mM, 6M urea, 2mM DTT 1M NaCl, pH 8
Buffer E	Tris-HCl 20mM, 6M urea, 2mM DTT 1M NaCl, pH 8
Refolding buffers	
	20Mm Tris-HCl, 0,5M NaCl, 3 M urea, 3mM DTT, 10% glycerol, pH 8 20Mm Tris-HCl, 0,5M NaCl, 3 M urea, 3mM DTT, 10% glycerol, pH 8 20Mm Tris-HCl, 0,5M NaCl, 2mM DTT, 10% glycerol, pH 8

Table 2. Buffers used in purification and refolding of NSP5wt and its mutants.

2.9.2. FPLC purification using AKTÄ system

NSP5wt and its mutants were purified under denaturing conditions. Chromatography was performed using FPLC-AKTÄ system (Amersham Pharmacia). First, proteins solubilized from inclusion bodies were purified using immobilized metal affinity chromatography. The Ni^{2+}-NTA column (1 ml His-Trap, Amersham Pharmacia), was washed with water and equilibrated with 10 volumes of buffer A. Prepared inclusion bodies solution was applied on the column that was then washed with 10 volumes of buffer B. The proteins were eluted using a linear gradient of buffer C in buffer B by increasing concentration of imidazole.

Fractions were collected and checked on SDS-PAGE gel. Those containing NSP5 were pooled, dialysed against buffer D at $+4^{o}C$ and subjected to the second purification step, ion exchange chromatography. Samples were applied to the anion exchange monoQ HR5/5 column (Amersham Pharmacia) equilibrated before with 10 volumes of buffer D. The column was washed with 10 vol. of buffer D and next proteins were eluted with a linear gradient 0-100 % of buffer E (growing NaCl concentration). Fractions were collected, peaks were analysed on SDS-PAGE gel, and those containing pure protein were pooled.

Expression of the proteins
in *E.coli* BL21[DE3]

Inclusion bodies purification
and proteins solubilization

Affinity chromatography in denaturing
conditions- His-Trap column

Anion-exchange chromatography in
denaturing conditions- monoQ column

Protein refolding by dialysis
against buffer lacking urea
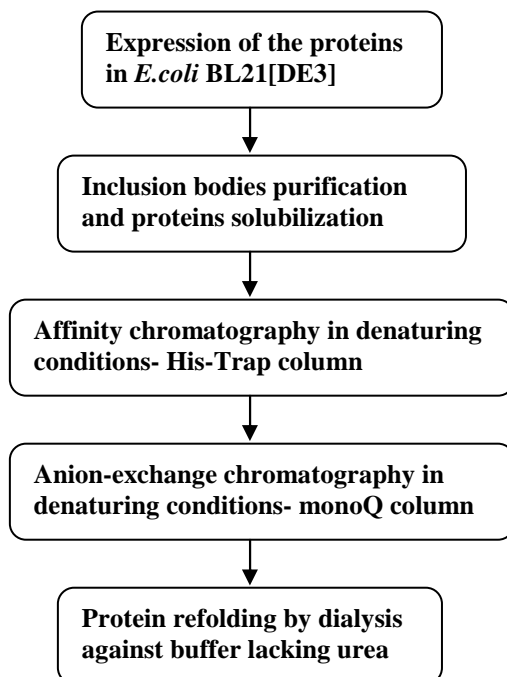
Fig. 24. General scheme for production, purification and refolding of His-tagged NSP5.

2.9.3. Proteins refolding

The samples were transferred to the membrane tubing (Spectrum Laboratories), cut off 12,000-14,000 Da and diluted to the concentration of 0,1 mg/ml in buffer E. Refolding was carried out by very slow (~48 h) dialysis in +4°C by exchanging the buffer containing urea to the one where urea was not present. First, protein was dialysed against buffer containing 3M urea. After 12 hours content of urea in dialysis buffer was decreased to 1,5M. During last 12 hours protein was dialysed against buffer without urea. 10% glycerol was always present in the buffer. Buffer exchange was performed in small volumes to ensure slow rate of denaturant concentration decrease. Buffer was changed often and usually only partially in purpose to avoid step gradient and make it more linear instead. Reducing agent (DTT) has always been added fresh to prevent forming of disulfide bridges. After refolding samples were concentrated using ultracentrifugation tubes (Sartoris). The correct molecular sizes of the purified products were confirmed by ESI-MS single quadrupole mass spectrometry (API-150EX, Applied Biosystems).

2.10. Size exclusion chromatography

Size exclusion chromatography was performed on FPLC-AKTÄ system (Amersham Pharmacia) using Superdex 200 column (Amersham Pharmacia) in buffer containing 20 mM Tris-HCl, pH 8.0, NaCl 150mM, DTT 2mM at a flow rate of 0.5 ml/min. Detection was at 280 nm. The column was precalibrated with seven protein standards of known masses and standard curve was made. Eluted fractions were collected, concentrated by speed-vac and Western blotting was performed.

2.11. Trypsin limited proteolysis

The mass ratio between trypsin and recombinant protein as well as reaction time was established empirically. Trypsin, 1μg, was used for digestion of 130 μg of recombinant

protein during time of 1h in room temperature. The used reaction buffer was Tris-HCl 20mM pH 8.0, NaCl 150mM, DTT 1mM. From the total reaction volume of 100 µl, samples of 10 µl in different time points were taken. To stop the digestion excess of SDS-PAGE loading buffer was added. All experiments were analysed on SDS-PAGE Coomassie-stained gels. Determination of the peptides masses was done thanks to the Termofinnigan LCQ deca Electrospray Ionization Mass Spectrometer. Band remained from the limited proteolysis of NSP5 wt was isolated from the gel and subjected to complete trypsin digestion and compared with parallel analysis of full-length protein.

2.12. Circular dichroism measurements

Proteins concentration was determined by UV absorbance measurement at 280 nm (Ultrospec3000, Amersham Pharmacia) using the calculated for every protein extinction coefficient values. CD measurements were performed on J-810 spectropolarimeter (Jasco Inc.) in the range of 250- 190 nm, using a 0.1 cm path length quartz cuvette (Hellma) in case of wt and S67D recombinant proteins and 0.02cm for and truncated mutants. Spectra were done at the concentration of 3.5, 4.5 and 5.0 µM respectively for S67D, wt and Δ32 mutant while ΔC19 protein required higher concentration and its measurements were recorded at concentration of 50 µM. All the spectra were averaged from 5 scans of 0.1 nm steps at 10 nm/min. The measured signals were converted to mean residue ellipticity (MRE, deg·cm^2/dmol) on the basis of a mean amino acid residue weight using the Dichroweb [236] server. All spectra were analyzed in terms of secondary structure using Dichroweb tools. The best fitting were represented by CONTIN[204] and CDSSTR programs.

2.13. Binding assay based on ELISA

Refolded NSP5 was tested in ELISA assay for binding activity of its molecular partner, NSP2, derived from MA104 cells infection with recombinant vaccinia virus VT7/LacOI/NSP2. ELISA plate (Nunc) was coated O/N on +4°C with 100 µl of NSP5 diluted to the concentration of 10µg/ml in PBS. In the morning the solution containing not

bound NSP5was discarded and plate was blocked by 1h incubation with PBS-BSA 1%. Then plate was washed with PBS-0.1%Tween20. NSP2 and mock-infected extracts (controls) were used to make serial dilutions. Plate was incubated in 37°C for 1,5 h and washed three times with PBS-0.01% Tween20. After that mouse anti-NSP2 sera was added at the dilution of 1:1000 and reaction was incubated for next 1h in 37°C. Plate was washed three times with PBS-0.01%Tween20 and secondary antibody, peroxidase conjugated goat anti-mouse IgGγ (KPL) at the dilution 1:5000, was added (Fig. 25). After 1h plate was washed three times with PBS-0.01%Tween20 and ELISA reaction was developed by adding 3.3',5,5'-tetramethyl-benzidine liquid substrate and then blocked with 2M HSO$_4$. Result was read in a plate reader (Model 550, Biorad) and plotted by Excel program.

anti IgG guinea pig HRP-conjugated Abs

guinea-pig sera anti NSP2

cellular extract containing NSP2

recombinant NSP5
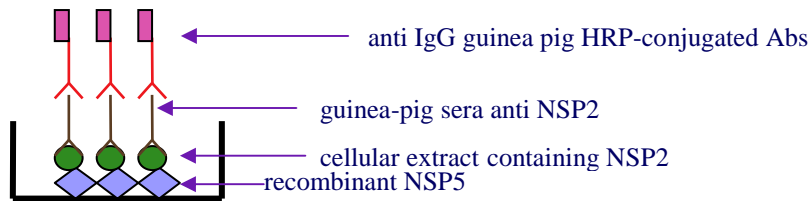
Fig. 25. Schematic representation of ELISA binding assay.

2.14. Peptides synthesis for the NSP5 phosphorylation by CK1α

Peptides were synthesized by a solid-phase method by using fluorenylmethoxycarbonl (Fmoc) chemistry. Peptide DMA-GG was synthesized by using Fmoc-*N*,*N*-dimethylarginine(Mts)-OH as described [237]. Peptides were purified by RPHPLC on a C-18 column (Waters RCM, 25 x100 mm) by using a linear gradient of acetonitrile in water (0–35% over 70 min) containing 0.1% trifluoroacetic acid, and their identity was confirmed by electrospray ionization MS (PerkinElmer SCIEX API- 150EX) and stored as lyophilized powder.

2.15. Expression and purification of His-6-tagged proteins for the *in vitro* kinase assay with recombinant CK1α

The proteins were prepared as described in [6]. pQE-Δ3, pQE-Δ3*a*, and pQE-Δ3/S67A were expressed in *Escherichia coli* M15[pREP4]. Culture was induced with isopropyl β-D-thiogalactoside (1 mM), grown for 4 h, and centrifuged at 3,500 rpm for 10 min. Pellet was resuspended in 6 ml of PBS and incubated for 15 min on ice with 0.1 mg_ml lysozyme, 5mMDTT, 1.5% laurylsarcosine, and mixture inhibitor protease (Sigma). Lysate was sonicated (six times for 10 s each) and centrifuged at 10,000 rpm for 10 min. Supernatant was supplemented with 1% Triton X-100 and loaded onto a nickel column (Novagen) previously equilibrated in 5 vol of loading buffer (20 mM imidazol in PBS). The beads were washed in 10 vol of washing buffer (35 mM imidazol in PBS) and eluted in 2 vol of elution buffer (250 mM imidazol in PBS). The eluted protein was dialyzed against PBS and analyzed by SDS-PAGE and Coomassie blue staining. pQE-wtNSP5, pQE-NSP5/(S63,65A,S67D), and pQE-NSP5a were expressed as described above. Pellet was resuspended in 4 ml of 20 mM Tris_HCl (pH 8.0) and inhibitor protease mixture, sonicated (four times for 10 s each), and centrifuged at 10,000 rpm for 10 min. Inclusion bodies were resuspended in 3 ml of cold 2 M urea, 20mMTris-HCl, 0.5MNaCl, and 2% Triton X-100 (pH 8.0), sonicated, and centrifuged as above. A second wash in urea was performed, and a last wash was performed in a buffer lacking urea. The pellet was resuspended in 5 ml of 20mMTris-HCl, 0.5MNaCl, 5 mM imidazol, and 6Mguanidine hydrochloride (pH 8.0), stirred for 45 min at room temperature, and centrifuged 15 min at 10,000 rpm. Supernatant was passed on a 0.45- μ m filter and loaded onto a preequilibrated nickel column (Novagen). Column was washed five times in 20 mM Tris-HCl, 0.5 M NaCl, 20 mM imidazol, and 6 M urea (pH 8.0) and eluted in 10 ml of 20 mM Tris_HCl, 0.5 M NaCl, 6 M urea, and 150 mM imidazol, pH 8.0. The eluted protein was refolded by consecutive dialysis in 3, 1.5, and 0M urea in 20mM Tris-HCl (pH 8.0), 250 mM NaCl, and 1 mM DTT.

2.16. *In vitro* CK1α phosphorylation assay

The assay has been described by Marin *et al.* [238]. In brief, synthetic peptides substrates (0.2 mM) were phosphorylated by incubation in a medium (30 μl of final volume) containing 50 mM Hepes buffer (pH 7.5), 10 mM MgCl2, 150 mM NaCl, and 50 μM γ-ATP (specific radioactivity 4,500– 8,000 cpm/pmol). The reaction was started with addition of 16 units (107 units/pmol) of recombinant CK1α from zebrafish (*Danio rerio*) [239]. After incubation at 30°C for the times indicated, the mixture was spotted onto phosphocellulose paper, which was washed three times in cold 75 mM H3PO4. Filter-associated radioactivity was measured by scintillation counting. Kinetic constants were determined by regression analysis of double-reciprocal plots constructed from initial-rate measurements. Protein phosphorylation was carried out in a reaction buffer containing 50 mM Hepes buffer (pH 7.5), 10mMMgCl2, 150mMNaCl, 50 μM[32P]- γ- ATP (12,000 cpm/pmol), and 1–2 μM of protein substrate, in a total volume of 14μl. Then 23 units of recombinant CK1α (65 units/ pmol) was added to initiate reactions; incubation was for 30 min at 30°C. Samples were analyzed by 12.5% SDS_PAGE and radioactive bands were detected in a Molecular Imager FX (Bio-Rad) apparatus.

3. RESULTS

The results presented in this dissertation have been divided into three parts. The first chapter demonstrate the effort of the laboratory that was focused on studying the mechanism of NSP5 hyperphosphorylation. The work presented here has been done with a kind help of Dr. Catherine Eichwald and in collaboration with Prof. Jorge E. Allende from Instituto de Ciencias Biomedicas, Programa de Biologia Celular y Molecular, Facultad de Medicina, University of Chile and resulted in a paper issued in PNAS [6].

In the second chapter I have described my work on mapping of post-translational modifications of NSP5 where we used MALDI-TOF/TOF spectroscopy as a tool. The collected data allowed us to draw certain conclusions how particular regions of NSP5 are modified.

Finally, in the third part, I have presented structural studies on NSP5, with the use of circular dichroism spectroscopy, that let us to draw first conclusions concerning the structure of this viral protein.

Part 1: Studies on hyperphosphorylation mechanism of NSP5

NSP5 was originally characterized to have a molecular mass of 26kDA on SDS-PAGE. Further studies demonstrated that the 26kDa form is a precursor for subsequent processing. SDS-PAGE analysis of rotavirus infected cells shows that NSP5 appears as two major bands of 26 and 28kDa and as series of higher bands whose apparent molecular mass spans from 30 to 34kDa. It was shown that the 26 and 28kDa forms are phosphorylated and glicosylated by addition of O-linked N-acetyloglucosamine (O-GlcNAc) monosaccharide residues to serines and threonines, while the highest bands are present mostly due to hyper-phosphorylation. Experiments performed in our laboratory have suggested that NSP5 hyperphosphorylation is the consequence of a complex autoregulatory mechanism. The process involves mobilization of cellular kinases and since it is unlikely that NSP5 posses kinase activity by itself, it may have a capability of promoting the activation of cellular kinase(s) [149], following interaction with NSP2.

We speculated the existence in NSP5 of regions specific for the activation of a cellular kinase and for its own phosphorylation. With the aim of understanding the molecular bases of this process, we decided to study a number of different NSP5 deletion mutants on their ability to become phosphorylated in vivo. These studies revealed, that two clearly distinct properties of this protein, related to its own phosphorylation could be uncoupled: one such property was related to its capacity to be a substrate and the other one was related to its capacity to become an activator of the phosphorylation process.

To test this hypothesis in vivo, a transfection assay where two mutants of NSP5 were chosen as a substrate, was developed. One mutant, $\Delta2$, was used since it is not phosphorylated in vivo [152], while it is a good substrate in vitro. The second mutant, $\Delta4$, was used since it is not phosphorylated in vivo [152] and is not a substrate in vitro. To distinguish these two mutants from other mutants used as activators of cellular kinases in a co-transfection assay, the eleven amino acid SV5 tag was added at the N-terminus of the substrate proteins. A co-transfection assay was performed using mutants that behave as activators, $\Delta1/\Delta3$ and $\Delta3$, with the substrates SV5-$\Delta2$ or SV5-$\Delta4$. To visualize the effect of the activators with the different substrates, the cellular extracts were loaded in a SDS-PAGE gel and a Western blot anti-SV5 was performed. The results presented in figure 26 show that the SV5-$\Delta2$ substrate is able to produce a mobility shift in the presence of the kinase activators (figure 26B, lanes 2 and 3). In

contrast, the SV5-Δ4 substrate was not able to produce this mobility shift (lanes 5 and 6). A treatment with λ-Ppase confirmed that the mobility shift present in these samples corresponded to hyper-phosphorylation of the SV5-Δ2 substrate (figure 26C) [6].

The conclusion of these experiments is, that the deletion of regions 1, 3, or both (mutants Δ1, Δ3, or Δ1/Δ3) renders transfected NSP5 able to be hyperphosphorylated, suggesting that these two domains act as internal inhibitors in the phosphorylation mechanism [49] [152] [149]. The mutant lacking region 2 (Δ2) is a substrate in this process and the activator molecule was acting in trans in relation to the substrate [6].
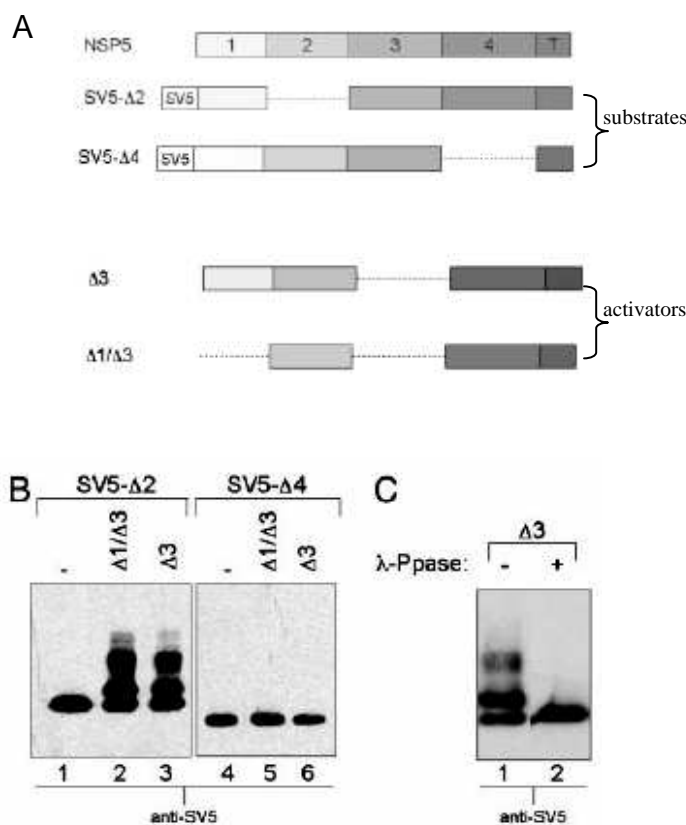
Fig. 26. Coexpression of SV5-Δ2 or SV5-Δ4 mutants with activators Δ1/Δ3 and Δ3. (*A*) Schematic representation of NSP5 and the deletion mutants constructs. (*B*) Coexpression assay in MA104 cells of SV5-Δ2 (lanes 1–3) or SV5-Δ4 (lanes 4–6) substrates with or without the indicated activators. (*C*) λ-Ppase treatment of SV5-Δ2 cotransfected with activator Δ3. Samples were loaded in a 15% SDS-PAGE, and substrates were visualized by Western blot. Modified from [6].

Ser-67 and the Activation Function

From the results presented above, activation and substrate activities could be clearly distinguished. As described above, SV5-Δ2 was a good substrate both in the in vivo hyper-phosphorylation assay as well as in the in vitro translation/phosphorylation kinase assay. On the other hand, Δ2 was unable to induce its own phosphorylation. Thus, we decided to study role of the region 2 in the activation of the putative cellular kinase. We therefore focused our attention into two different Ser-rich motifs within this domain, namely: Ser-Asp-Ser-Ala-Ser from residue 63 to residue 67 and Ser-Phe-Ser-Ile-Arg-Ser from residue 73 to residue 78, that we termed motifs *a* and *b*, respectively (Fig. 27A). We hypothesize that a mechanism of hierarchical phosphorylation could be at the basis of the activation step. The hierarchical phosphorylation implies that certain residues need to be phosphorylated first in order to make available other sites. To investigate whether either of the two motifs were involved in the activation step, they were independently mutated, in the activator construct Δ3, transforming Ser residues into Ala. As shown in Fig. 27, mutation of all three serines of motif a (Ser-63, Ser-65, and Ser-67, construct Δ3a) completely abolished activation function, whereas mutation of Ser in motif *b* (construct Δ3b) had no effect. A more detailed analysis of Ser in motif *a* revealed a key role for Ser-67. In fact, the single Ser67Ala mutation eliminated activation, suggesting that phosphorylation of this residue was required for activation function (Fig. 27C). This was further confirmed with a construct in which Ser-67 was mutated to Asp to mimic a phosphorylated residue, with the two other Sers in motif *a*, Ser-63 and Ser-65, mutated to Ala. As shown in Fig. 27C, lane 5, the Ser67Asp mutation (construct Δ3/S63,65A/S67D) restored activation. Furthermore, the single Ser-67Asp mutation was sufficient to render, the otherwise inactive full-length NSP5, able to be hyperphosphorylated despite the presence of the inhibitory domains 1 and 3 (Fig. 27D).
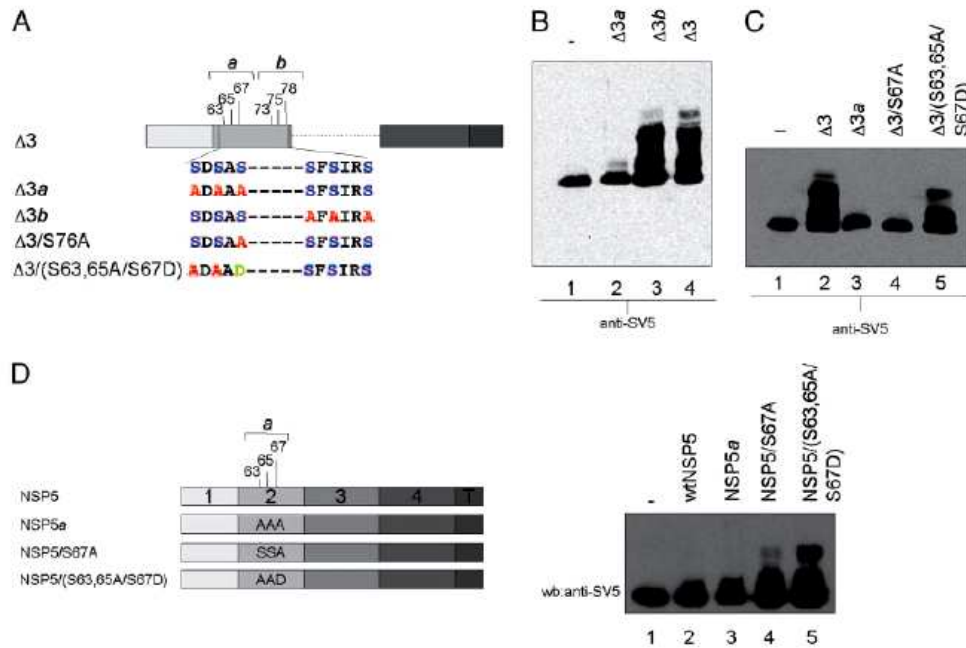
Fig. 27. Ser-67 from motif *a* is essential for activation. (A) Scheme of point mutations in the activator Δ3. Motifs *a* and *b* are indicated. The S to A and S to 3D modifications are indicated in red and green, respectively. (B) Coexpression of SV5- Δ2 substrate with activators Δ3a, Δ3b, or Δ 3. (C) Coexpression of SV5- Δ2 substrate with activators Δ3, Δ3a, Δ3ΔS67A, or Δ3(S63,65A,S67D). (D) Scheme of full-length NSP5 with point mutations on Sers in motif a (Left), and coexpression of SV5-Δ2 with the NSP5 point mutants (Right). Cellular extracts were loaded on a 15% SDS-PAGE, and substrate was visualized by Western blot [6].

Taken together these results demonstrated that in a first step, phosphorylation of Ser-67 is necessary for NSP5 to become an activator: once phosphorylation of Ser-67 has taken place, both domains 1 and 3 do not hinder hyperphosphorylation any longer. It appears that these domains prevent phosphorylation of Ser-67, probably by impairing its accessibility. This interpretation was supported by the fact that wtNSP5, and more importantly the activation negative NSP5a, (with Ser in motif *a* mutated to Ala), were efficiently hyperphosphorylated when coexpressed with an activator (such as Δ1/Δ3), but not when expressed alone (Fig. 28A). Furthermore, transfected NSP5/S67D, but not NSP5/S67A or NSP5a, became hyperphosphorylated, in a way similar to the protein obtained from virus-infected cells (Fig. 28B).
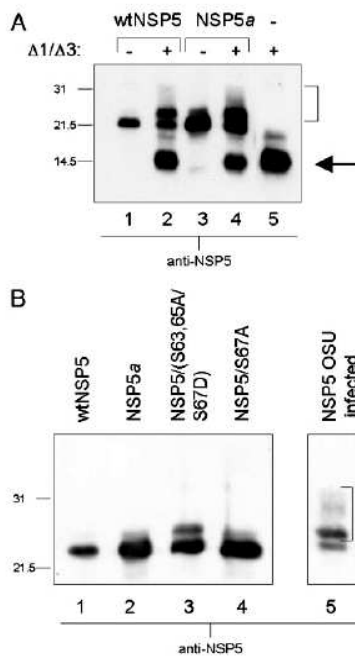
Fig. 28. Activator-dependent wt NSP5 phosphorylation. (A) Coexpression in MA104 cells of wtNSP5 or NSP5a with the activator Δ1/ Δ3. Immunoblot developed with anti-NSP5 serum. The arrow indicates the position of Δ1/ Δ3. Bracket indicates NSP5 phosphorylated forms. (B) Expression of full-length wtNSP5 and NSP5 point mutants. Lane 5 corresponds to NSP5 from rotavirusinfected MA104 cells. Samples were migrated in a 15% SDS/PAGE and were visualized by Western blot [6].

Ser-67 Phosphorylation and CK1

The canonical consensus sequence for phosphorylation by CK1α is D/EXXS*/T*, where X is any amino acid, D is aspartic acid, E glutamic acid and S*/T* are the target serine or threonine that is phosphorylated [240]. The sequence similarity between the consensus phosphorylation sites of protein kinase 1, also known as CK1 and motif *a*, suggested that CK1 could be involved in the phosphorylation of Ser-67. This possibility was investigated in an *in vitro* kinase assay by using recombinant CK1α (from zebrafish) and two synthetic substrate peptides of 14 residues encompassing motif *a*, from residues E58 to L71: peptide *NSP5wt* had the NSP5 wt sequence whereas peptide *NSP5(S67A)* contained a Ser to Ala mutation in the position corresponding to Ser-67. As shown in Fig. 29A, *NSP5wt* was a substrate of CK1α, whereas peptide *NSP5(S67A)* was not phosphorylated at all. When compared to a CK1 consensus specific substrate peptide (RRKHMIGDDDDAYSITA) [239], phosphorylation of peptide *NSP5wt* by CK1α showed a similar *K*m (specific peptide K_m=283μM; *NSP5wt* K_m=218μM) but a lower catalytic constant of 1.7 min^{-1} for the *NSP5wt* peptide vs. 50 min^{-1}

80

for the specific substrate. This result indicated CK1 as a candidate enzyme involved in the phosphorylation of Ser-67 to render NSP5 susceptible to hyperphosphorylation. Similar experiments with recombinant CK2 showed that this kinase does not phosphorylate Ser-67 (data not shown). However, when the involvement of CK1 in NSP5 hyperphosphorylation was further investigated in the *in vitro* assay by using as substrate the bacteria-expressed proteins, a surprising result was obtained. CK1α was able to phosphorylate only the activation positive substrates Δ3 or NSP5/(S63/65A/S67D) generating mobility shifted species with higher apparent molecular weights. The whole process depended on the phosphorylation of Ser-67 because the Ser-67Ala point mutation (construct Δ3/S67A) as well as mutation into Ala of all three Ser residues of motif *a* (constructs Δ3*a* and NSP5*a*) completely abolished phosphorylation activity (Fig. 29 B). This result strongly suggested that CK1 or a CK1-like activity, directly participates in NSP5 phosphorylation *in vivo*.
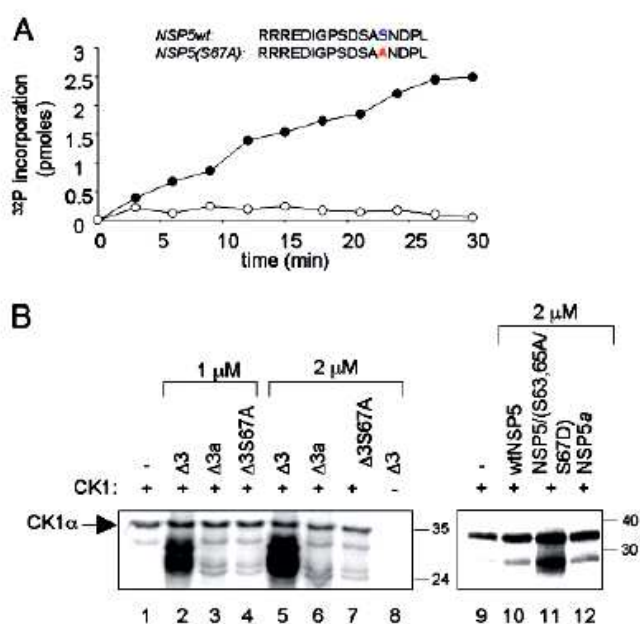
Fig. 29. NSP5 phosphorylation by recombinant CK1 from zebrafish. (*A*) Time course of CK1α phosphorylation of motif *a* peptides derived from NSP5. The amino acidic sequence of both peptides is indicated. *NSP5wt*(●)a; *NSP5*(*S67A*) (○). (*B*) SDS-PAGE autoradiography of *in vitro* phosphorylation assay of mutants Δ3, Δ3/S67A, and Δ3*a* (*Left*) or wtNSP5, NSP5/(S63,65A/S67D) and NSP5*a* (*Right*), with recombinant CK1α and [³²P]-γATP. The 38-kDa band corresponds to autophosphorylated CK1α. Reactions were carried out at the indicated concentrations of substrate proteins. Negative control reactions containing either CK1α alone (lanes 1 and 9) or substrate Δ3 alone (lane 8) are shown [6].

From the obtained data we would like to propose a model of NSP5 hyperphosphorylation (Fig. 30). In this model the phosphorylation of Ser-67 by CK1α is

followed by, or happen concomitantly with protein dimerization. Phosphorylated Ser-67 component activates the other monomer of the dimeric complex rendering it (or both) susceptible to hyperphosphorylation, by the same CK1 or a CK1-like enzyme. It is known, that the tail is involved in dimerisation [161] [149], and since the presence of C-terminal part of NSP5 was shown to be necessary for both, substrate and activator function (experiments performed by dr. Catherine Eichwald), we suggest that the interaction between an activator and a substrate is required for the final hyperphosphorylation effect [6].
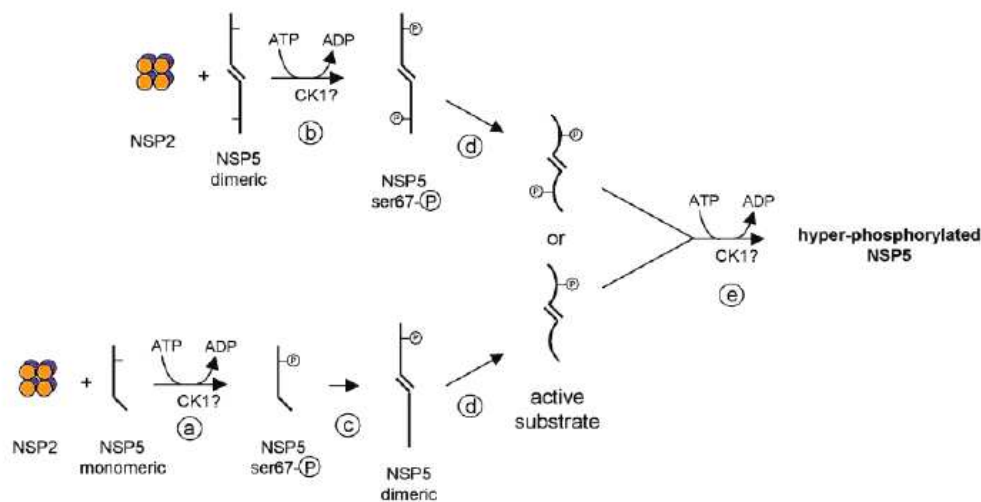
Fig. 30. Model for rotavirus NSP5 hyperphosphorylation. Interaction of NSP2 with monomeric (*a*) or dimeric (*b*) NSP5 promotes phosphorylation of Ser-67 by CK1 (in either one or both subunits), which in the case of the monomer may induce dimerization (*c*). A phosphorylated Ser-67 in the dimer induces conformational changes (*d*) to render the other partner molecule (and probably both) susceptible to hyperphosphorylation by the CK1-like cellular kinase (*e*) [6].

Part 2: Post-translational modification studies: mapping phosphorylation and glycosylation regions in NSP5

It was previously reported that NSP5 is modified by two types of post-translational modifications: O-glycosylation [146] and phosphorylation [3] [49] [241]. Glycosylation of NSP5 consists in the addition of O-linked N-acetyloglucosamine (O-GlcNAc) monosaccharide residues to serines and threonines, performed by cytoplasmic O-GlcNAc transferase [147]. Although not fully understood, the importance of O-GlcNAcylation is apparent, since many of known GlcNAc-modified proteins are also phosphoproteins [148]. In many cases the sites of phosphorylation and GlcNAcylation are localized to the same or neighboring residues (serines or threonines). Additionally, O-GlcNAc modification is transient (due to the activity of neutral β-N-acetylglucosaminidase [147]) so direct interplay between glycosylation and phosphorylation can take place, regulating the protein activity. Studies performed in our laboratory showed that NSP5 hyper-phosporylation is a complex process that is triggered by phosphorylation of serine 67 by casein kinase 1α (CK1α) [6] [242].

The correlation between these two types of post-translational modifications, their range and meaning for the activity and function of NSP5 have not yet been fully elucidated.

With the objective to map posttranslational modifications of NSP5, we performed studies on the protein using MALDI TOF/TOF spectroscopy. Different eukaryotic expression systems that could be used in our laboratory for the production of NSP5, like recombinant vaccinia virus or Sindbis virus replicon expression system did not ensure appropriate expression levels as well as did not offer the same profile of modifications that occur in this protein during rotavirus infection. Since these modifications could play an important role on NSP5 activity during the viral life cycle and, in particular, during the replication process, we preferred to use rotavirus infection as a source of NSP5.

To ensure the accuracy of analysis different experiments were performed with about 50 millions infected cells every time. In some experiments cells were infected with the porcine strain OSU, and in others with the bovine RF strain. The sequences of NSP5 encoded by these strains differ slightly, so peptides resulting from the trypsin digestion of the same regions of the protein have different masses. Additionally, samples were prepared for MALDI analysis by two different methods (described in detail in the "Material and methods" chapter). In both cases NSP5 was immunoprecipitated with an anti-NSP5 polyclonal antibodies and

protein-A beads but further processing was different. In the first method, protein was trypsinized while still bound to the beads. This method allowed analysis of the entire range of NSP5 modifications present in the infected cell. In the second method, the immunoprecipitated protein was resolved by PAGE and then different bands corresponding to isoforms with different load of post-translational modifications were excised, digested and analyzed (Fig. 31). Peptides coming from the lower bands of the gel should contain less phosphate than the ones deriving from the upper bands.

The diversity of the sequence of the analyzed proteins as well as the sample preparation methods were used with the purpose to diminish occurrence of false results and to contribute to their robustness. There was a chance that peptides coming from contaminants or differently modified peptides from different regions of NSP5 could have very similar masses that could cause confusion in data interpretation.
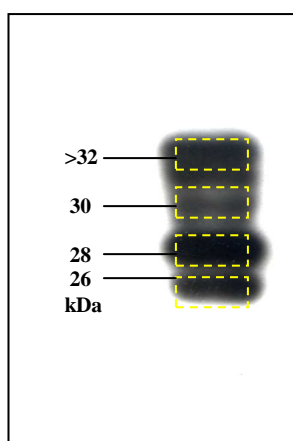
Fig. 31. Western blotting of immunoprecipitated NSP5 derived from rotavirus infection of MA104 cells. Dashed squares represent way of protein fractionating and excited fragments of the gel that were next trypsinized. The peptides were analyzed by MALDI TOF/TOF spectrometry.

The spectra were recorded on a 4800 MALDI TOF/TOF™ tandem TOF technology analyzer (Applied Biosystems/MDS SCIEX). The masses of the peaks were then analyzed by MASCOT tool (http://www.matrixscience.com/), which is based on a probabilistic model of the number of matching masses [235]. Unfortunately, because of low intensity of NSP5 peptides derived peaks, contaminations coming from cellular proteins covered NSP55 signal and the peptides masses were not associated with NSP5 with high scores. After spectra cleaning by erasing peaks present also in the control samples (i.e.: immunoprecipitations of mock infected cells), the program clearly assigned peptides as belonging to rotavirus NSP5 (Fig. 32).
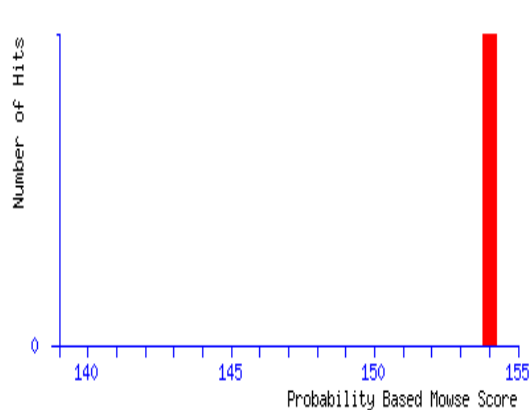
Fig. 32. Mascot Search Results- original output of the program.

Search title   : NSP5
Database   : MSDB 20060831 (3239079 sequences;
1079594700 residues)
Taxonomy   : Other viruses (305319 sequences)
Top Score   : 154* for S30217, NS26 protein - human
rotavirus A
*Protein score is -10*Log(P), where P is the probability that the
observed match is a random event.   Protein scores greater than
67 are significant (p<0.05).

The example of a spectrum, in this case resulting from the analysis of samples prepared by the direct trypsinization method, is shown in Fig. 33. The masses associated with dashed arrows perfectly fit the theoretical masses of the peptide that would derive from a tryptic digestion encompassing the fragment of NSP5 sequence spanning from amino acid 56 until 77, either non modified (Fig. 33B) or with one site phosphorylated (Fig. 33C). That peptide contain serine 67 that was earlier recognized to be part of a non canonical phosphorylation site for casein kinase 1α as shown in part 1 of this thesis and in reference [6], and it is highly probable that phosphorylation on this peptide occurs exactly at this serine.

B

SPEDIGPSDSA**S**NDPLTSFSIR

C

SPEDIGPSDSA**S**NDPLTSFSIR**+(Phos)₁**
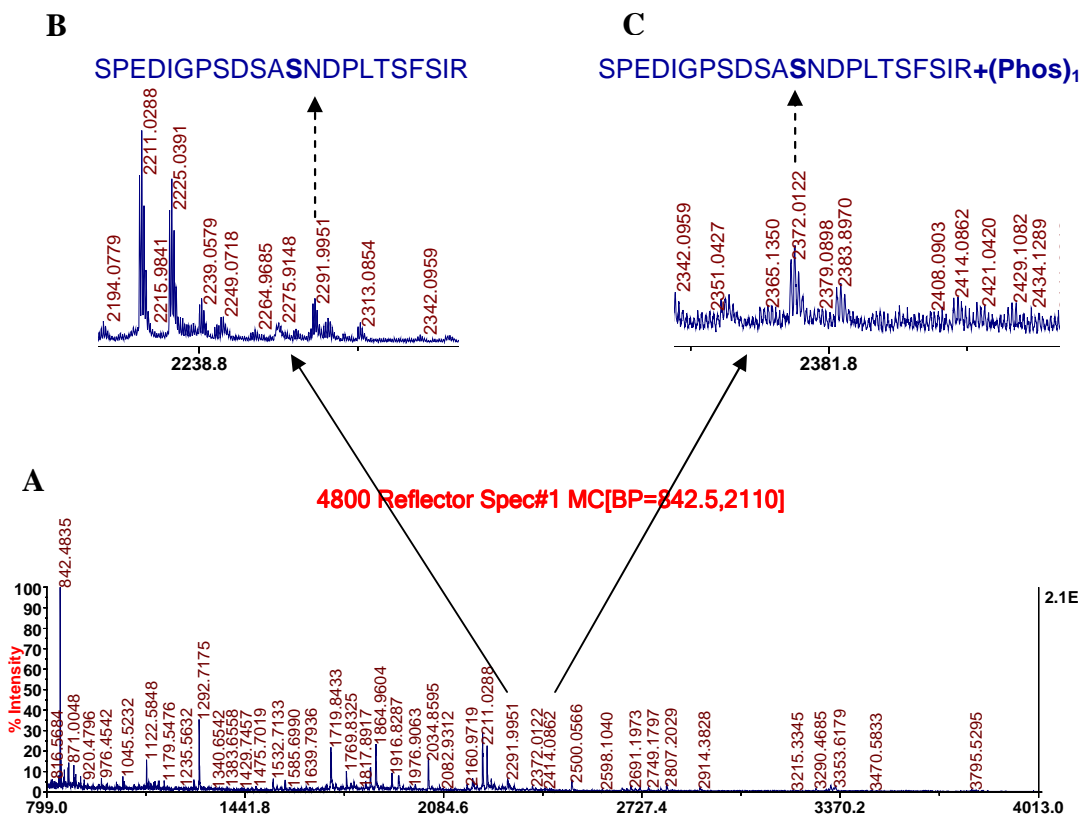
Fig. 33. Fragment of the experimental spectrum resulting from the analysis of samples derived from direct trypsin digestion of immunoprecipitated NSP5 sample (A). The upper panels show peptides that represent the region that encompass serine 67 (marked in bold), not modified (B) and with one of serines or treonine phosphorylated (C).

Selected masses were subjected to further manual processing with help of the PeptideMass program (www.expasy.org) that allows executing theoretical trypsin proteolysis of chosen protein with the purpose to find among them nonmodified NSP5 peptides. Subsequently masses were also analyzed by the GlycoMod program (available also on www.expasy.org site) that allows to identify phosphorylation and glycosylation modifications. Since NSP5 was demonstrated to bear phosphate residues and O-linked N-acetyloglucosamine monosaccharide residues (at serines and threonines), initially only these modifications were especially taken into consideration. The difference between theoretical and experimental values was established to not exceed 0.2 Da. Moreover, recently in our laboratory one new experiment with involvement of another rotavirus strain, simian SA11, was performed. Unfortunately the analysis revealed the absence of many peptides, that were found in previous studies. This fact convinced us to extend the range of considered posttranslational modification to the possibility of occurrence of two others- methylation and

acetylation. The inclusion of these two modifications to the analysis let us to reveal the presence of additional peptides of the protein that previously were not found in experiments, nevertheless the potency of this findings must be verified by biochemical experiments performed on infected cells.

Protein methylation is a post-translational modification by which a methyl group from S-adenosylmethionine is added to a protein by enzymes from the protein arginine methyltransferases family (PRMTs). PRMTs are located in the cytoplasm, in the nucleus or in both of them [243]. In eukaryotes, proteins can be methylated on the side chain nitrogens of arginine, lysine, and histidine residues or on the carboxyl groups of proteins. Methylation on side chain nitrogens is considered largely irreversible while methylation of the carboxyl groups is potentially reversible [244]. Arginine methylation is implicated in many cellular processes including transcription, RNA processing and transport, translation, signal transduction, DNA repair, apoptosis etc. It increases bulkiness of the affected residue and its hydrophobicity blocks hydrogen bonding but, in contrast to phosphorylation or acetylation, does not alter its charge. The maintenance of charge is particularly important because positively charged arginines often participate in binding other proteins or nucleic acids. In addition to its involvement in the pathology of diseases, protein methylation has also been shown to be important for virus replication and infectivity in a variety of viruses. For example Herpes simplex virus (HSV) replication is regulated, in part, by methylation of the RNA binding domain in the HSV ICP27 protein [245].

Acetylation is the post-translational modification of histone and non-histone proteins that is part of an important cellular signaling process, controlling a wide variety of functions in both the nucleus and the cytoplasm. It is a reversible process, in which histone acetyltransferases (HATs) transfer the acetyl moiety from acetyl coenzyme A to the amino groups of internal, highly conserved lysine residues [246]. Recent investigations designate this signaling pathway as one of the primary targets of viral proteins after infection. Indeed, specific viral proteins have acquired the capacity to interact with cellular acetyltransferases (HATs) and deacetylases (HDACs) and consequently to disrupt normal acetylation signaling pathways, thereby affecting viral and cellular gene expression.

The calculations encompassing these two additional modifications were done with higher stringency. The peptides supposed to carry methylation and/or acetylation were also considered as possibly true, however were distinguished in tables in *italic*. Results of all performed experiments, divided by strain, are collected in tables 3, 4 and 5.

OSU strain.

The sequence:
MSLSIDVTSLPSISSSIFKNESSSTTSTLSGKSIGRSEQYISPDAEAFNKYMLSKSPEDIGP
SDSASNDPLTSFSIRSNAVKTNADAGVSMDSSTQSRPSSNVGCDQVDFSLTKGINVSA
NLDSCVSISTDNKKEKSKKDKSRKHYPRIEADSDSEDYVLDDSDSDDGKCKNCKYKK
KYFALRMRMKQVAMQLIEDL

Abbreviations:
MC- trypsin digestion missed cleavage
HexNAc- N-acetyloglucosamine
Phos- phosphate
Acetyl- acetylation
Methyl- methylation
MSO- methionine oxidized
Cys_PAM-cystein with acrylamide adduct (only in- gel digestion method)
[dir]- direct trypsinization method
[26/28/ 30 / <32] – in- gel trypsinization method and band fraction where the peptide was found

Sequence of the peptide	Peptide position	MC	Theoretical mass	Experimental mass	Modification
MSLSIDVTSLPSISSSIFK	1-19	0	2012.06	2011.99[dir]	no modifications
			2295.10	2295.24[dir]	$(HexNAc)_1$ $(Phos)_1$
			2717.26	2717.36[dir]	MSO $(HexNAc)_3$ $(Phos)_1$
			3326.50	3326.46[dir]	MSO $(HexNAc)_6$ $(Phos)_1$
			2272.139	*2272.16[dir]*	*MSO $(HexNAc)_1$ $(Acetyl)_1$*
			2634.218	*2634.273[dir]*	*$(HexNAc)_3$ $(Methyl)_1$*
NESSSTTSTLSGK	20-32	0	1298.60	1298.68[26]	no modifications
			2110.92	2110.93[dir]	$(HexNAc)_4$
			2190.89	2191.05[dir]	$(HexNAc)_4$ $(Phos)_1$
NESSSTTSTLSGKSIGR	20-36	1	2118.08	2118.01[28]	$(HexNAc)_2$
			1995.013	1994.892[28]	$(HexNAc)_1$ $(Phos)_1$

SIGRSEQYISPDAEAFNK	33-50	1	2011.97	2011.99[dir]	no modifications
SIGRSEQYISPDAEAFNKYMLSK	33-55	2	2874.18 *2915.189*	2874.25[dir] *2915.285[dir]*	(Phos)$_3$ *(HexNAc)$_3$ (Methyl)$_3$*
SEQYISPDAEAFNK	37-50	0	1598.73 1678.70 1801.81	1598.74 [30/> 32] 1678.70 [>32] 1801.90 [26]	no modifications (Phos)$_1$ (HexNAc)$_1$
SEQYISPDAEAFNKYMLSK	37-55	1	2221.04 2440.12 2437.135	2221.05[dir] 2440.12[dir] 2437.214[dir]	no modifications MSO (HexNAc)$_1$ *(HexNAc)$_1$ (Methyl)$_2$*
SPEDIGPSDSASNDPLTSFSIR	56-77	0	2292.06 2372.02	2292.05[dir] 2372.06 [30/>32]	no modifications (Phos)$_1$
YMLSKSPEDIGPSDSASNDPLTSFSIR	51-77	1	2914.37	2914.44 [dir]	no modifications
TNADAGVSMDSSTQSRPSSNVGCDQVDFSLTK	83-114	0	*3822.668*	*3822.854[dir]*	*(HexNAc)$_2$ (Acetyl)$_1$ (Methyl)$_1$*
GINVSANLDSCVSISTDNK	115-133	0	2219.97 2370.97	2220.17[dir] 2371.13[26]	(HexNAc)$_1$ (Phos)$_1$ Cys_PAM (HexNAc)$_1$ (Phos)$_2$
GINVSANLDSCVSISTDNKK	115-134	1	2419.05 2243.014	2419.13[>32] *2242.93[dir]*	Cys_PAM (HexNAc)$_1$ (Phos)$_1$ *(Acetyl)$_1$(Phos)$_1$*
SKKDKSR	137-143	3	1051.57	1051.63 [dir]	(HexNAc)$_1$

KDKSRK	139-144	3	761.46 *788.486*	761.46[dir] *788.494[dir]*	no modifications *(Methyl)$_1$*
SRKHYPR	142-148	2	943.52 1023.48	943.6[dir] 1023.60[dir]	no modifications (Phos)$_1$
SRKHYPRIEADSDSEDYVLDD SDSDDGK	142-169	3	*3443.371*	*3443.504[dir]*	*(Phos)$_2$* *(Acetyl)$_1$* *(Methyl)$_2$*
HYPRIEADSDSEDYVLDDSD SDDGK	145-169	1	3452.41	3452.39[dir]	(HexNAc)$_3$
IEADSDSEDYVLDDSDSDDG KCKNCKYKK	149-177	4	3285.40	3285.48[dir]	no modifications
CKNCKYK	170-176	2	886.42	886.49[dir]	no modifications
KYFALRMRMK	178-187	3	1343.74	1343.86[dir]	no modifications
YFALRMR	179-185	1	*983.537*	*983.571[dir]*	*(Methyl)$_2$*
MRMKQVAMQLIEDL	184-197	2	1705.87	1705.93[dir]	no modifications
MKQVAMQLIEDL	186-197	1	1418.73	1418.73[dir]	no modifications

Tab. 3. Collection of peptide masses observed upon OSU strain infections together with fitting posttranslational modifications. Possible modification that were considered are phopshorylation and glycosilation, masses containing methylation and acetylation were conditionally included and are written in *italic*.

RF strain.

The sequence:

MSLSIDVTSLPSISSSIYKNESSSTTSTLSGKSIGRSEQY ISPDAEAFSKYMLSKSPEDI
GPSDSASNDPLTSFSIRSNAVKTNADAGVSMDSSTQSRPSSNVGCDQVDFSFNKGISM
NANLDSSISISTSSKKEKSKSDHKSRKHYPKIEAESDSDDYILDDSDSDDGKCKNCKY
KRKY FALRMRMKQVAMQLIEDL

Abbreviations:
MC- trypsin digestion missed cleavage
HexNAc- N-acetyloglucosamine
Phos- phosphate
Acetyl- acetylation
Methyl- methylation
MSO- methionine oxidized
Cys_PAM-cystein with acrylamide adduct (only in- gel digestion method)
[dir]- direct trypsinization method
[26/28/ 30 / <32] – in- gel trypsinization method and band fraction where the peptide was found

Sequence of the peptide	Peptide position	MC	Theoretical mass	Experimental mass	Modification
MSLSIDVTSLPSISSSIYK	1-19	0	2247.13	2247.14[28/30]	MSO (HexNAc)$_1$
			2637.29	2637.31[dir]	(HexNAc)$_3$
			2653.29	2653.45[dir]	MSO (HexNAc)$_3$
			2327.10	2327.13[dir]	MSO (HexNAc)$_1$ (Phos)$_1$
			2717.26	2717.37[dir]	(HexNAc)$_3$ (Phos)$_1$
			3299.34	3299.49[dir]	MSO (HexNAc)$_5$ (Phos)$_3$
			3342.49	3342.56[>32]	MSO (HexNAc)$_6$ (Phos)$_1$
			2551.00	2550.84[dir]	(HexNAc)$_1$ (Phos)$_4$
			2636.287	*2636.307[dir]*	*(HexNAc)$_2$ (Phos)$_2$ (Methyl)$_1$*
			3152.422	3152.298[dir]	*(HexNAc)$_5$ (Phos)$_1$ (Methyl)$_1$*

MSLSIDVTSLPSISSSI YKNESSSTTSTLSGK	1-32	1	3510.72	3510.73[dir]	(HexNAc)$_1$
			3713.676	*3713.607[dir]*	*(HexNAc)$_1$ (Phos)$_2$ (Methyl)$_2$*
NESSSTTSTLSGK	20-32	0	2110.92	2110.88[dir]	(HexNAc)$_4$
			1661.61	1661.79[dir]	(HexNAc)$_1$ (Phos)$_2$
			2270.85	2270.98[dir]	(HexNAc)$_4$(Phos)$_2$
NESSSTTSTLSGKSIGR	20-36	1	1914.92	1915.11[dir]	(HexNAc)$_1$
			1994.89	1995.00[dir]	(HexNAc)$_1$ (Phos)$_1$
			2118.00	2118.05[dir]	(HexNAc)$_2$
			2807.20	2807.12[28]	(HexNAc)$_5$ (Phos)$_1$
			2631.153	*2631.094[dir]*	*(HexNAc)$_4$ (Phos)$_1$ (Methyl)$_1$*
SIGRSEQYISPDAEAFSK	33-50	1	2314.00	2314.07[>32]	MSO
			2224.86	2225.02[dir]	(Phos)$_3$
SIGRSEQYISPDAEAFSK YMLSK	33-55	2	3435.58	3435.54[dir]	MSO, (HexNAc)$_4$
			3638.66	3638.65[dir]	MSO, (HexNAc)$_5$
			3312.47	3312.49[dir]	MSO, (HexNAc)$_3$ (Phos)$_1$
			2826.35	2826.24[>32]	MSO (HexNAc)$_1$
			3462.674	*3462.611[dir]*	*(HexNAc)4 (Methyl)$_1$*
SEQYISPDAEAFSK	37-50	0	1571.72	1571.67[28/30/>32]	no modifications

			1651.68	1651.88[dir]	(Phos)$_1$
			1731.65	1731.60[dir]	(Phos)$_2$
SEQYISPDAEAFSKYMLSK	37-55	1	2493.07	2492.97[dir]	MSO (HexNAc)$_1$ (Phos)$_1$
YMLSK	51-55	0	860.40	860.59[32]	(HexNAc)$_1$
YMLSKSPEDIGPSDSA SNDPLTSFSIR	51-77	1	3320.53	3320.44[dir]	(HexNAc)$_2$
			3726.69	3726.79[dir]	(HexNAc)$_4$
			3929.77	3929.89[dir]	(HexNAc)$_5$
			3720.36	3720.56[dir]	(HexNAc)$_2$ (Phos)$_5$
YMLSKSPEDIGPSDSA SNDPLTSFSIRSNAVK	51-82	2	3429.64	3429.56[dir]	MSO
SPEDIGPSDSASNDPL TSFSIR	56-77	0	2292.06	2291.98[dir/26/30]	no modifications
			2372.02	2372.01[>32]	(Phos)$_1$
TNADAGVSMDSSTQSRPSSN VGCDQVDFSFNK	83-114	0	*3827.637*	*3827.825 [dir]*	*(HexNAc)$_2$ (Methyl)$_1$*
GISMNANLDSSISISTSSK	115-133	0	2151.83	2151.88[dir]	(Phos)$_3$
			2210.97	2211.01[>32]	MSO (HexNAc)$_1$ (Phos)$_1$
			2820.21	2820.22[dir]	MSO (HexNAc)$_4$ (Phos)$_1$
			2884.18	2884.25[dir]	(HexNAc)$_4$ (Phos)$_2$
			2738.101	*2738.208[dir]*	*(HexNAc)$_3$ (Phos)$_2$ (Acetyl)$_1$*
GISMNANLDSSISISTSSKK	115-134	1	2040.03	2040.12 [dir]	no modifications
			2056.02	2055,91 [dir]	MSO

			2809.19	2809.17[dir]	(HexNAc)$_3$ (Phos)$_2$
			2825.19	2825.25[28]	MSO (HexNAc)$_3$ (Phos)$_2$
			3354.46	3354.63 [30]	MSO, (Phos)$_6$
			3338.47	3338.57 [>32]	(HexNAc)$_6$ (Phos)$_1$
			2850.201	*2850.293[dir]*	*(HexNAc)$_3$ (Phos)$_2$ (Acetyl)$_1$*
GISMNANLDSSISISTSSKKEK	115-136	2	2297.16	2297.09[26]	no modifications
			2922.39	2922.38[>32]	MSO, (HexNAc)$_3$
GISMNANLDSSISISTSSKKEKSK	115-138	3	3703.61	3703.69[26]	MSO, (HexNAc)$_5$ (Phos)$_2$
KEKSK	134-138	2	822.45	822.45[28/>32]	(HexNAc)$_1$
KEKSKSDHK	134-142	3	*1113.614*	*1113.617[dir]*	*(HexNAc)1 (Acetyl)$_1$*
EKSKSDHK	135-142	2	1161.57	1161.53[26]	(HexNAc)$_1$
IEAESDSDDYILDDSDSD DGKCK	150-172	1	2535.09	2535.20[dir]	no modifications
IEAESDSDDYILDDSDSDDG KCKNCKYKR	150-178	4	3327.42	3327.57[dir]	no modifications

Tab. 4. Collection of peptide masses observed upon RF strain infections together with fitting modifications. Possible modification that were considered are phopshorylation and glycosilation, masses containing methylation and acetylation were conditionally included and are written in *italic*.

<u>SA11 strain.</u>

MSLSIDVTSLPSIPSTIYKNESSSTTSTLSGKSIGRSEQYISPDAEAFNKYMLSKSPEDIG
PSDSASNDPLTSFSIRSNAVKTNADAGVSMDSSAQSRPSSNVGCDQVDFSLNKGLKV
KANLDSSISISTDTKKEKSKQNHKSRKHYPRIEAESDSDDYVLDDSDSDDGKCKNCK
YKKKYFALRMRMKQVAMQLIEDL

Abbreviations:
MC- trypsin digestion missed cleavage
HexNAc- N-acetyloglucosamine,
Phos- phosphate
Acetyl- acetylation
Methyl- methylation
MSO- methionines oxidized
[dir]- direct trypsinization method

Sequence of the peptide	Peptide position	MC	Theoretical mass	Experimental mass	Modification
NESSSTTSTLSGKSIGR	20- 36	1	*1818,836*	*1818,77[dir]*	*(Phos)$_1$ (Methyl)$_2$*
SIGRSEQYISPDAEAFNK	33- 50	1	*2104,946*	*2104,925[dir]*	*(Phos)$_1$ (Acetyl)$_1$*
SEQYISPDAEAFNKYMLSK	37- 55	1	*2393,989*	*2393,954[dir]*	*(Phos)$_2$ (Acetyl)$_1$*
			2437,978	*2437,984[dir]*	*MSO (Phos)$_2$ (Acetyl)$_1$*
			2531,96	*2531,982[dir]*	*MSO (Phos)$_2$ (Acetyl)$_1$*
SPEDIGPSDSASNDPLTSFSIR	56- 77	0	2654,067	2654,039[dir]	(Phos)$_2$ (HexNAc)$_1$
			2828,015	2828,125[dir]	(Phos)$_2$ (HexNAc)$_1$ (Methyl)$_1$
EKSKQNHK	135- 142	2	*1214,636*	*1214,616[dir]*	*(HexNAc)$_1$ (Methyl)$_1$*
			1228,615	*1228,627[dir]*	*(HexNAc)$_1$ (Methyl)$_2$*

EKSKQNHKSR	135- 144	3	1240.663	1240.637[dir]	no modifications
			1268.695	*1268.661[dir]*	*(Methyl)₂*
			1457,758	*1457,824[dir]*	*(HexNAc)₁ (Methyl)₁*
SKQNHKSR	137-144	2	*1214,636*	*1214,616[dir]*	*(HexNAc)₁ (Methyl)₂*
			1228,652	*1228,627[dir]*	*(HexNAc)₁ (Methyl)₃*
SKQNHKSRK	137-145	3	*1219,618*	*1219,62[dir]*	*(Phos)₁ (Methyl)₂*
SRKHYPRIEAESDSDDYVLDDS DSDDGK	143-170	3	*3283,438*	*3283,277[dir]*	*(Acetyl)₁ (Methyl)₂*
			3443,371	*3443,299[dir]*	*(Phos)₂ (Acetyl)₁ (Methyl)₂*
KKYFALRMR	178-186	3	*1269,701*	*1269,667[dir]*	*MSO (Acetyl)₁*
KYFALRMRK	179-188	3	*1458,783*	*1458,834[dir]*	*MSO (Acetyl)₁ (Methyl)₃*
YFALRMRK	180-188	2	1214,641	1214,616[dir]	no modifications
			1228,657	*1228,627[dir]*	*(Methyl)₁*

Tab. 5. Collection of peptide masses observed upon SA11 strain infections together with fitting modifications. Possible modification that were considered are phopshorylation and glycosilation, masses containing methylation and acetylation were conditionally included and are written in *italic*.

Although observations coming from all the experiments allowed drawing several conclusions, one important thing must be mentioned. The data analyzed in the these studies derived from naturally infected cells, so represent the dynamic state of NSP5 modification level, that changes during the course of infection. Saying shortly- the same peptide may contain e.g. one, two or three sugar moieties as well as can be phosphorylated, depending on the time-point of infection. The analyses of masses let to identify fragments that belong to almost all parts of NSP5 sequence, either modified or without modifications. From these observations we deduced the possible range of modification of this important rotaviral protein. The schematic view on NSP5 sequence and its modifications depicted in figure 34, was drawn based on the analysis of peptides masses, that were observed in at least two different experiments performed by two different methods. In that scheme we showed possible modifications, that do not have to be present in the protein at the same time, however the tendency of particular regions to be particularly modified is demonstrated. All depicted modifications met the highest (red) stringency criteria, nevertheless the peptides containing methylation or acetylation were believed less certain and were shown in blue, as we do not have the confidence, that these modifications, in contrary to phosphorylation and glycosilation, occur in NSP5. The other data, that did not feet all the stringency conditions, because for instance occurred among the analyzed results with a lower frequency, can also be discussed and are shown in tables 3 and 4. Data from table 5, where results from a single experiment with SA11 strain were placed, are less reliable and must be confirmed with subsequent studies.

In summary, from all the information obtained, few remarks can be made. The peptides assigned to the region 1 were well represented. Among all the peptides, we found these, that feet perfectly the amino acids 1-19 of region 1 not modified, carrying one, two, five and six N-acetyloglucosamines and with one or two phosphorylation sites. Additionally there is a high possibility, that the N-terminal part is also methylated. Other masses reflected perfectly the masses of peptides encompassing amino acids 20-32, bearing exactly four sugars but also four sugars and one and two phosphates. The general view on the region 1 suggest high level of glycosilation of this part of NSP5.

In the region 2, two potential phosphorylation sites were assigned. The peptide fitting to the region 56-77 was found without any modifications, but also with one serine or threonine phosphorylated. There is a big probability, that phosphorylation occurred at serine 67. An interesting observation is that the phosphorylated peptide derived only from the upper bands of the gel and was absent (or not found) in the lower bands of fractionated NSP5.

Serine 67 is a key residues from which initiates all the phosphorylation mechanism of NSP5 and the absence of the phosphorylation in lower bands may signify, that the hyperphosphorylation of NSP5 is a very quick cascade process.

Moreover, the first part of region 3, spanning from the position 80 to 115 was not represented in most of the experiments, suggesting that this peptide was not read by the machine because of its low level of ionization that depends on the sequence. Another, less possible reason for that fact, is that this peptide is so highly modified that its mass was out of an efficient range of the spectrum reading. Adding methylation as a possible modification to the selection criteria, increased hits assigned to this NSP5 region, and indicated the possibly of occurrence of two sugar moieties and one methylated site there.

In contrast, sequence between aa 115 and 134, assigned also to region 3, was abundantly recognized in all the experiments, non modified and modified. This region of NSP5 sequence is different for both strains (OSU and RF). In the bovine RF strain, eight potential modification sites are present, and in its case three added sugar moieties and two phosphorylation sites were characterized. OSU strain contains only four serines and one threonine in the same peptide and we managed to map in these region one glycosylation and one phosphorylation site. We found this part of the protein also modified in more extended way, for example we found peptide with six N-acetyloglucosamines and one phosphate and also with six phosphates but this result can not be treated as secure, because was not found in all the experiments.
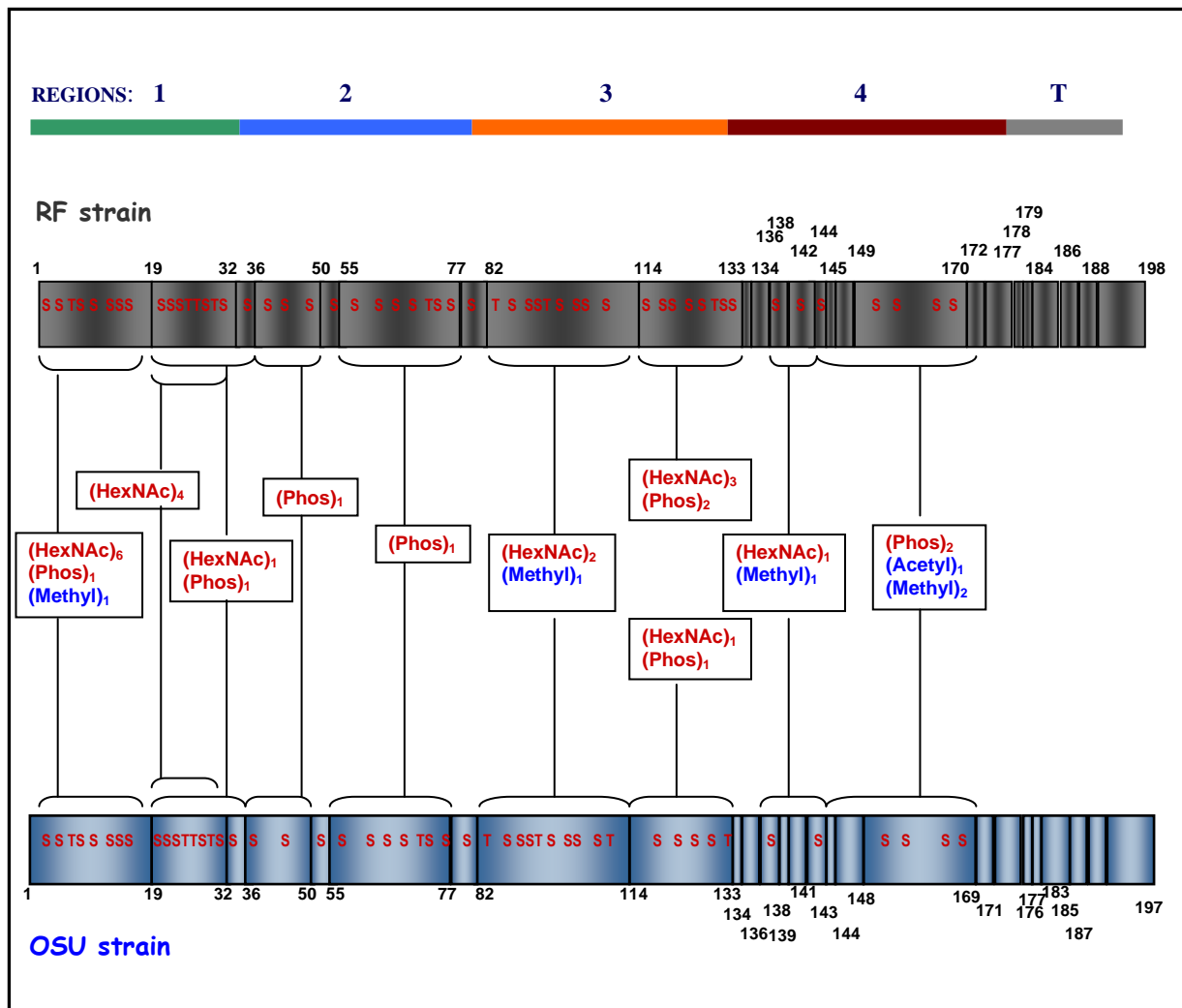
Fig. 34 Scheme showing the schematic representation of NSP5. The regions of the protein correspond to the bar above the scheme. Localisation of serines and threonines are depicted, the sites of trypsin cleavages are indicated by numbers. The modifications shown here were selected with the high stringency criteria, and these very probable are shown in red (phosphorylation and glycosilation), the ones that still require confirmation (methylation, acetylation) are marked in blue.

The beginning of region 4 revealed addition of one N-acetyloglucosamine in all the three strains. Moreover methylation of this part of the protein was deliberated as possible, because was found in all the experiments. Furthermore, in region 4, consistently to previous data obtained in our laboratory, suggesting that this part of NSP5 is phosphorylated, two phosphorylation sites were determined and perhaps some methylation and acetylation sites as well. Finally, all the peptides that form the end of the region 4 and the tail were completely reconstituted in both strains. This region of NSP5 was found as not modified and this is concordant with the fact that it lacks serines and threonines.

The method of in gel digestion (fig. 35) should allow to separate differently modified isoforms of NSP5 and help to understand how the protein is changed. Up to the previous studies [3], NSP5 derived from highest bands of SDS-PAGE gel should be hyperphosphorylated and rather not glycosilated. Although, the phosphorylation contributes to polypeptide charge and changes peptide's migration in the gel, the separation depends not only on the amount of phosphate added but also on the sequence where the phosphate moiety was attached. We found many phosphorylated peptides in the upper bands (including the one encompassing serine 67) and some in the lower. The fact that NSP5 isoforms visible in the lowest bands contain certain level of phosphorylation was known earlier and is concordant with our studies. However the presence of sugars in the upper bands is surprising, because is contradictory to the previous data obtained in our laboratory by Afrikanova and et al. We suspect, that this result may be explained with the fact, that in gel digestion method did not let us to collect enough material and to obtain more complex analysis next experiments should be made.

OSU:	37-50	(Phos)1
	56-77	(Phos)1
	115-134	(HexNAc)1(Phos)1
RF:	1-19	(HexNAc)6(Phos)1
	33-55	(HexNAc)1
	51-55	(HexNAc)1
	56-77	(Phos)1
	115-133	(HexNAc)1(Phos)1
	115-134	(HexNAc)6(Phos)1
	115-136	(HexNAc)3
	134-138	(HexNAc)1
	145-172	(Phos)2

OSU:	56-77	(Phos)1
RF:	1-19	(HexNAc)1
	56-77	no modifications
	115-134	(Phos)6
	134-138	(HexNAc)1

OSU:	20-36	(HexNAc)2(Phos)1
RF:	1-19	(HexNAc)1
	20-36	(HexNAc)5(Phos)1
	115-134	(HexNAc)3(Phos)2
	134-138	(HexNAc)1

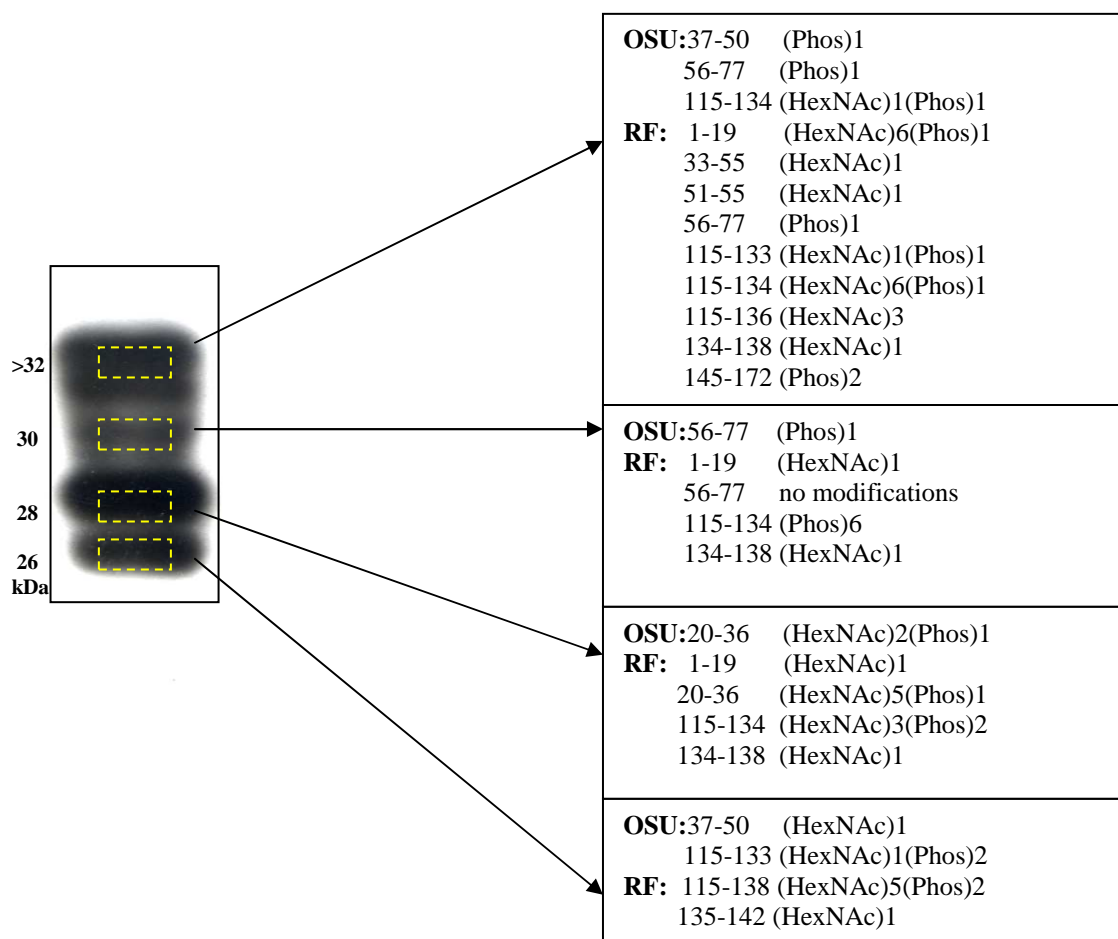|		
OSU:	37-50	(HexNAc)1
	115-133	(HexNAc)1(Phos)2
RF:	115-138	(HexNAc)5(Phos)2
	135-142	(HexNAc)1

Fig. 35. Western blotting of immunoprecipitated NSP5 derived from rotavirus infection of MA104 cells. Dashed squares represent way of protein fractionating. The scheme shows which peptide were found in which bands of separated by SDS-PAGE NSP5.

Part 3: Structural studies on NSP5

The aim of this part of the thesis was to express, purify, and analyze NSP5 protein. With our studies we wanted to add new data to very limited information about NSP5 structural organization and protein conformation. To achieve this, in addition to the full length protein, two different truncation mutants and one point mutation mutant were constructed, expressed and purified. Proteins secondary structure was analyzed by circular dichroism spectroscopy and the provided general information about protein conformation.

Expression of NSP5wt in *E.coli*

In order to facilitate the recombinant protein purification the cDNA encoding Nsp5 was cloned into a vector containing a tag of six histidines which is used in a metal-chelating affinity chromatography purification. First, for Nsp5 expression the pQE system (Qiagen) was chosen. High level of expression of his-tagged proteins in *E.coli* using pQE vectors is based on a powerful phage T5 promoter transcription-translation system. The promoter is recognized by *E.coli* RNA polymerase. There are two *lac* operator sequences which increase *lac* repressor binding and ensure efficient repression of the T5 promoter. The extremely high transcription level can be only efficiently regulated and repressed by the presence of high levels of *lac* repressor protein. *E.coli* host strain M15 [pRep4] utilized in that expression system uses a *lac* repressor gene in trans to the gene that is supposed to be expressed. The strain contains the low-copy plasmid pREP4 which posses kanamycin resistance and constitutively expresses the *lac* repressor, encoded by *lac I* gene. Expression of recombinant protein encoded by pQE vectors is rapidly induced by the addition of isopropyl-β-D-thiogalactoside (IPTG) which binds to the *lac* repressor protein and inactivates it. Once the lac repressor is inactivated the host cell's RNA polymerase can transcribe the sequence downstream from the promoter and production of the recombinant protein starts.

In spite of the relatively small size of the 6xHis-tag, it has been reported to affect the activity of some purified recombinant proteins. Considering that the C-terminus of NSP5 is necessary for the protein dimerization [161] [149], we decided that the tag should rather be

attached to the N-terminus of the recombinant protein, in particular because the protein was designated for structural studies.

The vector pQE30 codes for a polyhistidine tag, which is localized before a multicloning site. Therefore bacteria, transformed with the plasmid carrying a foreign gene, produce a recombinant protein that has a tag situated at its N-terminus. The pQE30-NSP5wt construct was prepared (Fig. 36A) [6] and viral protein was efficiently expressed in M15 [pRep4] strain of *E. coli* (Fig. 36B).
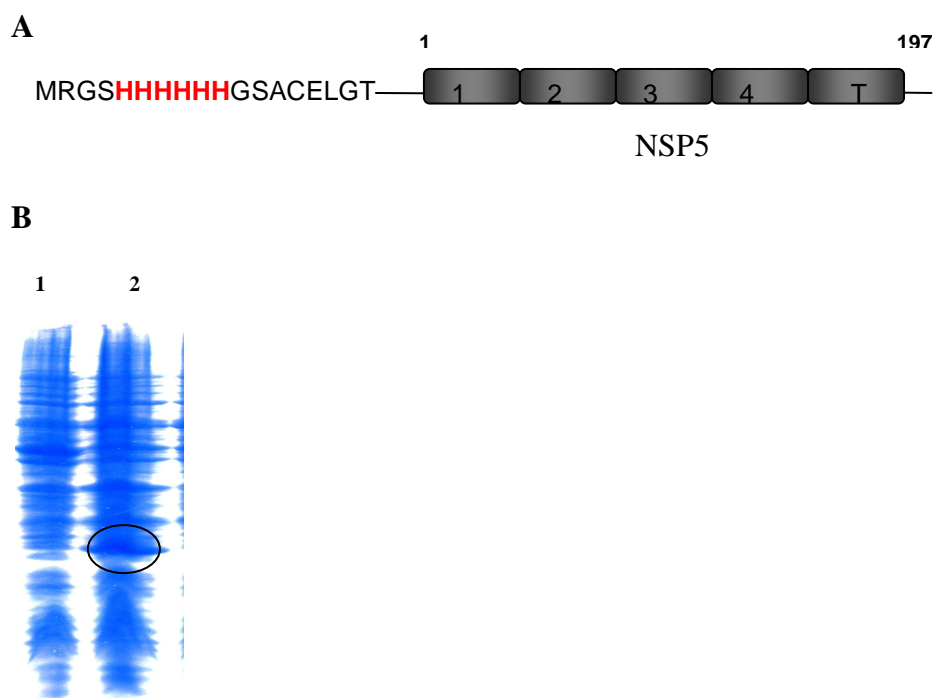
Fig. 36. Expression of His$_6$NSP5-pQE30 in *E.coli* M15 [pRep4]. (A) A schematic representation of NSP5 in pQ30 plasmid. (B) SDS-PAGE analysis of protein production. Line 1 represents protein of M15[Rep4] carrying NSP5 gene, uninduced control; line 2- the same bacteria induced with 1mM IPTG; circle indicate NSP5. Proteins were analyzed in 12% SDS-PAGE with Coomassie blue staining.

Unfortunately practically all the recombinant NSP5 was localized in the non soluble fraction of the cells, accumulating in aggregates called inclusion bodies. In these aggregates only a small fraction of the protein is folded correctly, and the rest is in a denatured and usually nonfunctional form [247] [248].

The maximized yield of soluble protein can be achieved by optimizing several factors such as the expression temperature, *E. coli* strain used, pH of the growth medium, a folding

partner the protein is fused with, rate and expression level and co-expression with chaperones and foldases [249] [250]. To increase soluble fraction of NSP5 we tested different conditions of bacteria cultivation that included growing in a lower temperature (25°C) and slowing down the rate of production by inducing a recombinant protein production with a very low concentration of IPTG. Also, another strain of *E. coli*, BL21 [DE3], was tested. To express NSP5 in this strain the pET plasmids system (Novagene) was used.

The pET system is driven by the promoter of the bacteriophage T7 . It was pointed out that T7 RNA polymerase is very specific for T7 promoters and it does not recognize DNA from other sources, since these promoter sequences are very rare. Also, termination signals for T7 RNA polymerase are rare too, therefore long transcripts can be made without premature truncation. The T7 RNA polymerase is about five times faster than *E.coli* RNA polymerase, so genes controlled by T7 promoters can be overexpressed. In order for this system to work, the T7 RNA polymerase must be supplied to the cells. A special strain called BL21 [DE3] has been made for this purpose. The BL21 [DE3] cells are lysogenic for a fragment of the phage DE3. This fragment contains the *lacI* gene, the *lac* UV5 promoter, the start of *lacZ* (beta galactosidase) and the T7 RNA polymerase gene. The lac UV5 promoter, inducible by IPTG, is responsible for driving the expression of T7 RNA polymerase. This fragment is inserted into the *E.coli* chromosome. In addition to the T7 RNA polymerase, there is a vector that carries the cloned gene behind a T7 promoter, and termination signals are downstream of the gene to stop the transcription (pET vector). In theory, the gene can only be made if IPTG is added to activate the lacUV5 promoter and turn on synthesis of T7 RNA polymerase. This in turn transcribes the gene in the pET vector. In practice though, the *E. coli* RNA polymerase does make a small amount of T7 RNA polymerase without induction. It may bind upstream of the *lac* UV5 promoter and read through it. For that reason it is said to be a 'leaky' system. This usually causes a small amount of the cloned gene to be expressed without IPTG induction.

In this study the pET23d (+) plasmid was used. As in the case of previous system, the construct was designated in the way, that polyhistidine tag was localized on the N-terminus of NSP5 (Fig. 38A). In this system control of inducible expression in case of NSP5 fails and the protein was produced with or without the presence of IPTG. Although it was very efficiently overexpressed, practically all the produced protein, as in case of the pQE/M15 [Rep4] system, formed aggregates.

Very low expression (nanograms per liter of bacterial culture) of soluble NSP5 was only obtained with the M15 [Rep4] strain. The purity degree of NSP5 after metal affinity

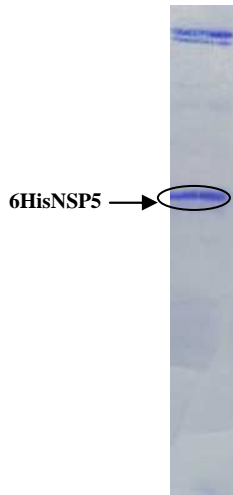chromatography was not sufficient (Fig. 37) and it had to be subjected to subsequent purification steps.

6HisNSP5 →

Fig. 37. His$_6$NSP5 expressed in M15 [Rep4] and purified by metal affinity chromatography in nondenaturing conditions

One of possible solutions was to change the expression system. We tested two different eukaryotic expression systems, recombinant vaccinia virus expression system and Sinbis virus replicon-based DNA expression system. The results of experiments were not satisfactory, since these systems did not ensure appropriate expression level of NSP5.

It was then decided to produce the protein in *E. coli* and to recover it from inclusion bodies. Protein derived from inclusion bodies can be refolded *in vitro* to give a functional protein. Sometimes the formation of inclusion bodies can be even advantageous, since the expressed level of desired protein is very high and protein can be purified simply by washing and pelleting the aggregated bodies. Because of the convenient cloning sites, short linker between polyhisidine tag and the viral gene and a good expression level the pET/BL21 system was chosen (Fig. 38).
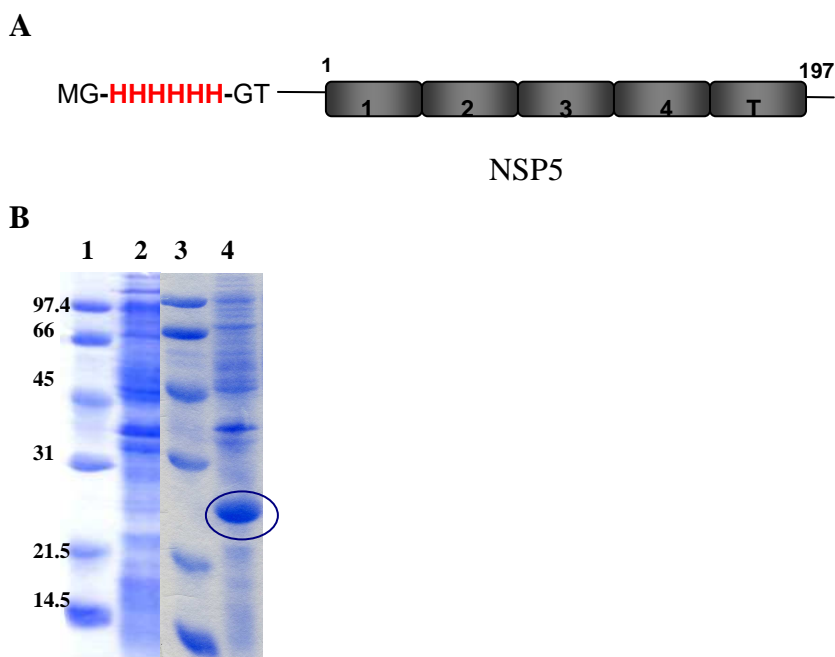
A

MG-**HHHHHH**-GT

1   197

1   2   3   4   T

NSP5

B

Fig. 38. Expression of His$_6$NSP5-wt in *E. coli* BL21 [DE3]. (A) A scheme of His$_6$NSP-wt construct in pET23d(+) vector. (B) Expression of His$_6$NSP5-wt in BL21[DE3]. Lines left to right: marker; BL21 [DE3] not transformed; marker; BL21 [DE3] transformed with pET23d (+) -His$_6$NSP5-wt.

NSP5 wt purification and refolding.

In the first attempts two methods for the purification and refolding of 6His-tagged NSP5 protein were used. Both of them were based on solubilizing inclusion bodies and in-column refolding of the bound protein that was performed by buffer exchange, using a linear gradient from a buffer containing 8 M urea to a non-denaturing buffer (0 M urea). Finally, bound, renaturated proteins are eluted with a gradient of imidazole. In case of some protein those protocols work and functional proteins can be obtained. [251] [252] [253]. Unfortunately in the case of NSP5 problems of heavy precipitation during refolding occurred, and in addition, the purity of the product was too low. However protein obtained in this way served to guinea pigs immunizations and anti-NSP5 serum generation.

In this situation we decided to go through all the purification steps in denaturing conditions and to refold the protein at the end by the dialysis method. The protocol for inclusion bodies purification, detailed purification steps and renaturation procedures are

described in "Material and methods" chapter. Shortly, the cocktail of denatured proteins was passed through affinity chromatography column (His-Trap, Amersham Pharmacia). Immobilized metal affinity chromatography (IMAC) systems have three basic components: an electron donor group, a solid support and a metal ion. The metal ion (usually Ni^{2+}, Co^{2+}, Cu^{2+} or Zn^{2+}) is restrained in a coordination complex where it still retains significant affinity towards macromolecules [254]. To purify NSP5 we used His-Trap affinity column of Amersham Pharmacia that utilizes Ni^{2+} metal ion. Protein was eluted using imidazole gradient in the presence of urea (Fig. 39A), fractions of protein were checked on SDS-PAGE gel and those containing NSP5 (Fig. 39B) were pooled and subjected to the second purification step- ion exchange chromatography.
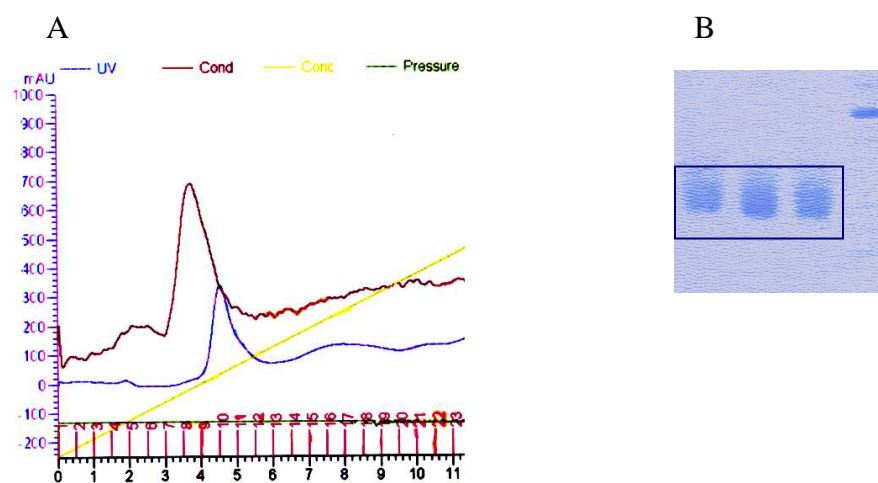
Fig. 39. Affinity chromatography purification of His$_6$-NSP5 wt. (A) elution profile and (B) SDS-PAGE gel demonstrating fractions containing recombinant protein (marked with square).

The ion exchangers are insoluble solid matrices containing fixed ionogenic groups which bind reversibly to sample molecules (proteins, etc.). Desorption is then brought about by increasing the salt concentration or by altering the pH of the mobile phase. The two major classes of ion exchangers are cation exchangers and anion exchangers, having negatively and positively charged functional groups, respectively. In case of His$_6$-NSP5, since the isoelectric point, (pI), is about 6.4, an anion exchange chromatography was applied. The protein was eluted with increasing NaCl gradient (Fig. 40A), a peak of pure recombinant His$_6$-NSP5wt was collected and sample was checked on SDS-PAGE.

The refolding step is complex and depends greatly on the renaturing conditions, such as, pH, rate of dialysis, protein concentration and redox conditions. The protein was refolded by dialysis, according to the protocol described in the chapter [2.8]. Because glycerol stabilizes proteins and often facilitates renaturation allowing the return to native secondary structure as well as preventing aggregation, it was added to the refolding buffer.

NSP5 has four cysteins but as it is a cytoplasmic protein, disulfide bonds should not be formed. For this reason a reducing agent (DTT) was present during refolding process. Cysteins in NSP5, when present in an oxidizing environment, were found to be very reactive. It is noticeable that disulfide bridges can be formed during polyacrylamide gel running even though the sample buffer contained a reducing agent like β-mercaptoethanol or DTT, what result in dimer formation (Fig. 40B and C- arrow). Including a reducing agent into a SDS-PAGE running buffer eliminates this phenomenon.
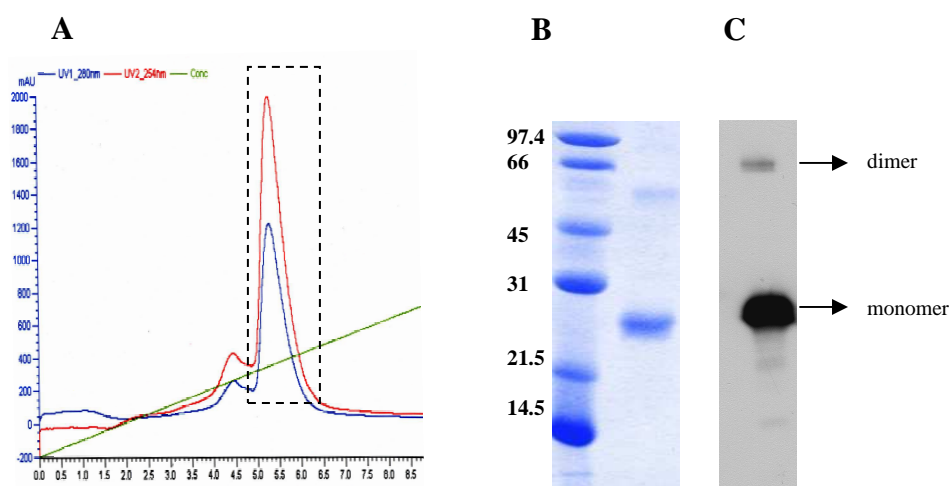
Fig. 40. Second step of His$_6$-NSP5wt purification and final product after renaturation. (A) Protein was eluted from anion exchange monoQ HR5/5 column with an increasing gradient of NaCl. The peak containing recombinant NSP5 was collected (dashed square). (B) Final product of purification: SDS-PAGE (12 %) gel, reducing conditions, stained with Coomassie blue and Western blotting anti-NSP5 (C). To bands visible- lower, representing a monomer, and a dimer formed by hard to reduce disulfide bridges.

The yield and homogeneity of purified refolded His6-NSP5wt protein was high. The final product was checked by SDS-PAGE, Western blotting (Fig. 40B and C) and its correct molecular mass (small differences may occur due to a calibration of the system) was confirmed by ESI-MS single quadrupole mass spectrometry (Fig. 41).
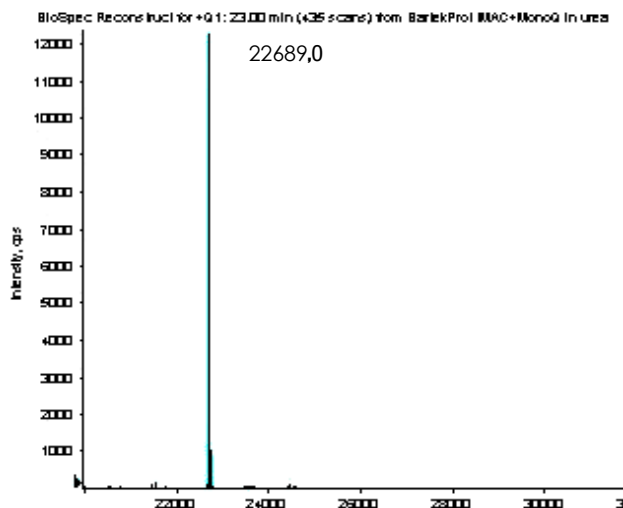
Fig. 41. ESI-MS single quadrupole mass spectrometry result of purified His₆NSP5-wt; mass calculated- 22693.0 Da, mass obtained- 22689.0 Da

ELISA-based binding assay.

Sine NSP5 is not an enzyme it was hard to say if the renaturated protein was well folded and active. In order to check its biological activity, a binding assay to NSP2, a well established partner of NSP5, was designed. The plate was coated with recombinant NSP5 and cellular extract containing NSP2 was then added. Significant binding signal, decrease of signal in wells with serial dilutions of NSP2 extract as well as a big difference between controls was noticed (Fig. 42).

NSP5-NSP2 ELISA binding assay
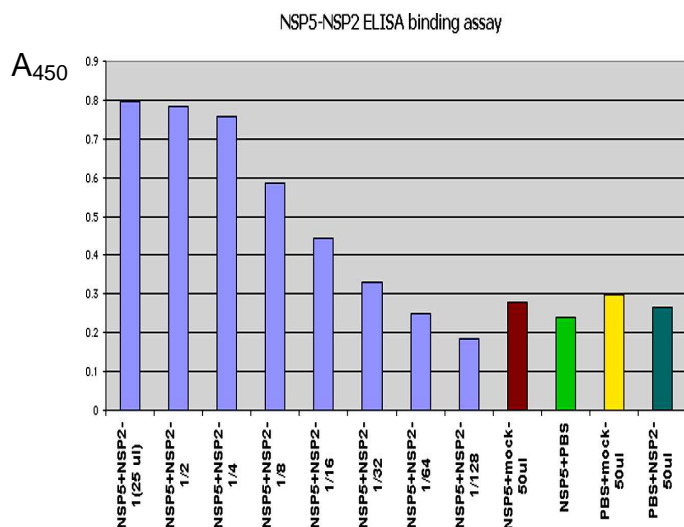
Fig. 42. Schematic representation of NSP5-NSP2 ELISA binding assay result.

This result showed that the refolded protein retained its ability to interact with NSP2, suggesting a biologically active form.

Trypsin limited proteolysis assay.

Although the full-length NSP5 was produced and a satisfactory protein purity was achieved, significant precipitation was observed and hence such preparation was not suitable for more advances structural studies.

Limited proteolysis involves the incomplete digestion of a protein by a protease, which can be examined as a function of time or ratio of protease to protein concentration. The partial digestion of a protein defines those fragments resistant to proteolysis and, therefore, are expected to have a more compact structure. In contrast, flexible and exposed to solvent regions, are prone to be digested. Information from these experiments can aid the design of constructs for structural biology studies.

Purified His_6-NSP5wt was subjected to limited proteolysis by trypsin as described under "Materials and Methods". With the time course of the reaction a peptide with an apparent molecular mass in PAGE lower than 20 kDa remained, indicating it was more resistant to trypsin proteolysis than the full length protein (Fig. 43A). That band as well as the full-length protein band were isolated from the gel, subjected to complete trypsin digestion and peptide's masses were estimated by mass spectrometry. After analysis it was clear that the

trypsin-resistant part of NSP5 initiated at position 32, exactly at the beginning of region 2 and encompasses all the protein length up to the C-terminus (Fig. 43B).

A   **B**

MGHHHHHHHGTMSLSIDVTSLPSISSSIFK**NES SSTTSTLSGK**SIGRSEQYISPDAEAFNKYMLS KSPEDIGPSDSASNDPLTSFSIRSNAVKTNADA GVSMDSSTQSRPSSNVGCDQVDFSLTKGINVS ANLDSCVSISTDNKKEKSKKDKSRKHYPRIEA DSDEDYVLDDSDSDDGKCKNCKYKKKYFAL RMRMKQVAMQLIEDL
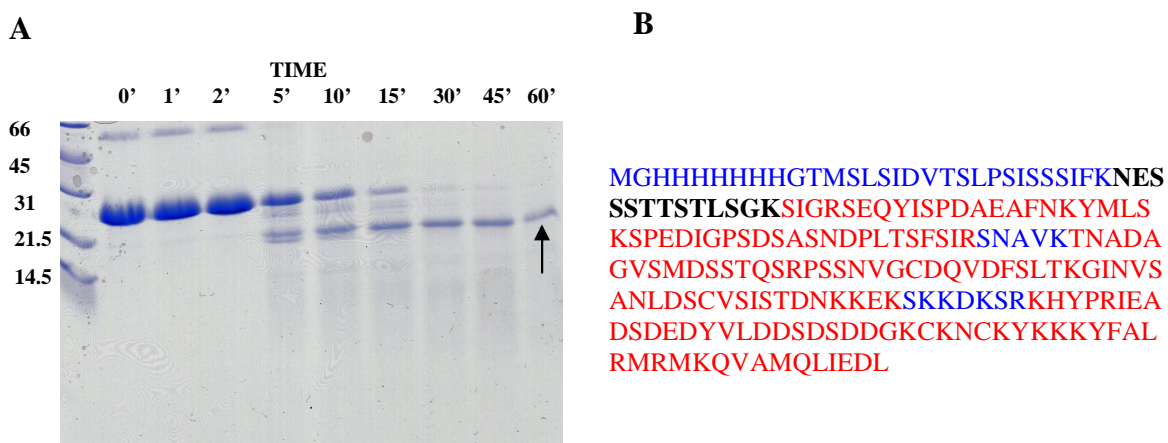
Fig 43. (A)Time course of trypsinolysis of His₆-NSP5wt, arrow shows a trypsin-resistant band; (B) Representation of a mass spectrometry analysis of the full-length and trypsin-resistant fragment. Sequence in blue represents regions that were not found in neither full-length nor trypsin resistant part of NSP5; in red – peptides found in both proteins; in black- fragments recognized only in a full-length protein analysis.

Expression, purification and refolding of NSP5 mutants.

To study structure of NSP5 three different mutants were prepared (the schemes of the constructs are shown in Fig. 45A). The N-terminus truncated mutant, His₆-NSP5ΔN32, in which the region one was deleted, was constructed on the basis of observations derived from trypsin limited proteolysis experiment. The protein with truncated N-terminal part was much more resistant to trypsin than the full length protein, suggesting that it might behave better in the solution, showing lower levels of precipitation and being more suitable for structural studies. In addition, taking into consideration the secondary structure predictions, a mutant deleted on the last 19aa was also constructed (His₆-NSP5 ΔC19). Secondary structure prediction was obtained running Jpred (http://www.compbio.dundee.ac.uk/~www-jpred/) - a

consensus method available on Expasy web server. All the programs concordantly predict the presence of around 20 aa long α-helix in the C-terminal region of NSP5 (Fig. 44). Although it was demonstrated, that this region is necessary for the protein dimerization [161] [149], the recombinant protein without the C-terminal tail may still fold into a soluble version of NSP5, indicating many interesting structural features.

A third mutant (S67D) was also constructed. This mutant was relevant because of its previous characterization in the context of NSP5 hyperphosphorylation mechanism (see page 77). We wanted to check if that feature affects the secondary structure content and reflects in change of the structural organization in comparison to the wild type protein.

Sequence:

MSLSIDVTSLPSISSSIFKNESSSTTSTLSGKSIGRSEQYISPDAEAFNKYMLSKSPEDIGPSDSASNDPLTSFSIRSNAVKTNADAGVSMDSSTQSRPSSNVGCDQVDFSLTKGINVSANLDSCVSISTDNKKEKSKKDKSRKHYPRIEADSDSEDYVLDDSDSDDGKCKNCKYKKKYFAL
RMRMKQVAMQLIEDL

Methods:

```
jalign   : -----E---------EEE-------EE--------------HHHHHHHH------------------EEE---EEEE----EEE---------------EEE------EE-----EEEE-----------------EE-----EEEE---------HHHHHHHHHHHHHHHHHHHHHHHHH---- : jalign
jfreq    : -------------EEEE-------EEE-------E----HHHHHHHHH-----------------EEEEE--HHE--------------------EEEEEEE--EEEEE----EEEEE------------------EE------EEE---------HHHHHHHHHHHHHHHHHHHHHHHHH--- : jfreq
jhmm     : -----EE---------EEE-------EE---------E----HHHH------------------------E----------EEEE----------------EEEEE----EE-----EEEE------------------EEE----EEEEE-------------HHHHHHHHHHHHHHHHHH---- : jhmm
jnet     : --EEEEE-------EEEE----------------EEE---HHHHHHHHH------------------EEEE--EE------EE--------------EEEEEE--EEEEE----EEEEE----------------EEE------EEE----------HHHHHHHHHHHHHHHHHHHHHHHHH--- : jnet
jpssm    : --EEE-------------------------------------HHHH------------------------EE------------------------EEE------EE-----EEEEE------------------EE--------------------HHHHHHHHHHHHHHHHHHH--- : jpssm

jpred    : -----E---------EEE--------E----------------HHHHHHHH--------------------EEE----EE------EE--------------EEEEE----EEE-----EEEEE-----------------EEE------EEE----------HHHHHHHHHHHHHHHHHHHHHHHHH--- : jpred
```

Jnet Rel :6745422146888820342888878820177667887132369023317772579999989999999998805163453023897331136787789998898653512531532646255006996457642224899999877736127997524634998888876114678418999999999998979399

Code for alignment:

Shades of red	- high level of conservation of chemical properties of the amino acids
jalign	- Jnet alignment prediction
jfreq	- Jnet PSIBLAST frequency profile prediction
jhmm	- Jnet hmm profile prediction
jnet	- Jnet prediction
jpssm	- Jnet PSIBLAST pssm profile prediction
jpred	- Consensus prediction over all methods
Jnet Rel	- Jnet prediction of prediction accuracy, ranges from 0 to 9, bigger is better.
H	- alpha helix
E	- beta sheet

Fig. 44. NSP5 secondary structure predictions by JPRED.

DNA constructs were generated (Fig. 45A) and the proteins were expressed, purified and refolded using exactly the same protocol as in case of wild type protein. Expression level in BL21 [DE3] strain was high (Fig. 45B) and the produced proteins also formed inclusion bodies. The purified proteins were refolded as mentioned before and final products were analyzed by SDS-PAGE and Western blotting (Fig. 46). As in the case of wild type protein, during the run of the samples in the gel (despite of the presence of reducing agent in the loading buffer), disulfide bridges are formed and the dimmer formation can be seen. Addition of DTT to the running buffer inhibits this process (Fig. 46, line 2).
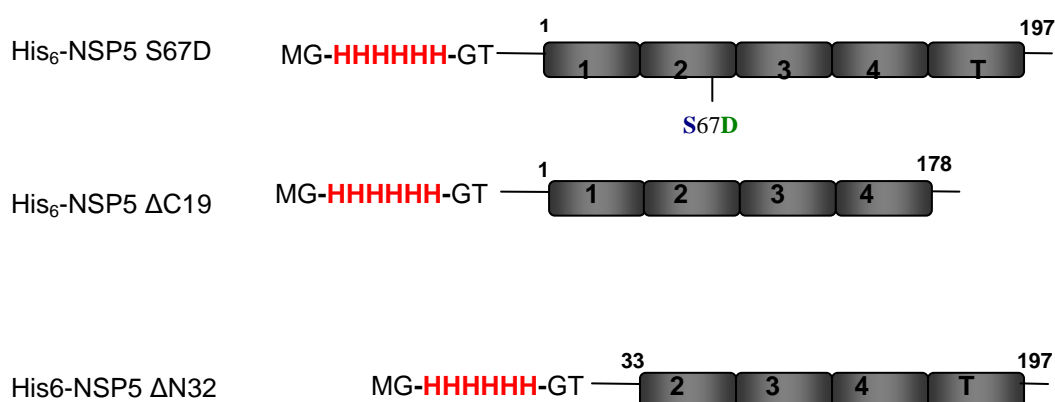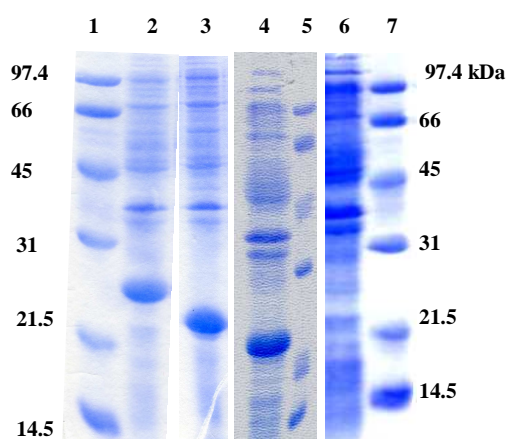
Fig. 45. Schemes of NSP5 constructs and proteins overexpression in *E.coli* BL21 DE3 strain. (A) Schematic representation of NSP5 mutants constructs cloned into pET23d (+) vector. (B) SDS-PAGE analysis of proteins production; lines 1, 5 and 7- marker; line 6- control, BL21DE3 non transformed; other lines - BL21DE3 transformed with pET23d(+)-His$_6$-NSP5S67D (line 2), His$_6$-NSP5ΔN32 (line 3) and His$_6$-NSP5ΔC19 (line 4). The proteins were analyzed in 12% SDS-PAGE gel stained with Coomassie blue dye.

Fig. 46. Analysis of purified NSP5 mutants. Proteins were analyzed by, respectively, SDS-PAGE Coomassie blue staining and anti-NSP5 Western blotting. Lines 1 and 5- His$_6$-NSP5 S67D; 2 and 4, His$_6$-NSP5 Δ19; 3 and 5- His$_6$-NSP5 ΔN32; In all the lines, both of Coomassie and WB, the artificially formed dimers are noticeable, except line 2, where DTT was present in loading and running buffer.

Gel filtration chromatography of purified NSP5 wt and mutants

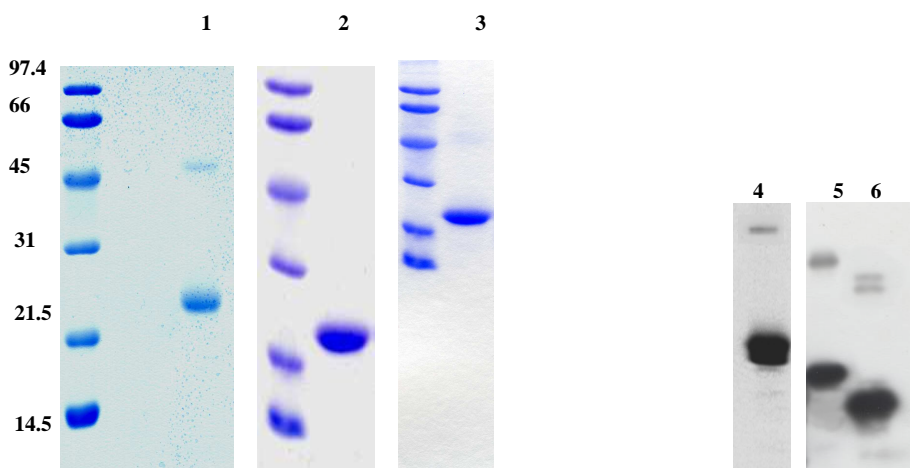The principle of size exclusion chromatography (gel filtration chromatography) is based on molecular volumes. The small molecules which diffuse into the gel beads (the porous three dimensional matrix that acts as a steric barrier) are delayed in their passage down the column compared to the large molecules which cannot diffuse into the gel and move continuously down the column with the eluent. Therefore the large molecules leave the column first followed by the smaller molecules in the order of their size. The gel filtration chromatography can be used to estimate approximate molecular weight of purified proteins in native condition, in contrary to SDS-PAGE. It can also give important information about protein like precipitation or aggregation, dimer formation, etc., that are crucial for advanced structural studies like crystallography.

Size exclusion chromatography of NSP5wt and mutants was performed using Superdex 200 column (Amersham Pharmacia), calibrated with a set of standards prior to the experiment (Fig. 47B). Unfortunately all the proteins were eluted in the void volume, confirming our presumption about high aggregation level. The only protein that behaved differently, due to its heterogeneous migration, and retained partially its size was the NSP5 ΔN32 mutant. Although the main portion of the protein was eluted in the void volume as

aggregated or precipitated protein, there was also some fraction that reflected the size of a dimeric NSP5 mutant. The collected fractions were subjected to Western blot analysis. The peak that left the column with the void volume (Fig. 47A; fractions 13 and 14) was recognized as NSP5, but in the 'hill' around 14 ml of the elution volume NSP5 was also found (Fig. 47A, C), suggesting that some part of the refolded protein did not aggregate and managed to form a dimer, maintaining its proper structure. The fact that the peak is not sharp might be explained by certain level of protein degradation, that could occur during refolding. In fact, the WB show that the bands of fractions 27, 28 and 29 are not perfectly homogenous. The possible explanation for that phenomenon is, that the dimer formation may require a shorter NSP5 construct, and some level of protein degradation (truncation from N- or C-terminus) enabled its formation, suggesting that for structural studies purpose even a shorter NSP5 mutant is required.

A

B

	LOGMW	El. Vol
Ferritine	440	2.643453	9.21
Catalase	232	2.365488	10.98
Aldolase	158	2.198657	11.17
Albumin	67	1.826075	12.68
Ovalbumin	43	1.633468	13.62
ChymotripA	25	1.39794	15.52
Ribonuc A	13.7	1.136721	16.21

C

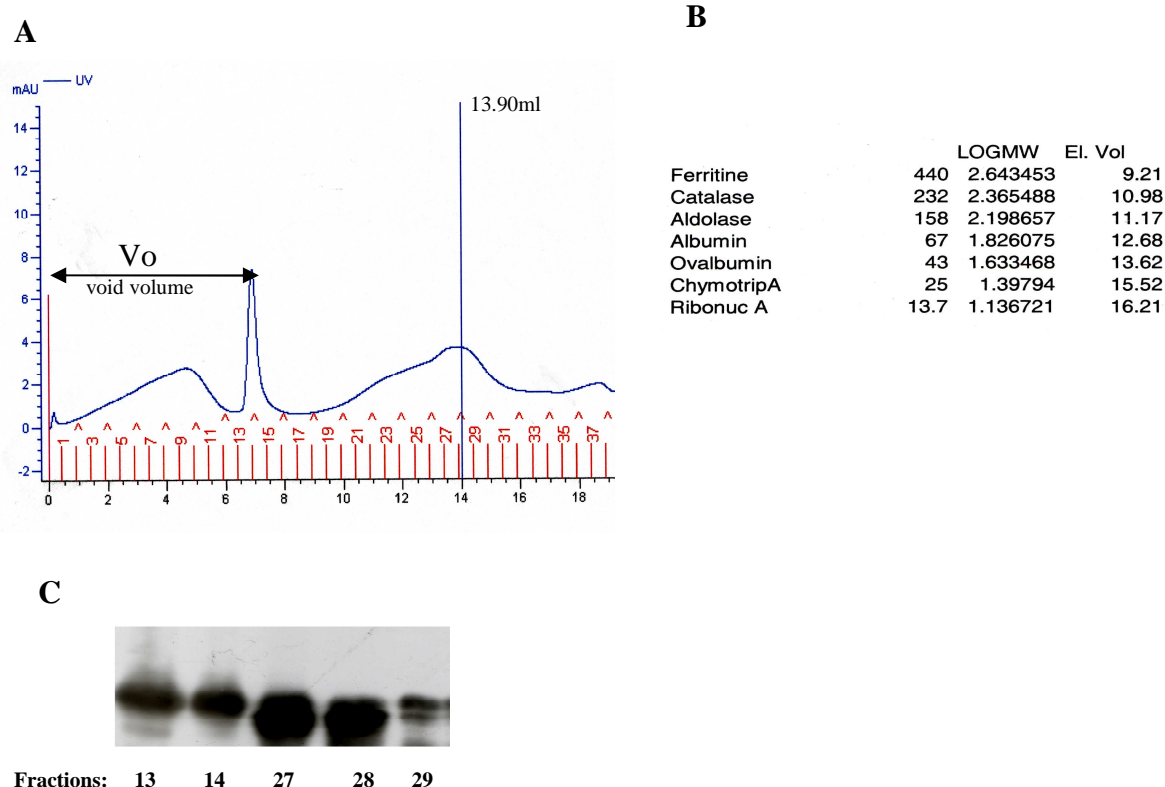Fractions: 13   14   27   28   29

Fig. 47. Size exclusion chromatography of His₆-NSP5 ΔN32. (A) elution profile, two peaks visible, one within Vo and one around 14ml of elution volume.   (B) calibration standards of Superdex S200 column; their molecular weights in kDa (first column) and elution volume (third column) are listed. (C) Western blotting (anti-NSP5) analysis of collected fractions.

Circular dichroism far-UV experiments

The purified NSP5wt and mutants were analyzed by circular dichroism (CD) spectroscopy. CD spectroscopy is an important method, complementary to crystallography in studies of protein structure, stability and folding. The sensitivity of far-UV allows estimating the content of secondary structure elements present in the protein and also assess if the protein is folded, which is especially an important issue in the case of refolded proteins.

For all analyzed proteins the percentage of secondary structure was estimated through deconvolution of the CD spectra in the range 190-240 nm using several methods available on the Dichroweb site. The best fittings between the experimental and calculated spectra were obtained using ContinLL and CDSSTR tools. The references set in these methods consist of spectra of proteins of known structure and includes also spectra of unfolded proteins [255].

NSP5wt

The spectrum of wild type NSP5 reflects good fitting when analyzed with ContinLL tool. Calculated contents of secondary structure elements for all tested proteins are collected in table 5. The spectrum reflects characteristics of an α-helical component profile (Fig. 48). The estimated percentage of α-helix (9.8%) is concordant with secondary structure predictions data. Quite high content of β-strands was also found. However accuracy for assessment of protein secondary structure by CD is more accurate for α-helical elements. Noticeable high participation of unordered elements that reaches almost 50%, could indicate a non- globular conformation of NSP5.
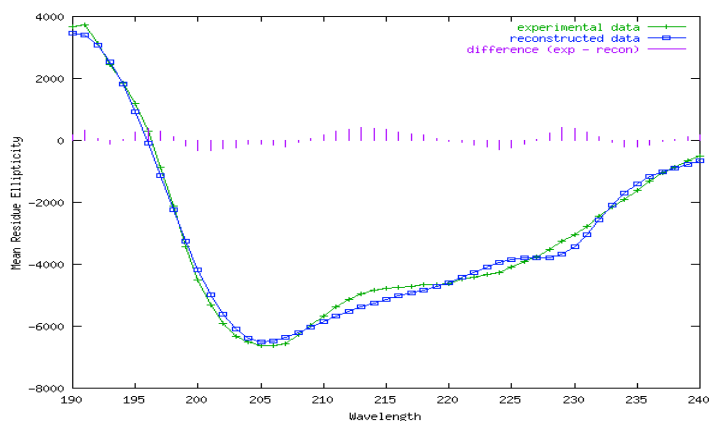
Fig. 48. Far-UV NSP5wt spectra. Experimental data (green) and reconstructed by ContinLL (blue). Differences in fitting marked in pink.

NSP5 S67D

We constructed, produced and purified S67D mutant, because it revealed to be more eager to become hyperphosporylated than the wild type protein. It was possible that introducing in that position a negatively charged residue, aspartic acid, changes protein conformation and affect content of secondary structure elements. However, the data obtained do not confirm this hypothesis, as the spectrum for the mutated protein does not significantly differ from the wild type protein (Fig. 48, 49).

Fig. 49. NSP5 S67D spectra of experimental data reconstructed by ContinLL data.

To make these result even more clear, spectra of both proteins were superimposed (Fig. 50). The differences in an intensity (not in spectral shape) might be affected by protein

concentration change due to precipitation prior to the experiment. These observations are in agreement with secondary structure content calculations tha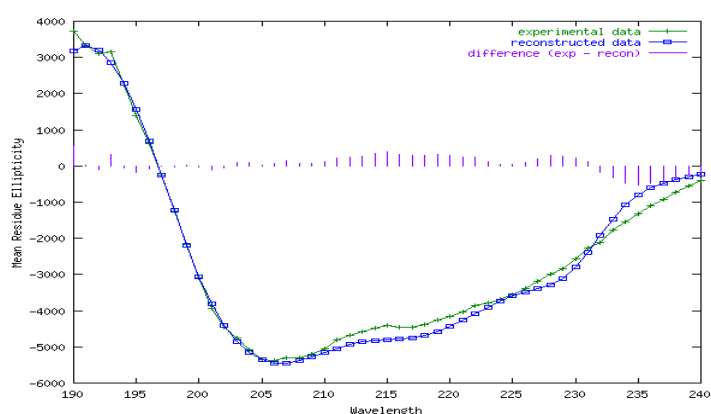t are practically identical. The small differences could derive rather from not identical experimental and reconstructed data coverage. This mean that the wild type protein an mutant have the same content of secondary structure elements and the aspartic acid introduced in the position of serine 67 did not change it.
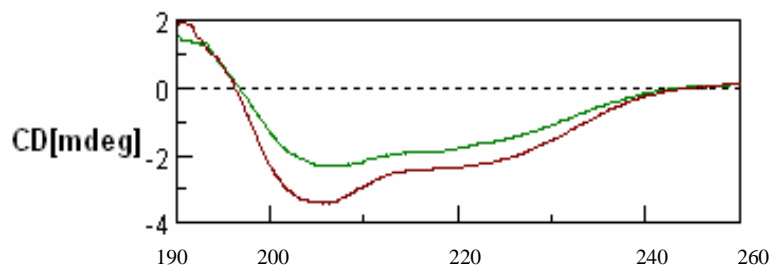
Fig. 50. Superimposition of far-UV CD spectra of NSP5wt and S67D mutant.

NSP5 ΔN32

The spectrum profile of ΔN32 mutant was also very similar to the one of wild type protein (Fig. 48, 50). This could be interpreted with the absence of organized structural elements in N-terminal part of NSP5, consistent with the limited proteolysis data. The shape of the spectrum is more α-helical than in the case of wild type protein since it has a more noticeable shoulder peak at about 222 nm, a typical pattern of α-helical conformation

The higher percentage of estimated α-helix content can take its origin from the fact, that helices localized in another part of the protein contribute more to the truncated protein as well as from a better fitting of reconstructed and experimental data. It is highly probable that the N-terminal part of NSP5 is mostly disordered and it appears not to be necessary for proper NSP5 folding.

119
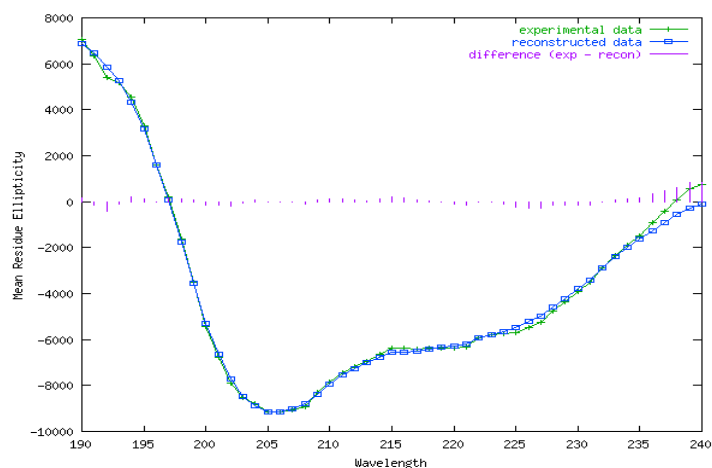
Fig. 50. NSP5 ΔN32 spectra of experimental data and fitting reconstruction calculated by CDSSTR tool.

NSP5 ΔC19

In contrast, removing of the last 19aa from the C-terminal part of NSP5 drastically changed the spectrum profile (Fig. 51). In His_6-NSP5 ΔC19 spectrum it is noticeable, a minimum at around 200nm characteristic for disordered proteins. Estimations computed by Contin LL tool (3.5% of α-helical component) had poor fitting and high NRMSD value. This parameter is an important measure of the correspondence between the experimental and calculated spectra and can be used to judge the quality of the results [256]. To confirm these findings spectra were also analyzed with another tools, including Yang.jwr software (0% of α-helix), that confirmed complete lack of α-helical structure elements that are present in the full length protein and NSP5 ΔN32 mutant. However other elements (β-sheets and turns) are maintained in calculations on the similar level than in other proteins, what can prove that C-terminal deletion mutant is folded. These observations, together with secondary structure predictions, lead us to speculate that practically all α-helical conformation motifs are present in C-terminal part of NSP5.
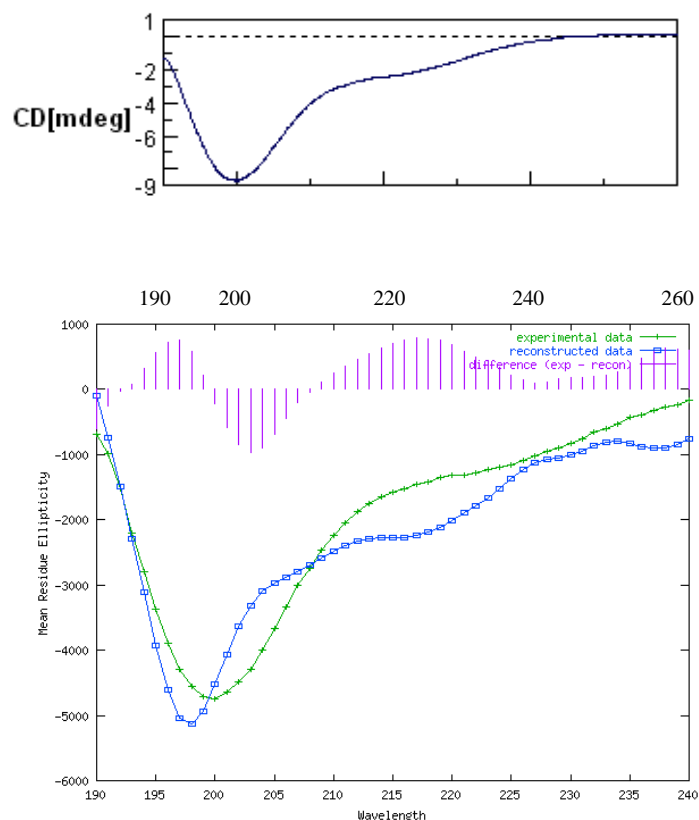
Fig. 51. NSP5 ΔC19 spectra. Experimental data in machine θ units (mdeg) and converted to MRE units data, reconstruction adjusted by CDSSTR tool.

In the table 6 the summary of all CD experiments is shown. The most significant observation is, that remove of C-terminus from the wild type protein (NSP5ΔC19 mutant) is reflected by a huge drop of total α-helices percentage, indicating, that most of α-helical elements are situated in the carboxy-terminal part of NSP5. This is concordant with the secondary structure predictions (Fig. 44) and confirmed by previous observations, that this region of the protein is involved in NSP5 dimerization. This result may suggest, that the dimerization may occur by formation of coiled-coil structure between two α-helices.

Protein	NRMDS	α-helices (%)	β-strands (%)	Turns (%)	Unordered(%)
NSP5wt	0.055	9.7	32	14	45
NSP5 S67D	0.069	9.0	29.9	17.4	44
NSP5ΔN32	0.042	18	22	16	43
NSP5ΔC19	0.213	3.5	33.2	16	48

Tab. 6. The contents of the secondary structures elements estimated by ContinLL and CDSSTR tools on the basis of far-UV circular dichroism experiments performed on NSP5 wt and its mutants.

4. DISCUSSION

Rotavirus NSP5 is phosphorylated and O-glycosylated serine- and threonine-rich nonstructural protein, that localizes in viroplasms of virus-infected cells. It undergoes a complex posttranslational hyperphosphorylation, generating species with reduced PAGE mobility. The phosphorylation of NSP5 is a complex process that requires in the initial steps the interaction of at least three different partners, the substrate NSP5 in a dimeric form, NSP2, as a possible inducer of conformational change in NSP5 to make Ser 67 accessible for phosphorylation, and the cellular kinase, CK1α, responsible for such activity. The presence of glycosilation and phosphorylation in NSP5 was already known, but there was no evidence about their range and contribution to the protein's function. In this work we attempted to map these modifications using MALDI-TOF/TOF spectroscopy as a tool. The collected data allowed us to add new information on how particular regions of NSP5 are modified. Structural organization of NSP5 was unknown and in this work we found new data, that can contribute to this subject. In this discussion section we wanted to address the questions and conclusions, that appeared during these studies, and find the common points between all the three parts of the thesis.

One of distinguishing features of NSP5 in infected cells is its complex migration pattern observed in SDS-PAGE, as a consequence of its hyperphosphorylation. In some of the previous studies NSP5 was presented to have a kinase activity, and the pattern of its phosphorylation in virus infected cells has been proposed to be, in part, due to autophosphorylation. However, no clear evidences of NSP5 enzymatic activity have been reported, so the data that support that theory are very poor and not consistent with the characteristics described for eukaryotic kinases [49] [257] [258] [161].

NSP5, when expressed alone, is mostly non-phosphorylated, whilst interaction with NSP2 was shown to induce a phosphorylation cascade [49]. The studies performed in our laboratory suggested, that NSP5 hyperphosphorylation is a multi-step process that requires activation of the NSP5 substrate. We speculated the existence in NSP5 of regions specific for the activation of a cellular kinase or the substrate. The use of NSP5 deletion mutants proved to be a useful strategy to understand the molecular bases of this process. We decided to study a number of different deletion mutants on their ability to become phosphorylated *in vivo*. These studies revealed that two clearly distinct properties of the protein, related to its own

phosphorylation could be uncoupled: one such property was related to its capacity to be a substrate and the other one was related to its capacity to become an activator of the phosphorylation process [6].

The conclusion from a series of experiments previously carried out in our laboratory is that regions 1, 3, or both, prevent transfected NSP5 from getting hyperphosphorylated when expressed alone, suggesting that these two domains act as internal inhibitors in the phosphorylation mechanism [49] [152] [149] [6]. In fact, deletion of either region 1 or 3 (or both) render NSP5 able to be phosphorylated. In addition, a mutant such as Δ2, which does not get phosphorylated at all, becomes an excellent substrate when co-expressed with Δ1 or Δ 1/ Δ 3, thus indicating that region 2 is involved in inititating the cascade. One could speculate that in virus-infected cells, interaction with NSP2 may neutralize the inhibition of the two inhibitory regions.

Thus, we intended to study the role of the region 2 in the activation of phosphorylation. We decided to investigate, if phosphorylation of serines present in this region could be involved in such process. We selected in this region two serine motifs. The first motif, called motif *a* (Ser-Asp-Ser-Ala-Ser), consists of three serines at positions 63, 65 and 67, and the second one named motif *b* (Ser-Phe-Ser-Ile-Arg-Ser), contained serines 73, 75 and 78. To study these motifs, the serines were point mutated to alanine in mutant Δ3, that was previously demonstrated to play an activator role in the hyperphosphorylation mechanism. We showed, that the point mutations in serines of motif *a*, blocked two different functions of NSP5- its hyper-phosphorylation and the cellular kinase activation. A precise analysis of the three serines present in this motif, revealed clearly that serine 67 was responsible for the lack of activation function. This was confirmed with the mutation from serine to aspartic acid (this amino acid can mimic phosphorylated serine, due to the length and the negative charge of the R group [259]. The results strongly suggested, that a phosphate in serine 67 was required to restore kinase activation and ability to be hyperphosphorylated. More importantly, when this mutation (S63, 65A and S67D) was introduced in the full-length NSP5, the protein gained the ability to become hyperphosphorylated, a property that was not possible for the NSP5 wt alone, but that is characteristic for NSP5 in virus-infected cells (that is, in the presence of NSP2). On the other hand, co-expression of NSP2 with NSP5a (S63A,S65A,S67A) did not produce its hyperphosphorylation, while NSP5 S67D (S63,65A,S67D) was hyper-phosphorylated even in the absence of NSP2. These results (experiments performed by Dr Catherine Eichwald) together with the fact that deletion mutants Δ1, Δ3 and Δ1/Δ3 are hyperphosphorylated in vivo in absence of other viral proteins

[49] [152], suggested that NSP5 hyperphosphorylation is due to conformational changes in NSP5, produced by the interaction with NSP2, a situation that can be mimicked by deletion of regions 1, 3 or both, that play an inhibitory effect. In this hypothesis, serine 67 becomes available for phosphorylation by a cellular kinase(s), and this event makes the protein a permissive substrate for phsophorylation of serines and threonines of other regions, especially in region 4.

The next question we pointed was, if we could determine the kinase that is responsible for serine 67 phosphorylation. Among phosphorylation sites of known kinases we did not find exactly the sequence encompassing the motif with serine 67 that fit the consensus. However, the sequence similarity between consensus phosphorylation sites of CK1 (D/EXXS*/T*, where X is any amino acid, D is aspartic acid, E glutamic acid and S*/T* are the target serine or threonines) and motif *a* (SDSAS), suggested the possibility that motif *a* was a non-classical phosphorylation site for CK1, suggesting that this kinase could be involved in phosphorylation of serine 67. Experiments performed in vitro using of a recombinant CK1α from zebrafish confirmed this hypothesis. In the experiments two synthesized peptides, encompassing motif *a* from wt protein and with serine 67 mutated to alanine, confirmed that this sequence was a substrate for CK1α. Subsequent experiments with recombinant proteins produced in *E. coli* used as substrates, supported this finding.

However, not only phosphorylation of serine 67 by CK1α has a crucial role in the hyperphosphorylation process of NSP5. The lack of the tail (T) in deletion mutants with well established activation function, like Δ1 and Δ1/Δ3, hindered their capacity to hyperphosphorylate the substrate SV5-Δ2. As described earlier [161], the last C-terminal 10 amino acids of NSP5 appear to be involved in the interaction with another NSP5 molecule, that results in the dimerization of the protein.

On this basis we proposed a model for NSP5 hyperphosphorylation. In this model (Fig. 28), NSP5 molecule interacts with NSP2 and this interaction renders NSP5 sensitive to phosphorylation by CK1 on the serine 67. This phosphorylation step is essential, because phosphorylated serine 67 switches a cascade of events leading to hyperphosphorylated protein. This event is followed by, or parallel with dimerization of activator and substrate through the C-terminal tailpiece. The Ser-67 phosphorylated component activates the other monomer of the dimeric complex rendering it (or both) susceptible to hyperphosphorylation. In the above mentioned *in vitro* studies with recombinant CK1α from zebrafish, we have suggested that the cellular kinase CK1α can be responsible for the whole hyperphosphorylation [6]. However recent studies by means of specific RNA interference

technique performed in our laboratory by Dr Michela Campagna showed, that CK1α is the enzyme that only initiates the cascade of NSP5 phosphorylation, indicating, that kinases other than CK1α are involved *in vivo* in the downstream process. Depletion of CK1α results in lack of NSP5 hyperphosphorylation, that however did not affected neither its interaction with the virus VP1 and NSP2, proteins normally found in viroplasms, nor the production of viral proteins. In contrast, the morphology of viroplasms was clearly altered in cells, in which CK1α was depleted and a moderate decrease in the production of double-stranded RNA and infectious virus was observed [242].

In the second part of the thesis I have presented our studies on posttranslational modifications of NSP5. It was previously known that NSP5 is post-translationaly modified by introduction of two types of modifications: O-glycosylation [146] and phosphorylation [3] [49] [241]. Glycosylation of NSP5 consists in the addition of O-linked N-acetyloglucosamine (O-GlcNAc) monosaccharide residues to serines and threonines. Although it is not fully clarified, the O-glycosilation can be important, since many of known GlcNAc-modified proteins are also phosphoproteins [148]. In many cases the sites of phosphorylation and GlcNAcylation are localized to the same or neighboring residues (serines or threonines). Moreover, O-GlcNAc modification is transient, so direct interaction between glycosylation and phosphorylation can take place, regulating the protein's function and activity. Although the phosphorylation of NSP5 is responsible for the different isoforms visualized in PAGE (λ-phosphatase treatment reduces all of them to a single migrating band), it is not known whether the different isoforms have also changes in the level of O-glycosylation.

With the objective to map posttranslational modifications of NSP5, we performed studies on the protein using MALDI TOF/TOF spectroscopy. As a source of NSP5 for the analysis we used rotavirus infected cells, because our aim was to study the original modifications occurring on NSP5 during the viral life cycle, especially during the replication process, that could play an important role on NSP5 activity. Unfortunately, the amount of protein we were able to obtain using this method was low. To ensure the accuracy of the experiments, the samples were prepared for MALDI analysis by two different methods and three different strains of rotavirus were used for infections. The sequences of NSP5 encoded by these strains differ slightly, so some peptides resulting from the trypsin digestion of the same regions of the proteins have different masses, therefore when comparing results we could observe similarities in modifications of certain regions of the protein and we could

diminish the possibility of false data generation. The samples analyses let us to find many interesting observations.

One of the findings was that among the collected data, the peptide encompassing serine 67, non-phosphorylated or with one phosphate added, was present. As described earlier serine 67 is phosphorylated by CK1α and this event starts the phosphorylation process of NSP5. Interestingly, we did not find the phosphorylated peptide in the lower 26kDa band, but only in the upper bands of the PAGE analysis (or among peptides prepared by direct trypsinization of the immunoprecipitate). One possibility is that since the phosphorylation of serine 67 triggers the cascade, subsequent phosphorylations should occur very fast. Therefore, after serine 67 phoshorylation by CK1, the protein would be immediately hyperphosphorylated and shifted to higher bands of SDS-PAGE gel. Previous experiments with λ-phosphatase treatment revealed that the 26kDa band still contains phosphate, demonstrating resistance of some sites to phosphatases [3]. Even so, the obtained data suggest that phosphate observed in 26kDa band does not belong to serine 67.

The whole analysis let us to conclude, that region 1 of NSP5 is very heavily glycosilated. The role of this glycosilation is not clear, but it was demonstrated in literature that GlcNAc attachment can play a protective role [260], preventing the N-terminal part from degradation by cellular proteases. Analysis of NSP5 sequence, performed by the YinOYang 1.2 server (http://www.cbs.dtu.dk/services/YinOYang/), that predicts O-GlcNAc acceptor sites, revealed the similarities between computed results (Fig. 52) and obtained experimental data (depicted in Fig. 34), increasing their reliability.
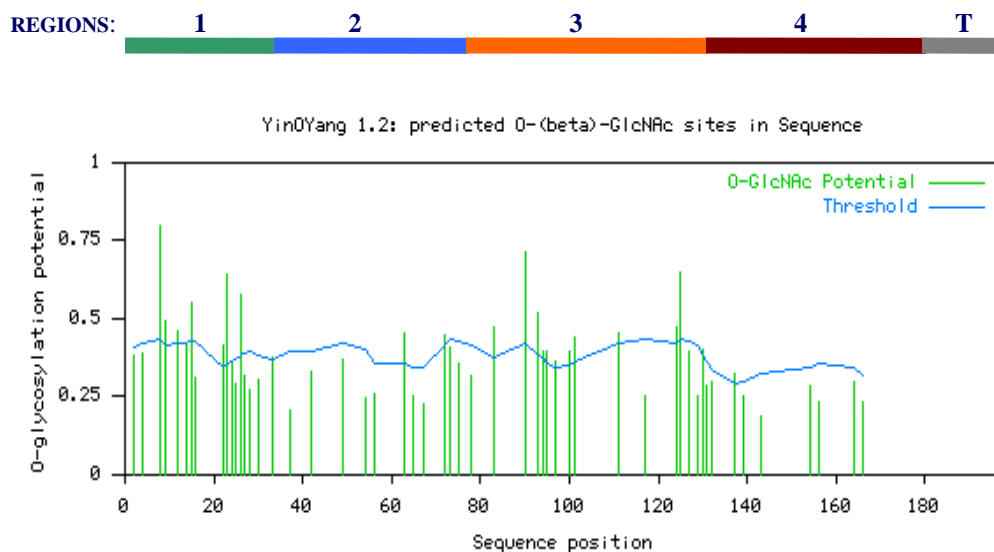
Fig. 52. Oryginal output of YinOYang 1.2 Prediction Tool, that computed O-GlcNAc acceptor sites in NSP5 sequence (the protein's regions are demonstrated as a bar above the graph). Up to the prediction the N-terminus of NSP5 is heavily-glycosilated, that is consistent with experimental data.

Furthermore, the first part of region 3, spanning from the position 80 to 115 was not represented in most of the experiments. The possible explanation for this fact can be, that this peptide may not be ionized well enough, so the machine was not able to read its signal. There are two main effects influencing the signal intensity of a peptide: first, the amount of peptides incorporated into the MALDI crystal depends on the peptide itself and on the sample preparation. Second, the capability to become ionized differs from peptide to peptide, depending on their sequence [261]. Another, possible reason for the absence of this peptide in most analyses can be the fact, that it is so highly modified that its mass was out of an efficient range of the spectrum reading. In contrast, sequence between aa 115 and 134, assigned also to region 3, was abundantly recognized in all the experiments both, non-modified and modified. This region of NSP5 sequence is different in the three analyzed strains (OSU, RF and SA11). In the bovine RF strain, eight potential modification sites are present, and in its case, three added sugar moieties and two phosphorylation sites were characterized. OSU strain contains only four serines and one threonine in the same peptide and we managed to map in these region one glycosylation and one phosphorylation site. Interestingly, glycosilation range of this region is are also similar to the result generated by the YinOYang 1.2 Prediction Server.

The beginning of region 4 revealed the addition of one N-acetyloglucosamine in all the three strains. Moreover, up to the previous data obtained in our laboratory [149], this

region was assigned to have main phosphoacceptor sites. Results obtained by MALDI TOF/TOF confirmed that this part of NSP5 is phosphorylated, because at least two phosphorylation sites were determined to be present there.

Finally, all the peptides that form the end of the region 4 and the tail were completely reconstituted and found as not modified, as this region of NSP5 lacks serines and threonines.

The idea of enclosing methylation and acetylation to selected posttranslational modifications increased the abundance of obtained data, nevertheless this eventuality must be still confirmed by biochemical studies on infected cells.

There was no evidence about the structure of NSP5, and since this kind of information often can add invaluable insight in understanding the function of a protein, we decided to study also this subject. For this purpose we produced recombinant NSP5wt and some mutants and used them to study their structural features. First, we tried to produce NSP5wt using an eukaryotic expression systems. We tested the use of vaccinia expression system and Sindbis virus replicon expression system, but these systems did not ensure appropriate expression levels of NSP5, required for structural studies. The yield of protein produced was extremely low and we then moved to the use of a bacterial expression system instead. NSP5 was produced in *E.coli* with a high yield, although we were unable to find conditions to produce the protein in soluble form. Standard changes of procedures, such as different strains used for protein production and lower temperature of bacteria cultivation, did not help to increase NSP5 solubility and prevent inclusion bodies formation. NSP5 production in bacteria was followed by development of an efficient two-steps purification protocol and refolding procedure. Refolding was performed by very slow remove of denaturing agent by dialysis, in constant presence of DTT, to prevent disulfide bonds formation because NSP5, as a cytoplasmic protein, does not have them.

The biological activity of renatured NSP5 was tested by the interaction with its molecular partner- NSP2. Designed ELISA-based binding assay demonstrated, that at least some pool of renatured protein is folded and has the ability to bind NSP2. Although the full-length NSP5 was produced with a satisfactory protein purity, significant precipitation was observed. This problem was also found in recently published literature [117], and hence this fact strikes off the protein for more detailed structural studies like crystallization. It was then decided to look for more suitable constructs.

To achieve this aim we used limited proteolysis assay, as a tool to design constructs better suited for structural biology studies. Limited proteolysis involves the incomplete

digestion of a protein by a protease. The partial digestion of a protein defines the fragments resistant to proteolysis and, therefore, expected to have a more compact structure. In contrast, flexible and exposed to solvent regions are prone to be digested.

The results coming from these experiments were quite surprising. As a protease we used trypsin, and despite the fact that NSP5 sequence contains a large number of cleavable sites we found that a big part of the protein was resistant to proteolysis. Examination by mass spectrometry of the bands remaining revealed that the part prone to digestion was the N-terminus. The link between this result and our previous observation, that in Eukaryotic cells the N-terminus of NSP5 is heavily glycosilated seems to be apparent. In bacteria the protein is not posttranslationaly modified, so carries no sugar. The sensibility of this region to digestion may signify that it does not have an organized structure, or is not buried inside the protein and so being exposed to the solvent is sensitive to the protease. This would be consistent with the hypothesis of the possible protective function of N-terminus glycosilation.

The part of the protein starting exactly at the beginning of region 2 (from aa 33) remained, in contrary, resistant to trypsinolysis. On the basis of this experiment we generated a new construct with the first 32 N-terminal residues deleted (His_6NSP5-ΔN32) that we subsequently produced, purified and refolded.

The suggestions from previous studies as well as predictions of secondary structures were strong enough to induce us to produce another mutant. Most of secondary structure prediction methods indicated that the last 25-30aa of NSP5 could be involved in the formation of an α-helical structure. Since it was demonstrated earlier that the C-terminus is involved in dimerization [161], we speculated that this α-helix could be the reason of protein dimerization by interaction between two α-helices. We also hypothesized, that outside of the cell environment and without the molecular partners, NSP2 and VP1, this C-terminal α-helices could oligomerize, causing problems, like protein's agregation. We decided to delete last 19aa in purpose to investigate if this kind of construct was a better candidate for obtaining some new information about NSP5 structure.

The production of His_6NSP5-S67D was inspired by the fact, that it has an ability to become hyperphoshorylated even in the absence of NSP2. An introduction of aspartic acid in the position of serine 67 mimics phosphorylated serine. Since, we hypothesized, that interaction with NSP2 change the conformation of NSP5 and renders it to become susceptible to phosphorylation, we wanted to check, if the different behavior of S67D mutant alter the secondary structure content, in comparison to the wt protein.

All the purified proteins were analyzed by size exclusion chromatography, to estimate their aggregation level. Although, we did not notice the precipitation of the newly produced truncation mutants, they were tested by gel filtration chromatography. Unfortunately, all proteins were eluted in the void volume, confirming that some level of aggregation was taking place. The only protein that was not eluted in a homogenous peak with the void volume, but generated also a second peak reflecting the size of a dimer was His$_6$NSP5-ΔN32. The collected peak was analyzed by Western blotting confirming that consists NSP5. Most likely a fraction of the refolded protein managed to maintain its proper structure without becoming aggregated. It is also possible that during refolding slight degradation occurred suggesting that even a shorter construct was required for structural studies, concordantly with recent studies performed by cryoelectron microscopy structures of NSP2-NSP5 complexes, where a construct encompassing aa from 66 to 188 were successfully used [117].

To investigate the secondary structure of NSP5 the proteins were analyzed by circular dichroism spectroscopy. Despite of some attempts, we did not manage to obtain promising NSP5 crystals and studies are still in the course. Having in our hands the full length protein, one point mutation and two truncation mutants, we were able to draw interesting conclusions about the structural organization of NSP5. Analysis of the results showed, that renaturated proteins were folded with a percentage of random coil (reaching almost 50%), which indicated a relevant content of unordered regions. Although introduction of a negatively charged residue at position 67 (aspartic acid) in S67D mutant changed NSP5 ability to get hyperphosphorylated without interaction with NSP2, it did not effect the spectrum shape, suggesting lack of secondary structure alteration. The analysis revealed significant amount of β-strands, but it is known that accuracy for assessment of protein secondary structure by CD is more accurate for α-helices. The very similar spectra of ΔN32 mutant and wt protein suggests that the N-terminal part does not fold into neither α nor β structures. The most meaningful observation was that removal of the C-terminus from the wild type protein (NSP5ΔC19 mutant) caused a drop of total α-helices percentage, pointing that most of the α-helical elements are situated in the carboxy-terminal part of NSP5, concordantly with predictions. We suppose, that the NSP5 dimerization may occur by formation of coiled-coil structure between two α-helices.

High amount of unordered elements and presence of distinctive C-terminal α-helix, obtained by far-UV CD experiments, was in accordance with a model (Fig. 53.) obtained by Porter, a new accurate server [262] for protein secondary and tertiary structure prediction.
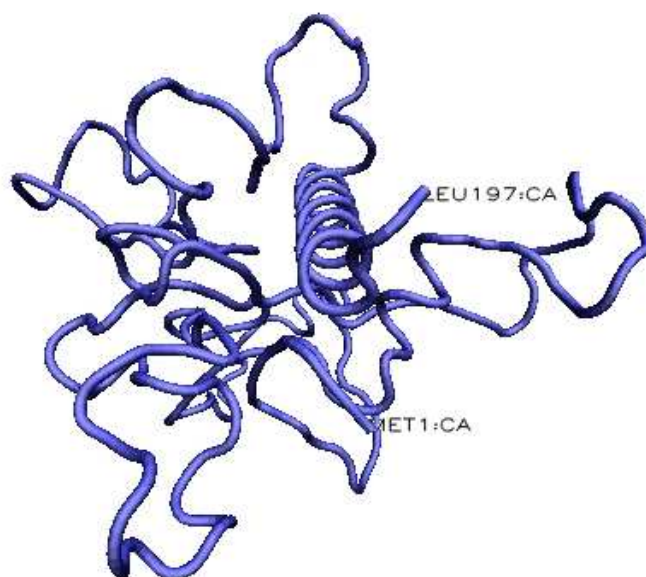
Fig. 53. Secondary and tertiary NSP5 structure prediction carried out by PORTER server. Visible C-terminal α-helix. Model pictured by VMD program.

The high percentage of disorder determined by CD experiments induced us to consider the possibility, that NSP5 is an intrinsically unfolded protein. Intrinsically unstructured/disordered or natively unfolded proteins have a broad occurrence in living organisms. They are characterized by the lack of stable secondary and tertiary structure under physiological conditions in the absence of a binding partner/ligand. Intrinsically disordered proteins fulfill essential functions, which are often linked with their disordered structural state. This kind of proteins tend to have a low sequence complexity, although it is not a general rule [263]. It has been shown recently that the more low-complexity regions an eukaryotic protein has, the less it is likely to be solubly expressed in bacteria. This can be true also in case of NSP5 and might be related to the fact that low-complexity regions, which are more frequent in eukaryotic proteins than in bacterial proteins, are more sensitive to proteolytic degradation [264]. Disordered regions often prevent crystallization of proteins, or the generation of interpretable NMR data. Therefore, structural biologists use disorder predictions to delineate compact domains to solve their 3D structure, or to dissect target sequences into a set of independently folded domains in order to facilitate tertiary structure and threading predictions [265] [266].

Taking into consideration these information we run a series of predictions of disordered regions in NSP5 (with a kind help of Dr. Sonia Longhi). The analysis was made using hints found in [265], all the addresses of used servers can also be found in the cited

paper. The graphed result of this analysis we demonstrated in figure 54. The level of disorder predicted by all the tools is very high, e.g. RONN indicates the regions lacking 3D structure in native conditions, and in case of NSP5, it is all the sequence length without its C-terminus and short region spanning from aa 110 and aa 125, that is predicted to form 3 short β-sheets. In yellow shading (aa 1-19, 47-54, 174-196) are shown the regions, that could fold in the presence of a molecular partner, so NSP2 or VP1. This regions are too short to be able to fold alone, so probably they are structured temporarily, when are in contact with their molecular partners.
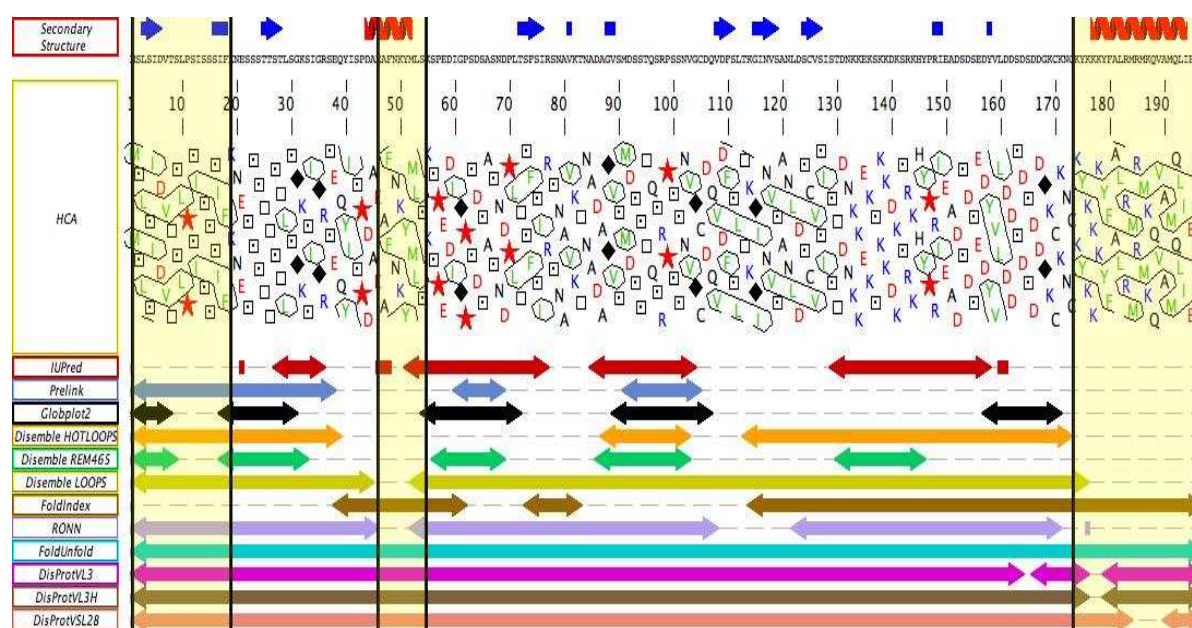
Fig. 54. Results of disorder regions for NSP5 (OSU strain) performed by different methods.

Summarizing, it is highly possible, that NSP5 is a naturally unstructured/disordered protein. Its significant portion can be flexible or disordered, and some regions of the protein can change the conformation under the interaction with its partners. NSP5 interacts with NSP2 and VP1 and it also binds nonspecifically to RNA, so its disordered nature may be necessary to impart the dynamics necessary to coordinate its functions in genome replication/packaging and assembly. Previous structural analysis of NSP2-RNA complexes indicated that RNA and NSP5 bind to very similar positively charged regions near the twofold axes of the NSP2 octamer [117]. It is known that the 48aa of C-terminus are essential for interaction with VP1 [50], the N- and C-terminal parts of NSP5 are required for the interaction with NSP2, 30 amino acids of region 3 could also be important [150]. These

regions may flexibly change its conformation, when they interfere with the ligands. The glycosilation of the N-terminus can play the protective role against cellular proteases when this part of the protein is not under interaction. The meaning of phosphorylation still remain unclear, however it can play some particular role in maintaining the dynamic equilibrium of conformation, required for proper contact with NSP2, VP1 and RNA.

REFERENCES

1.   Estes, M., *Rotaviruses and their replication*, in *Fields Virology*, D. Knipe and P. Howley, Editors. 2001, Lippincott Williams and Wilkins: New York. p. 1747-1785.
2.   Fasman, G.D., *Circular Dichroism and the Conformational Analysis of Biomolecules.* 1996.
3.   Afrikanova, I., et al., *Phosphorylation generates different forms of rotavirus NSP5.* J Gen Virol, 1996. **77 (Pt 9)**: p. 2059-65.
4.   Yeager, M., et al., *Three-dimensional structure of rhesus rotavirus by cryoelectron microscopy and image reconstruction.* J Cell Biol, 1990. **110**(6): p. 2133-44.
5.   Pesavento, J.B., et al., *The reversible condensation and expansion of the rotavirus genome.* Proc Natl Acad Sci U S A, 2001. **98**(4): p. 1381-6.
6.   Eichwald, C., et al., *Uncoupling substrate and activation functions of rotavirus NSP5: phosphorylation of Ser-67 by casein kinase 1 is essential for hyperphosphorylation.* Proc Natl Acad Sci U S A, 2004. **101**(46): p. 16304-9.
7.   Yoder, J.D.D., P.R., *Alternative intermolecular contacts underlie the rotavirus VP5* two- to three-fold rearrangement.* The EMBO Journal, 2006. **25**: p. 1559-1568.
8.   Glass, R., Parashar U.D., Bresee J.S., Turcios R., Fischer T.K., Widdowson M, Jiang B, Gentsch J.R., *Rotavirus vaccines: current prospects and future challenges.* Lancet, 2006. **368**: p. 323-32.
9.   Dormitzer, P.R., et al., *Structural rearrangements in the membrane penetration protein of a non-enveloped virus.* Nature, 2004. **430**(7003): p. 1053-8.
10.   Jayaram, H., et al., *Rotavirus protein involved in genome replication and packaging exhibits a HIT-like fold.* Nature, 2002. **417**(6886): p. 311-5.
11.   Bishop, R.F., et al., *Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis.* Lancet, 1973. **2**(7841): p. 1281-3.
12.   Parashar, U.D., et al., *Global illness and deaths caused by rotavirus disease in children.* Emerg Infect Dis, 2003. **9**(5): p. 565-72.
13.   Parashar U.D., B.J.S., Gentsch J.R., Glass R.I., *Rotavirus.* Emerging Infectious Diseases., 1998. **4**: p. 561-570.
14.   Santos, N.a.H., Y., *Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine.* Rev. Med. Virol, 2005. **15**(1): p. 29-56.
15.   Hoshino, Y.a.K., A.Z., *Rotavirus serotypes: Classification and importance in epidemiology, immunity, and vaccine development.* J. Health Popul. Nutr., 2000. **1**(18): p. 5-14.
16.   Zhang, W., et al., *Attenuation of dengue virus infection by adeno-associated virus-mediated siRNA delivery.* Genet Vaccines Ther, 2004. **2**(1): p. 8.
17.   Prasad, B.V. and M. Estes, *Molecular Basis of Rotavirus Replication. Structure-Function Correlations.*, in *Structural Biology of Viruses*, W. Chiu, R.M. Burnett, and G. R.L., Editors. 1997, Oxford University Press: New York. p. 239-268.
18.   Mindich, L., *Bacteriophage phi 6: a unique virus having a lipid containing membrane and genome composed of three dsRNA segments.* Adv. Virus Res., 1988. **35**: p. 137-176.
19.   Prasad, B.V., et al., *Three-dimensional structure of rotavirus.* J Mol Biol, 1988. **199**(2): p. 269-75.

20.   Jayaram, H., M.K. Estes, and B.V. Prasad, *Emerging themes in rotavirus cell entry, genome organization, transcription and replication.* Virus Res, 2004. **101**(1): p. 67-81.

21.   Prasad, B.V., et al., *Localization of VP4 neutralization sites in rotavirus by three-dimensional cryo-electron microscopy.* Nature, 1990. **343**(6257): p. 476-9.

22.   Yeager, M., et al., *Three-dimensional structure of the rotavirus haemagglutinin VP4 by cryo-electron microscopy and difference map analysis.* Embo J, 1994. **13**(5): p. 1011-8.

23.   Shaw, A.L., et al., *Three-dimensional visualization of the rotavirus hemagglutinin structure.* Cell, 1993. **74**(4): p. 693-701.

24.   Lawton, J.A., M.K. Estes, and B.V. Prasad, *Three-dimensional visualization of mRNA release from actively transcribing rotavirus particles.* Nat Struct Biol, 1997. **4**(2): p. 118-21.

25.   Prasad, B.V., et al., *Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus.* Nature, 1996. **382**(6590): p. 471-3.

26.   Mathieu, M., et al., *Atomic structure of the major capsid protein of rotavirus: implications for the architecture of the virion.* Embo J, 2001. **20**(7): p. 1485-97.

27.   Liu, M. and M.K. Estes, *Nucleotide sequence of the simian rotavirus SA11 genome segment 3.* Nucleic Acids Res, 1989. **17**(19): p. 7991.

28.   Poncet, D., C. Aponte, and J. Cohen, *Structure and function of rotavirus nonstructural protein NSP3.* Arch Virol Suppl, 1996. **12**: p. 29-35.

29.   Labbe, M., et al., *Expression of rotavirus VP2 produces empty corelike particles.* J Virol, 1991. **65**(6): p. 2946-52.

30.   Valenzuela, S., et al., *Photoaffinity labeling of rotavirus VP1 with 8-azido-ATP: identification of the viral RNA polymerase.* J Virol, 1991. **65**(7): p. 3964-7.

31.   Chen, D., et al., *Rotavirus open cores catalyze 5'-capping and methylation of exogenous RNA: evidence that VP3 is a methyltransferase.* Virology, 1999. **265**(1): p. 120-30.

32.   Liu, M., N.M. Mattion, and M.K. Estes, *Rotavirus VP3 expressed in insect cells possesses guanylyltransferase activity.* Virology, 1992. **188**(1): p. 77-84.

33.   Patton, J.T., *Structure and function of the rotavirus RNA-binding proteins.* J Gen Virol, 1995. **76 (Pt 11)**: p. 2633-44.

34.   Lawton, J.A., M.K. Estes, and B.V. Prasad, *Comparative structural analysis of transcriptionally competent and incompetent rotavirus-antibody complexes.* Proc Natl Acad Sci U S A, 1999. **96**(10): p. 5428-33.

35.   Lawton, J.A., M. Estes, and B.V. Prasad, *Mechanism of genome transcription in segmented dsRNA viruses.* Advances in virus research, 2000. **55**: p. 185-229.

36.   Pesavento, J.B., et al., *pH-induced conformational change of the rotavirus VP4 spike: implications for cell entry and antibody neutralization.* J Virol, 2005. **79**(13): p. 8572-80.

37.   Gouet, P., et al., *The highly ordered double-stranded RNA genome of bluetongue virus revealed by crystallography.* Cell, 1999. **97**(4): p. 481-90.

38.   Desselberger, U. and M.A. McCrae, *The rotavirus genome.* Curr Top Microbiol Immunol, 1994. **185**: p. 31-66.

39.   McCrae, M.A. and J.G. McCorquodale, *Molecular biology of rotaviruses. V. Terminal structure of viral RNA species.* Virology, 1983. **126**(1): p. 204-12.

40.   Hua, J., E.A. Mansell, and J.T. Patton, *Comparative analysis of the rotavirus NS53 gene: conservation of basic and cysteine-rich regions in the protein and possible stem-loop structures in the RNA.* Virology, 1993. **196**(1): p. 372-8.

41. Patton, J.T., et al., *Nucleotide and amino acid sequence analysis of the rotavirus nonstructural RNA-binding protein NS35.* Virology, 1993. **192**(2): p. 438-46.
42. Patton, J.T. and E. Spencer, *Genome replication and packaging of segmented double-stranded RNA viruses.* Virology, 2000. **277**(2): p. 217-25.
43. Cohen, J., et al., *Nucleotide sequence of bovine rotavirus gene 1 and expression of the gene product in baculovirus.* Virology, 1989. **171**(1): p. 131-40.
44. Patton, J.T. and D. Chen, *RNA-binding and capping activities of proteins in rotavirus open cores.* J Virol, 1999. **73**(2): p. 1382-91.
45. Patton, J.T., et al., *cis-Acting signals that promote genome replication in rotavirus mRNA.* J Virol, 1996. **70**(6): p. 3961-71.
46. Munoz, M., M. Rios, and E. Spencer, *Characteristics of single- and double-stranded RNA synthesis by a rotavirus SA-11 mutant thermosensitive in the RNA polymerase gene.* Intervirology, 1995. **38**(5): p. 256-63.
47. Patton, J.T., et al., *Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome.* J Virol, 1997. **71**(12): p. 9618-26.
48. Kattoura, M.D., X. Chen, and J.T. Patton, *The rotavirus RNA-binding protein NS35 (NSP2) forms 10S multimers and interacts with the viral RNA polymerase.* Virology, 1994. **202**(2): p. 803-13.
49. Afrikanova, I., et al., *Rotavirus NSP5 phosphorylation is up-regulated by interaction with NSP2.* J Gen Virol, 1998. **79 (Pt 11)**: p. 2679-86.
50. Arnoldi, F., Campagna, M., Eichwald, C., Desselberger, U., Burrone, O.R., *Interaction of Rotavirus Polymerase VP1 with Nonstructural Protein NSP5 Is Stronger than That with NSP2.* J. Virol., 2007. **81**: p. 2128–2137.
51. Gonzalez, S.A. and J.L. Affranchino, *Assembly of double-layered virus-like particles in mammalian cells by coexpression of human rotavirus VP2 and VP6.* J Gen Virol, 1995. **76 (Pt 9)**: p. 2357-60.
52. Zeng, C.Q., et al., *Characterization of rotavirus VP2 particles.* Virology, 1994. **201**(1): p. 55-65.
53. Zeng, C.Q., et al., *The N terminus of rotavirus VP2 is necessary for encapsidation of VP1 and VP3.* J Virol, 1998. **72**(1): p. 201-8.
54. Boyle, J.F. and K.V. Holmes, *RNA-binding proteins of bovine rotavirus.* J Virol, 1986. **58**(2): p. 561-8.
55. Labbe, M., et al., *Identification of the nucleic acid binding domain of the rotavirus VP2 protein.* J Gen Virol, 1994. **75 (Pt 12)**: p. 3423-30.
56. Mitchell, D.B. and G.W. Both, *Completion of the genomic sequence of the simian rotavirus SA11: nucleotide sequences of segments 1, 2, and 3.* Virology, 1990. **177**(1): p. 324-31.
57. Fukuhara, N., et al., *Nucleotide sequence of gene segment 1 of a porcine rotavirus strain.* Virology, 1989. **173**(2): p. 743-9.
58. Pizarro, J.L., et al., *Characterization of rotavirus guanylyltransferase activity associated with polypeptide VP3.* J Gen Virol, 1991. **72 (Pt 2)**: p. 325-32.
59. Hoshino, Y., et al., *Independent segregation of two antigenic specificities (VP3 and VP7) involved in neutralization of rotavirus infectivity.* Proc Natl Acad Sci U S A, 1985. **82**(24): p. 8701-4.
60. Fiore, L., H.B. Greenberg, and E.R. Mackow, *The VP8 fragment of VP4 is the rhesus rotavirus hemagglutinin.* Virology, 1991. **181**(2): p. 553-63.
61. Ciarlet M, L.J., Iturriza-Gomara M, Liprandi F, Gray JJ, and E.M. Desselberger U, *The initial interaction of rotavirus strains with N-acetyl-neuraminic (sialic) acid residues on the cell surface correlates with VP4 genotype, not species of origin.* J. Virol., 2002. **76**: p. 4087–4095.

62. Kirkwood, C.D., Bishop, R.F., Coulson, B.S., *Human rotavirus VP4 contains strain-specific and cross-reactive neutralization sites.* Arch. Virol., 1996. **141**: p. 587-600.

63. Hoshino, Y., A. Z. Kapikian A.Z, *Classification of rotavirus VP4 and VP7 serotypes.* Arch. Virol. Suppl., 1996. **12**: p. 99-111.

64. LaMonica, R., et al., *VP4 differentially regulates TRAF2 signaling, disengaging JNK activation while directing NF-kappa B to effect rotavirus-specific cellular responses.* J Biol Chem, 2001. **276**(23): p. 19889-96.

65. Dormitzer, P.R., et al., *Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core.* J Virol, 2002. **76**(20): p. 10512-7.

66. Monnier N., H.-M.K., Sun Z-Y. J., Prasad B.V., Taniguchi K., Dormitzer P.R, *High-Resolution Molecular and Antigen Structure of the VP8* Core of a Sialic Acid-Independent Human Rotavirus Strain.* J. Virol., 2006. **80**: p. 1513-1523.

67. Maass, D.R. and P.H. Atkinson, *Rotavirus proteins VP7, NS28, and VP4 form oligomeric structures.* J Virol, 1990. **64**(6): p. 2632-41.

68. Petrie, B.L., et al., *Ultrastructural localization of rotavirus antigens using colloidal gold.* Virus Res, 1984. **1**(2): p. 133-52.

69. Enouf, V., et al., *Interactions of rotavirus VP4 spike protein with the endosomal protein Rab5 and the prenylated Rab acceptor PRA1.* J Virol, 2003. **77**(12): p. 7041-7.

70. Bican, P., et al., *Purification and characterization of bovine rotavirus cores.* J Virol, 1982. **43**(3): p. 1113-7.

71. Charpilienne, A., et al., *Identification of rotavirus VP6 residues located at the interface with VP2 that are essential for capsid assembly and transcriptase activity.* J Virol, 2002. **76**(15): p. 7822-31.

72. Affranchino, J.L. and S.A. Gonzalez, *Deletion mapping of functional domains in the rotavirus capsid protein VP6.* J Gen Virol, 1997. **78 (Pt 8)**: p. 1949-55.

73. Thouvenin, E., et al., *Antibody inhibition of the transcriptase activity of the rotavirus DLP: a structural view.* J Mol Biol, 2001. **307**(1): p. 161-72.

74. Kabcenell, A.K. and P.H. Atkinson, *Processing of the rough endoplasmic reticulum membrane glycoproteins of rotavirus SA11.* J Cell Biol, 1985. **101**(4): p. 1270-80.

75. Arias, C.F., S. Lopez, and R.T. Espejo, *Gene protein products of SA11 simian rotavirus genome.* J Virol, 1982. **41**(1): p. 42-50.

76. Ericson, B.L., et al., *Identification, synthesis, and modifications of simian rotavirus SA11 polypeptides in infected cells.* J Virol, 1982. **42**(3): p. 825-39.

77. Kabcenell, A.K., et al., *Two forms of VP7 are involved in assembly of SA11 rotavirus in endoplasmic reticulum.* J Virol, 1988. **62**(8): p. 2929-41.

78. Both, G.W., J.S. Mattick, and A.R. Bellamy, *Serotype-specific glycoprotein of simian 11 rotavirus: coding assignment and gene sequence.* Proc Natl Acad Sci U S A, 1983. **80**(10): p. 3091-5.

79. Ericson, B.L., et al., *Two types of glycoprotein precursors are produced by the simian rotavirus SA11.* Virology, 1983. **127**(2): p. 320-32.

80. Stirzaker, S.C., et al., *Processing of rotavirus glycoprotein VP7: implications for the retention of the protein in the endoplasmic reticulum.* J Cell Biol, 1987. **105**(6 Pt 2): p. 2897-903.

81. Munro, S. and H.R. Pelham, *A C-terminal signal prevents secretion of luminal ER proteins.* Cell, 1987. **48**(5): p. 899-907.

82. Mattion, N., J. Cohen, and M. Estes, *The rotavirus proteins*, in *Viral Infections of gastrointestinal tract.*, A.Z. Kapikian, Editor. 1994, Marcel Decker: New York. p. 169-249.

83. Mirazimi, A. and L. Svensson, *ATP is required for correct folding and disulfide bond formation of rotavirus VP7.* J Virol, 2000. **74**(17): p. 8048-52.

84.   Mirazimi, A., M. Nilsson, and L. Svensson, *The molecular chaperone calnexin interacts with the NSP4 enterotoxin of rotavirus in vivo and in vitro.* J Virol, 1998. **72**(11): p. 8705-9.

85.   Beisner, B., Kool, D., Marich, A., Holmes, I.H., *Characterisation of G serotype dependent non-antibody inhibitors of rotavirus in normal mouse serum.* Arch. Virol., 1998. **143**: p. 1277–1294.

86.   Mendez, E., et al., *Entry of rotaviruses is a multistep process.* Virology, 1999. **263**(2): p. 450-9.

87.   Zarate, S., et al., *VP7 mediates the interaction of rotaviruses with integrin alphavbeta3 through a novel integrin-binding site.* J Virol, 2004. **78**(20): p. 10839-47.

88.   Graham, K.L., et al., *Integrin-using rotaviruses bind alpha2beta1 integrin alpha2 I domain via VP4 DGE sequence and recognize alphaXbeta2 and alphaVbeta3 by using VP7 during cell entry.* J Virol, 2003. **77**(18): p. 9969-78.

89.   Gajardo, R., et al., *Two proline residues are essential in the calcium-binding activity of rotavirus VP7 outer capsid protein.* J Virol, 1997. **71**(3): p. 2211-6.

90.   Ruiz, M.C., et al., *The concentration of Ca2+ that solubilizes outer capsid proteins from rotavirus particles is dependent on the strain.* J Virol, 1996. **70**(8): p. 4877-83.

91.   Dormitzer, P.R. and H.B. Greenberg, *Calcium chelation induces a conformational change in recombinant herpes simplex virus-1-expressed rotavirus VP7.* Virology, 1992. **189**(2): p. 828-32.

92.   Dormitzer, P.R., H.B. Greenberg, and S.C. Harrison, *Purified recombinant rotavirus VP7 forms soluble, calcium-dependent trimers.* Virology, 2000. **277**(2): p. 420-8.

93.   Hua, J., X. Chen, and J.T. Patton, *Deletion mapping of the rotavirus metalloprotein NS53 (NSP1): the conserved cysteine-rich region is essential for virus-specific RNA binding.* J Virol, 1994. **68**(6): p. 3990-4000.

94.   Gonzalez, R.A., et al., *In vivo interactions among rotavirus nonstructural proteins.* Arch Virol, 1998. **143**(5): p. 981-96.

95.   Brottier, P., et al., *Bovine rotavirus segment 5 protein expressed in the baculovirus system interacts with zinc and RNA.* J Gen Virol, 1992. **73 (Pt 8)**: p. 1931-8.

96.   Graff, J.W., et al., *Interferon regulatory factor 3 is a cellular partner of rotavirus NSP1.* J Virol, 2002. **76**(18): p. 9545-50.

97.   Silvestri, L.S., Z.F. Taraporewala, and J.T. Patton, *Rotavirus replication: plus-sense templates for double-stranded RNA synthesis are made in viroplasms.* J Virol, 2004. **78**(14): p. 7763-74.

98.   Patton, J.T., et al., *Effect of intragenic rearrangement and changes in the 3' consensus sequence on NSP1 expression and rotavirus replication.* J Virol, 2001. **75**(5): p. 2076-86.

99.   Taniguchi, K., K. Kojima, and S. Urasawa, *Nondefective rotavirus mutants with an NSP1 gene which has a deletion of 500 nucleotides, including a cysteine-rich zinc finger motif-encoding region (nucleotides 156 to 248), or which has a nonsense codon at nucleotides 153-155.* J Virol, 1996. **70**(6): p. 4125-30.

100.   Barro, M. and J.T. Patton, *Rotavirus nonstructural protein 1 subverts innate immune response by inducing degradation of IFN regulatory factor 3.* Proc Natl Acad Sci U S A, 2005. **102**(11): p. 4114-9.

101.   Sharma, S., tenOever, B. R., Grandvaux, N., Zhou, G. P., Lin, R. & Hiscott, and J., Science, 2003. **300**: p. 1148–1151.

102.   Knipe, D., C.E. Samuel, and P. Palese, *Virus-Host Interactions*, in *Fields Virology*, D. Knipe and P. Howley, Editors. 2001, Lippincott Williams and Wilkins: New York. p. 133-170.

103. Dunn, S.J., T.L. Cross, and H.B. Greenberg, *Comparison of the rotavirus nonstructural protein NSP1 (NS53) from different species by sequence analysis and northern blot hybridization.* Virology, 1994. **203**(1): p. 178-83.

104. Aponte, C., D. Poncet, and J. Cohen, *Recovery and characterization of a replicase complex in rotavirus-infected cells by using a monoclonal antibody against NSP2.* J Virol, 1996. **70**(2): p. 985-91.

105. Gallegos, C.O. and J.T. Patton, *Characterization of rotavirus replication intermediates: a model for the assembly of single-shelled particles.* Virology, 1989. **172**(2): p. 616-27.

106. Ramig, R.F. and B.L. Petrie, *Characterization of temperature-sensitive mutants of simian rotavirus SA11: protein synthesis and morphogenesis.* J Virol, 1984. **49**(3): p. 665-73.

107. Chen, D., J.L. Gombold, and R.F. Ramig, *Intracellular RNA synthesis directed by temperature-sensitive mutants of simian rotavirus SA11.* Virology, 1990. **178**(1): p. 143-51.

108. Taraporewala, Z.F., D. Chen, and J.T. Patton, *Multimers of the bluetongue virus nonstructural protein, NS2, possess nucleotidyl phosphatase activity: similarities between NS2 and rotavirus NSP2.* Virology, 2001. **280**(2): p. 221-31.

109. Taraporewala, Z.F., Jiang, X., Vasquez-Del Carpio, R., and H. Jayaram, Prasad, B. V. & Patton, J. T., *Structure-function analysis of rotavirus NSP2 octamer by using a novel complementation system.* J. Virol., 2006. **80**: p. 7984–7994.

110. Taraporewala, Z., D. Chen, and J.T. Patton, *Multimers formed by the rotavirus nonstructural protein NSP2 bind to RNA and have nucleoside triphosphatase activity.* J Virol, 1999. **73**(12): p. 9934-43.

111. Taraporewala, Z.F. and J.T. Patton, *Identification and characterization of the helix-destabilizing activity of rotavirus nonstructural protein NSP2.* J Virol, 2001. **75**(10): p. 4519-27.

112. Carpio, R.V., et al., *Role of the histidine triad-like motif in nucleotide hydrolysis by the rotavirus RNA-packaging protein NSP2.* J Biol Chem, 2004. **279**(11): p. 10624-33.

113. Schuck, P., et al., *Rotavirus nonstructural protein NSP2 self-assembles into octamers that undergo ligand-induced conformational changes.* J Biol Chem, 2001. **276**(13): p. 9679-87.

114. Carpio, R.V., Gonzalez-Nilo, F. D., Riadi, G., Patton, J. T., Taraporewala Z. F., *Histidine triad-like motif of the rotavirus NSP2 octamer mediates both RTPase and NTPase activities.* J.Mol. Biol., 2006. **362**: p. 539–554.

115. Lima, C.D., M.G. Klein, and W.A. Hendrickson, *Structure-based analysis of catalysis and substrate definition in the HIT protein family.* Science, 1997. **278**(5336): p. 286-90.

116. Kumar, M., Jayaram, H., Carpio, R., Jacobson, R., and J.T.P. Patton, B. V., *A novel viral NDP kinase with a cellular HIT fold.* 2006.

117. Jiang, X., Jayaram, H., Ludtke, S., Estes, M. & Prasad, and B. V., *Cryo-EM structures of rotavirus NSP2-NSP5 and NSP2-RNA complexes: Implications for genome replication.* J. Virol. In review., 2006.

118. Mattion, N.M., et al., *Characterization of an oligomerization domain and RNA-binding properties on rotavirus nonstructural protein NS34.* Virology, 1992. **190**(1): p. 68-83.

119. Poncet, D., S. Laurent, and J. Cohen, *Four nucleotides are the minimal requirement for RNA recognition by rotavirus non-structural protein NSP3.* Embo J, 1994. **13**(17): p. 4165-73.

120. Piron, M., et al., *Rotavirus RNA-binding protein NSP3 interacts with eIF4GI and evicts the poly(A) binding protein from eIF4F.* Embo J, 1998. **17**(19): p. 5811-21.

121. Piron, M., et al., *Identification of the RNA-binding, dimerization, and eIF4GI-binding domains of rotavirus nonstructural protein NSP3.* J Virol, 1999. **73**(7): p. 5411-21.

122. Vende, P., et al., *Efficient translation of rotavirus mRNA requires simultaneous interaction of NSP3 with the eukaryotic translation initiation factor eIF4G and the mRNA 3' end.* J Virol, 2000. **74**(15): p. 7064-71.

123. Deo, R.C., et al., *Recognition of the rotavirus mRNA 3' consensus by an asymmetric NSP3 homodimer.* Cell, 2002. **108**(1): p. 71-81.

124. Nibert, M.L., *Rotavirus translation control protein takes RNA to heart.* Structure (Camb), 2002. **10**(2): p. 129-30.

125. Poncet, D., C. Aponte, and J. Cohen, *Rotavirus protein NSP3 (NS34) is bound to the 3' end consensus sequence of viral mRNAs in infected cells.* J Virol, 1993. **67**(6): p. 3159-65.

126. Varani, G. and F.H. Allain, *How a rotavirus hijacks the human protein synthesis machinery.* Nat Struct Biol, 2002. **9**(3): p. 158-60.

127. Groft, C.M. and S.K. Burley, *Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization.* Mol Cell, 2002. **9**(6): p. 1273-83.

128. Montero H., A.C.F., Lopez S., *Rotavirus Nonstructural Protein NSP3 Is Not Required for Viral Protein Synthesis.* J. Virol., 2006. **80**: p. 9031–9038.

129. Gradi, A., H. Imataka, Y. V. Svitkin, E. Rom, B. Raught, S. Morino, and N. and Sonenberg, *A novel functional human eukaryotic translation initiation factor 4G.* Mol. Cell. Biol., 1998. **18**: p. 334–342.

130. Bergmann, C.C., Maass, D., Poruchynsky, M.S., Atkinson, P.H. and Bellamy, A.R. and, *Topology of the non-structural rotavirus receptor glycoprotein ns28 in the rough endoplasmic reticulum.* EMBO J. **8**(1695-1703.).

131. Ball, J.M., Mitchell, D.M., Gibbons, T.F. and Parr, R.D., *Rotavirus NSP4: A multifunctional viral enterotoxin.* Viral. Immunol., 2005. **18**: p. 27-40.

132. Bowman, G.D., et al., *Crystal structure of the oligomerization domain of NSP4 from rotavirus reveals a core metal-binding site.* J Mol Biol, 2000. **304**(5): p. 861-71.

133. Meyer, J.C., C.C. Bergmann, and A.R. Bellamy, *Interaction of rotavirus cores with the nonstructural glycoprotein NS28.* Virology, 1989. **171**(1): p. 98-107.

134. Au, K.S., et al., *Receptor activity of rotavirus nonstructural glycoprotein NS28.* J Virol, 1989. **63**(11): p. 4553-62.

135. Taylor, J.A., et al., *The RER-localized rotavirus intracellular receptor: a truncated purified soluble form is multivalent and binds virus particles.* Virology, 1993. **194**(2): p. 807-14.

136. Xu, A., A.R. Bellamy, and J.A. Taylor, *Immobilization of the early secretory pathway by a virus glycoprotein that binds to microtubules.* Embo J, 2000. **19**(23): p. 6465-74.

137. Lopez, T., et al., *Silencing the morphogenesis of rotavirus.* J Virol, 2005. **79**(1): p. 184-92.

138. Ball, J.M., et al., *Age-dependent diarrhea induced by a rotaviral nonstructural glycoprotein.* Science, 1996. **272**(5258): p. 101-4.

139. Zhang, M., et al., *Mutations in rotavirus nonstructural glycoprotein NSP4 are associated with altered virus virulence.* J Virol, 1998. **72**(5): p. 3666-72.

140. Dong, Y., Zeng, C.Q., Ball, J.M., Estes, M.K. and Morris, A.P., *The rotavirus enterotoxin NSP4 mobilizes intracellular calcium in human intestinal cells by stimulating phospholipase c-mediated inositol 1,4,5-trisphosphate production.* Proc. Natl. Acad. Sci. USA, 1997. **94**(8): p. 3960-3965.

141. Tian, P., Estes, M.K., Hu, Y., Ball, J.M., Zeng, C.Q. and Schilling, W.P., *The rotavirus nonstructural glycoprotein NSP4 mobilizes Ca2+ from the endoplasmic reticulum.* J. Virol., 1995. **69**: p. 5763-5772.

142. Michelangeli, F., et al., *Selective depletion of stored calcium by thapsigargin blocks rotavirus maturation but not the cytopathic effect.* J Virol, 1995. **69**(6): p. 3838-47.

143. Ruiz, M.C., J. Cohen, and F. Michelangeli, *Role of Ca2+in the replication and pathogenesis of rotavirus and other viral infections.* Cell Calcium, 2000. **28**(3): p. 137-49.

144. Poruchynsky, M.S., D.R. Maass, and P.H. Atkinson, *Calcium depletion blocks the maturation of rotavirus by altering the oligomerization of virus-encoded proteins in the ER.* J Cell Biol, 1991. **114**(4): p. 651-6.

145. Welch, S.K., S.E. Crawford, and M.K. Estes, *Rotavirus SA11 genome segment 11 protein is a nonstructural phosphoprotein.* J Virol, 1989. **63**(9): p. 3974-82.

146. Gonzalez, S.A. and O.R. Burrone, *Rotavirus NS26 is modified by addition of single O-linked residues of N-acetylglucosamine.* Virology, 1991. **182**(1): p. 8-16.

147. Wells L, H.G., *O-GlcNAc turns twenty: functional implications for post-translational modification of nuclear and cytosolic proteins with a sugar.* FEBS Lett., 2003. **546**(1): p. 154-8.

148. Comer F. I., H.G.W., *O-Glycosylation of Nuclear and Cytosolic Proteins. DYNAMIC INTERPLAY BETWEEN O-GlcNAc AND O-PHOSPHATE.* J. Biol. Chem., 2000. **275**: p. 29179.

149. Eichwald, C., et al., *Rotavirus NSP5: mapping phosphorylation sites and kinase activation and viroplasm localization domains.* J Virol, 2002. **76**(7): p. 3461-70.

150. Eichwald, C., J.F. Rodriguez, and O.R. Burrone, *Characterisation of rotavirus NSP2/NSP5 interaction and dynamics of viroplasms formation.* J Gen Virol, 2004. **85**: p. 625-634.

151. Poncet, D., et al., *In vivo and in vitro phosphorylation of rotavirus NSP5 correlates with its localization in viroplasms.* J Virol, 1997. **71**(1): p. 34-41.

152. Fabbretti, E., et al., *Two non-structural rotavirus proteins, NSP2 and NSP5, form viroplasm-like structures in vivo.* J Gen Virol, 1999. **80 (Pt 2)**: p. 333-9.

153. Estes, M.K. and J. Cohen, *Rotavirus gene structure and function.* Microbiol Rev, 1989. **53**(4): p. 410-49.

154. Taraporewala, Z.F. and J.T. Patton, *Nonstructural proteins involved in genome packaging and replication of rotaviruses and other members of the Reoviridae.* Virus Res, 2004. **101**(1): p. 57-66.

155. Vascotto, F., et al., *Effects of intrabodies specific for rotavirus NSP5 during the virus replicative cycle.* J Gen Virol, 2004. **85**(Pt 11): p. 3285-90.

156. Campagna, M., et al., *RNA interference of rotavirus segment 11 mRNA reveals the essential role of NSP5 in the virus replicative cycle.* J Gen Virol, 2005. **86**(Pt 5): p. 1481-7.

157. Vende, P., Z.F. Taraporewala, and J.T. Patton, *RNA-binding activity of the rotavirus phosphoprotein NSP5 includes affinity for double-stranded RNA.* J Virol, 2002. **76**(10): p. 5291-9.

158. Araujo, I.T., et al., *Rotavirus genotypes P[4]G9, P[6]G9, and P[8]G9 in hospitalized children with acute gastroenteritis in Rio de Janeiro, Brazil.* J Clin Microbiol, 2001. **39**(5): p. 1999-2001.

159. Patton, J.T., *Rotavirus VP1 alone specifically binds to the 3' end of viral mRNA, but the interaction is not sufficient to initiate minus-strand synthesis.* J Virol, 1996. **70**(11): p. 7940-7.

160.   Zeng, C.Q., et al., *Characterization and replicase activity of double-layered and single-layered rotavirus-like particles expressed from baculovirus recombinants.* J Virol, 1996. **70**(5): p. 2736-42.

161.   Torres-Vega, M.A., et al., *The C-terminal domain of rotavirus NSP5 is essential for its multimerization, hyperphosphorylation and interaction with NSP6.* J Gen Virol, 2000. **81**(Pt 3): p. 821-30.

162.   Ciarlet, M. and M.K. Estes, *Human and most animal rotavirus strains do not require the presence of sialic acid on the cell surface for efficient infectivity.* J Gen Virol, 1999. **80 (Pt 4)**: p. 943-8.

163.   Fukuhara, N., Yoshie, O., Kitaoka, S., Konno, T., *Role of VP3 in human rotavirus internalization after target cell attachment via VP7.* J. Virol., 1988. **62**: p. 2209–2218.

164.   Kalica, A.R., J. Flores, and H.B. Greenberg, *Identification of the rotaviral gene that codes for hemagglutination and protease-enhanced plaque formation.* Virology, 1983. **125**(1): p. 194-205.

165.   Mackow, E.R., et al., *The rhesus rotavirus outer capsid protein VP4 functions as a hemagglutinin and is antigenically conserved when expressed by a baculovirus recombinant.* J Virol, 1989. **63**(4): p. 1661-8.

166.   Coulson, B.S., S.L. Londrigan, and D.J. Lee, *Rotavirus contains integrin ligand sequences and a disintegrin-like domain that are implicated in virus entry into cells.* Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5389-94.

167.   Hewish, M.J., Y. Takada, and B.S. Coulson, *Integrins alpha2beta1 and alpha4beta1 can mediate SA11 rotavirus attachment and entry into cells.* J Virol, 2000. **74**(1): p. 228-36.

168.   Guerrero, C.A., et al., *Heat shock cognate protein 70 is involved in rotavirus cell entry.* J Virol, 2002. **76**(8): p. 4096-102.

169.   Zarate, S., et al., *Interaction of rotaviruses with Hsc70 during cell entry is mediated by VP5.* J Virol, 2003. **77**(13): p. 7254-60.

170.   Sanchez-San Martin, C., et al., *Characterization of rotavirus cell entry.* J Virol, 2004. **78**(5): p. 2310-8.

171.   Cohen, J., et al., *Activation of rotavirus RNA polymerase by calcium chelation.* Arch Virol, 1979. **60**(3-4): p. 177-86.

172.   Stacy-Phipps, S. and J.T. Patton, *Synthesis of plus- and minus-strand RNA in rotavirus-infected cells.* J Virol, 1987. **61**(11): p. 3479-84.

173.   Pizarro, J.M., et al., *Effect of nucleotide analogues on rotavirus transcription and replication.* Virology, 1991. **184**(2): p. 768-72.

174.   Pesavento, J.B., Estes, M.K., Prasad, B.V.V., *Structural organization of the genome in rotavirus.* Viral Gastroenteritis; Desselberger, U., Gray, J., 2003. **9**: p. 115-128.

175.   Patton, J.T. and C.O. Gallegos, *Structure and protein composition of the rotavirus replicase particle.* Virology, 1988. **166**(2): p. 358-65.

176.   Chen, D., et al., *Template-dependent, in vitro replication of rotavirus RNA.* J Virol, 1994. **68**(11): p. 7030-9.

177.   Wentz, M.J., J.T. Patton, and R.F. Ramig, *The 3'-terminal consensus sequence of rotavirus mRNA is the minimal promoter of negative-strand RNA synthesis.* J Virol, 1996. **70**(11): p. 7833-41.

178.   Chen, D., et al., *Features of the 3'-consensus sequence of rotavirus mRNAs critical to minus strand synthesis.* Virology, 2001. **282**(2): p. 221-9.

179.   Chen, D. and J.T. Patton, *Rotavirus RNA replication requires a single-stranded 3' end for efficient minus-strand synthesis.* J Virol, 1998. **72**(9): p. 7387-96.

180. Tortorici, M.A., et al., *Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus.* J Biol Chem, 2003. **278**(35): p. 32673-82.

181. Chizhikov, V. and J.T. Patton, *A four-nucleotide translation enhancer in the 3'-terminal consensus sequence of the nonpolyadenylated mRNAs of rotavirus.* Rna, 2000. **6**(6): p. 814-25.

182. Carlos F. Arias, M.A.D., Lorenzo Segovia, Tomás López, Minerva Camacho, and R.E. Pavel Isa, Susana López, *RNA silencing of rotavirus gene expression.* Virus Research, 2004. **102**: p. 43-51.

183. Petrie, B.L., et al., *Localization of rotavirus antigens in infected cells by ultrastructural immunocytochemistry.* J Gen Virol, 1982. **63**(2): p. 457-67.

184. Poruchynsky, M.S. and P.H. Atkinson, *Rotavirus protein rearrangements in purified membrane-enveloped intermediate particles.* J Virol, 1991. **65**(9): p. 4720-7.

185. Tian, P., et al., *The rotavirus nonstructural glycoprotein NSP4 possesses membrane destabilization activity.* J Virol, 1996. **70**(10): p. 6973-81.

186. Delmas, O., et al., *Spike protein VP4 assembly with maturing rotavirus requires a postendoplasmic reticulum event in polarized caco-2 cells.* J Virol, 2004. **78**(20): p. 10987-94.

187. Jourdan, N., et al., *Rotavirus is released from the apical surface of cultured human intestinal cells through nonconventional vesicular transport that bypasses the Golgi apparatus.* J Virol, 1997. **71**(11): p. 8268-78.

188. Kosek M, B.C., Guerrant RL., *The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000.* Bull World Health Organ, 2003(81): p. 197-204.

189. Parashar U, G.C., Bresee JS, Glass RI., *Rotavirus and severe childhood diarrhea.* Emerg Infect Dis, 2005. **12**: p. 304-06.

190. Jain V, P.U., Glass RI, Bhan MK., *Epidemiology of rotavirus in India.* Indian J Pediatr, 2001(68): p. 855-62.

191. Zhou, Y., et al., *Characterization of human rotavirus serotype G9 isolated in Japan and Thailand from 1995 to 1997.* J Med Virol, 2001. **65**(3): p. 619-28.

192. Kapikian AZ, H.Y., Chanock RM., and Howley PM, *Rotaviruses. In: Knipe DM.* Fields virology, 2001. **2001**: p. 1787–833.

193. Kapikian AZ, H.Y., Chanock RM, Perez-Schael I., *Efficacy of a quadrivalent rhesus rotavirus-based human rotavirus vaccine aimed at preventing severe rotavirus diarrhea in infants and young children.* J Infect Dis, 1996. **174**(suppl 1): p. 65-72.

194. Peter, G. and M.G. Myers, *Intussusception, rotavirus, and oral vaccines: summary of a workshop.* Pediatrics, 2002. **110**(6): p. e67.

195. Parashar UD, H.R., Cummings KC, et al., *Trends in intussusception-associated hospitalizations and deaths among US infants.* Pediatrics, 2000. **106**: p. 1413-21.

196. De Vos, B., Vesikari, T., Linhares, AC., et al. and, *A rotavirus vaccine for prophylaxis of infants against rotavirus gastroenteritis.* Pediatr Infect Dis J, 2004. **23**(**10 suppl**): p. 179–82.

197. Ruiz-Palacios GM, P.-S.I., Velazquez FR, et al., *Safety and effi cacy of an attenuated vaccine against severe rotavirus gastroenteritis.* N Engl J Med, 2006. **354**: p. 11-22.

198. Woody, R.W.a.D., A.K., *Circular Dichroism and the Conformational Analysis of Biomolecules (Fasman, G.D., ed.).* 1996: p. 109-157.

199. Greenfield NJ, F.G., *Computed circular dichroism spectra for the evaluation of protein conformation.* Biochemistry, 1969. **8**: p. 4108–4116.

200. Chen YH, Y.J., *A new approach to the calculation of secondary structures of globular proteins by optical rotatory dispersion and circular dichroism.* Biochem Biophys Res Commun, 1971. **44**: p. 1285–1291.

201. Bolotina IA, C.V., Lugauskas VY, Finkel'shtein AV, Ptitsyn OB., *Determination of the secondary structure of proteins from the circular dichroism spectra. 1. Protein reference spectra for a-, b- and irregular structures.* Mol Biol (Eng Transl Mol Biol), 1980. **14**: p. 701-709.

202. Brahms S, B.J., *Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism.* J Mol Biol, 1980. **138**: p. 149-178.

203. Hennessey JP Jr, J.W.J., *Information content in the circular dichroism of proteins.* Biochemistry, 1981. **20**: p. 1085–1094.

204. Provencher SW, G.J., *Estimation of protein secondary structure from circular dichroism.* Biochemistry, 1981. **20**: p. 33-37.

205. Pancoska P, Y.S., Keiderling TA., *Statistical analysis of the vibrational circular dichroism of selected proteins and relationship to secondary structure.* Biochemistry, 1991. **30**: p. 5089–5103.

206. Böhm G, M.R., Jaenicke R., *Quantitative analysis of protein far UV circular dichroism spectra by neural networks.* Protein Eng, 1992. **5**: p. 191–195.

207. Sciences, N.A.o., *Defining the Mandate of Proteomics in the Post-Genomics Era: Workshop Report.*, in *Molecular & Cellular Proteomics.* 2002. p. 763-780.

208. Barber M., B.R.S. Sedgwick R. D, Tyler A. N., Nature, 1981. **293**: p. 270–275.

209. Fenn, J.B., M. Mann, C. K. Meng, S. F. Wong, and C. M. Whitehouse, *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**: p. 64-71.

210. Thogersen D.F., S.R.P.a.M.R.D., Biochem. Biophis. Res. Commun., 1974. **60**: p. 616-621.

211. Karas, M., F. Hillenkamp, *Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons.* Anal. Chem., 1988. **60**: p. 2299-2301.

212. Tanaka, K., H. Waki, Y. Ido, S. Akita, Y. Yoshida, and T. Yoshida, *Protein and polymer analysis up to m/z 100,000 by laser ionization time-of-flight mass spectrometry.* Rapid Commun. Mass Spectrom, 1988. **2**: p. 151-153.

213. Aebersold, R., Mann, M., *Mass spectrometry-based proteomics.* Nature, 2003. **422**: p. 198-207.

214. Gygi, S.P., Y. Rochon, R. Franza, and R. Aebersold, *Correlation between Protein and mRNA Abundance in Yeast.* Mol. Cell. Biol., 1999. **19**: p. 1720-1730.

215. Gygi, S.P., B. Rist, S. A. Gerber, F. Turecek, M. H. Gelb, and R. Aebersold, *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nat. Biotechnol., 1999. **19**: p. 994-999.

216. Wilkins, M.R., E. Gasteiger, A. A. Gooley, B. R. Herbert, M. Molley, P. A. Binz, K. Ou, J. C. and A.B. Sanchez, K. L. Williams, and D. F. Hochstrasser, *High Throughput Mass Spectrometric Discovery of Protein Post-translational Modifications.* J. Mol. Biol., 1999. **289**: p. 645-657.

217. Ashman, K., Moran, M. F., Sicheri, F., Pawson, T., Tyers, M., *Cell signalling- the proteomics of it all.* Science's STKE <http://stke.sciencemag.org/cgi/content/full/sigtrans;2001/103/pe33>, 2001.

218. Rappsilber, J., Siniossoglou, S., Hurt, E. C., Mann, M., *A generic strategy to analyze the spatial organization of multi-protein complexes by cross-linking and mass spectrometry.* Anal. Chem., 2000. **72**: p. 267–275.

219. Mann, M., C. K. Meng, and J. B. Fenn, *"Interpreting Mass Spectra of Multiply Charged Ions."* Anal. Chem., 1989. **61**: p. 1702-1708.

220. Wilm M, M.M., *Analytical properties of the nanoelectrospray ion source.* Anal Chem., 1996. **68**: p. 1-8.
221. Aebersold, R., Goodlett, D. R., *Mass spectrometry in proteomics.* Chem. Rev., 2001. **101**: p. 269–295.
222. Deutzmann, R., *Structural characterization of proteins and peptides.* Methods Mol Med., 2004. **94**: p. 269-97.
223. Beavis RC, C.B., *High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry.* Anal Chem., 1990. **62**: p. 1836-40.
224. Zhou J, E.W., Standing KG, Verentchikov A., *Kinetic energy measurements of molecular ions ejected into an electric field by matrix-assisted laser desorption.* Rapid Commun Mass Spectrom., 1992. **6**: p. 671-8.
225. Brown RS, L.J., *Mass resolution improvement by incorporation of pulsed ion extraction in a matrix-assisted laser desorption/ionization linear time-of-flight mass spectrometer.* Anal Chem., 1995. **67**: p. 1998-2003.
226. Mamyrin, B.A., *Ref Type: Patent.* 1966: Russia.
227. Krutchinsky AN, K.M., Chait BT., *Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer.* Anal Chem., 2001. **73**: p. 5066-77.
228. Loboda, A.V., A. N. Krutchinsky, M. Bromirski, W. Ens, and K. G. Standing, *A tandem quadrupole/time-of-flight mass spectrometer with a matrix- assisted laser desorption/ionization source: design and performance.* Rapid Commun. Mass Spectrom., 2000. **14**: p. 1047-1057.
229. Medzihradszky, K.F., J. M. Campbell, M. A. Baldwin, A. M. Falick, P. Juhasz, M. L. Vestal, and A. L. Burlingame, *The characteristics of peptide collision-induced dissociation using a high-performance MALDI-TOF/TOF tandem mass spectrometer.* Anal. Chem., 2000. **72**: p. 552-558.
230. Fuerst, T.R., et al., *Eukaryotic transient-expression system based on recombinant vaccinia virus that synthesizes bacteriophage T7 RNA polymerase.* Proc Natl Acad Sci U S A, 1986. **83**(21): p. 8122-6.
231. Estes, M.K., et al., *Simian rotavirus SA11 replication in cell cultures.* J Virol, 1979. **31**(3): p. 810-5.
232. Gray J, D.U., *Rotaviruses. Methods and Protocols.* 2000.
233. Ward, G.A., et al., *Stringent chemical and thermal regulation of recombinant gene expression by vaccinia virus vectors in mammalian cells.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 6773-7.
234. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(259): p. 680-5.
235. Perkins, D.N., D. J. C. Pappin, D. M. Creasy, and J. S. Cottrell, *"Probability-based protein identification by searching sequence databases using mass spectrometry data."* Electrophoresis, 1999. **20**: p. 3551-3567.
236. Lobley, A., Whitmore, L., and Wallace, B. A., *DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra.* Bioinformatics, 2002. **18**: p. 211-2.
237. Szekelt, Z., Zakhariev, S., Guarnaccia, C., Antcheva, N. & Pongor, S., Tetrahedron Lett., 1999. **40**: p. 4439–4442.
238. Marin O, B.V., Boschetti M, Meggio F, Allende CC, Allende JE, Pinna LA., *Structural features underlying the multisite phosphorylation of the A domain of the NF-AT4 transcription factor by protein kinase CK1.* Biochemistry, 2002. **41**(2): p. 618-27.

239. Marin, O., Bustos, V. H., Cesaro, L., Meggio, F., Pagano, M. A., Antonelli, M., Allende, C. C., Pinna, L. A. & Allende, J. E., *A noncanonical sequence phosphorylated by casein kinase 1 in beta-catenin may play a role in casein kinase 1 targeting of important signaling proteins.* Proc. Natl. Acad. Sci. USA, 2003. **100**): p. 10193-10200.

240. Pulgar, V., et al., *Optimal sequences for non-phosphate-directed phosphorylation by protein kinase CK1 (casein kinase-1)--a re-evaluation.* Eur J Biochem, 1999. **260**(2): p. 520-6.

241. Blackhall, J., et al., *Serine protein kinase activity associated with rotavirus phosphoprotein NSP5.* J Virol, 1997. **71**(1): p. 138-44.

242. Campagna M, B.M., Arnoldi F, Desselberger U, Allende JE, Burrone OR. and, *Impaired hyperphosphorylation of rotavirus NSP5 in cells depleted of casein kinase 1alpha is associated with the formation of viroplasms with altered morphology and a moderate decrease in virus replication.* J Gen Virol., 2007. **88**: p. 2800-10.

243. Frankel A, Y., N Lee J, Branscombe T., Clarke S, Bedford MT, *The novel human protein arginine N-methyltransferase PRMT6 is a nuclear enzyme displaying unique substrate specificity.* J. Biol. Chem., 2002. **277**: p. 3537–3543.

244. Aletta JM, C.T., Ettinger MJ, *Protein methylation: a signal event in post-translational modification.* Trends in Biochemical Sciences, 1998. **23**: p. 89-91.

245. Mears WE, R.S., *The RGG box motif of the herpes simplex virus ICP27 protein mediates an RNA-binding activity and determines in vivo methylation.* Journal of Virology, 1996. **70**: p. 7445-7453.

246. Cheung, B.a.A., *Acetylation and chromosomal functions.* Curr.Opin. Cell Biol., 2000. **12**: p. 326-333.

247. Geisow, M.J., *Both bane and blessinginclusion bodies.* Trends Biotechnol., 1991. **9**: p. 368-369.

248. Valax, P.a.G., G., *Molecular characterization of beta-lactamase inclusion bodies produced in Escherichia coli.* Biotechnol. Progr., 1993. **9**: p. 539-547.

249. Zhang, Y., Olsen, D.R., Nguyen, K.B., Olsen, P.S., Rhodes, E.T. and Mascarenhas, D., *Expression of eukaryotic proteins in soluble form in Escherichia coli.* Protein Expr. Purif., 1998. **12**: p. 159-165.

250. Murby, M., Uhlén, M. and Ståhl, S., *Upstream strategies to minimize proteolytic degratation upon recombinant pruduction in Escherichia coli.* Protein Expr. Purif., 1996. **7**: p. 129-136.

251. Holzinger A, P.K., Weaver TE., *Single-step purification/solubilization of recombinant proteins: application to surfactant protein B.* Biotechniques, 1996. **20**(5): p. 804-6, 808.

252. Amersham, *Rapid and efficient purification and refolding of a (His)6-tagged recombinant protein produced in E. coli as inclusion bodies.* 1999-09. **18**: p. 1134-37.

253. Colangeli R, H.A., Williams AM, Manca C, Chan J, Lyashchenko K, Gennaro ML., *Three-step purification of lipopolysaccharide-free, polyhistidine-tagged recombinant antigens of Mycobacterium tuberculosis.* J Chromatogr B Biomed Sci Appl., 1998. **142**(2): p. 23-35.

254. Porath, J., *Immobilized metal ion affinity chromatography.* Protein Expr. Purif., 1992. **3**: p. 263-281.

255. Sreerama, N., and Woody, R. W., *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.* Anal. Biochem., 2000. **287**: p. 252-60.

256. Whitmore L., W.B.A., *DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data.* Nucleic Acids Research, 2004. **32**: p. 668-673.

257. Blackhall, J., A. Fuentes, and G. Magnusson, *Genetic stability of a porcine rotavirus RNA segment during repeated plaque isolation.* Virology, 1996. **225**(1): p. 181-90.

258. Blackhall, J., et al., *Analysis of rotavirus nonstructural protein NSP5 phosphorylation.* J Virol, 1998. **72**(8): p. 6398-405.

259. Jabbur, J.R., P. Huang, and W. Zhang, *Enhancement of the antiproliferative function of p53 by phosphorylation at serine 20: an inference from site-directed mutagenesis studies.* Int J Mol Med, 2001. **7**(2): p. 163-8.

260. Lefebvre T, P.S., Guérardel C, Deltour S, Martin-Soudant N, Slomianny MC, Michalski JC, Leprince D., *The tumor suppressor HIC1 (hypermethylated in cancer 1) is O-GlcNAc glycosylated.* Eur J Biochem, 2004. **271**: p. 3843-54.

261. Olumee, Z., M. Sadeghi, X. Tang, and A. Vertes, *Amino Acid Composition and Wavelength Effects in Matrix-Assisted Laser Desrption/Ionisation.* Rapid Commun. Mass Spectrom, 1995. **9**: p. 744-752.

262. Pollastri G., M.A., *Porter: a new, accurate server for protein secondary structure prediction.* Bioinformatics., 2004. **21**: p. 1719-1720.

263. Romero P, O.Z., Li X, Garner EC, Brown CJ, Dunker AK., *Sequence complexity of disordered protein.* Proteins, 2001. **42**: p. 38-48.

264. Dyson MR, S.S., Vincent KJ, Perera RL, McCafferty J. and, *Production of soluble mammalian proteins in Escherichia coli:identification of protein features that correlate with successful expression.* BMC Biotechnol, 2004. **4**: p. 32.

265. Ferron F, L.S., Canard B, Karlin D, *A Practical Overview of Protein Disorder Prediction Methods.* PROTEINS: Structure, Function, and Bioinformatics, 2006. **65**: p. 1-14.

266. Friedberg I, J.L., Ye Y, Godzik A., *The interplay of fold recognition and experimental structure determination in structural genomics.* Curr Opin Struct Biol, 2004. **14**: p. 307-312.

Acknowledgements

It is very hard to give voice to my feelings. Almost five years is not a blink, it was a very special time of my life and will remain under my skin forever.

First of all I would like to thank my supervisor, Oscar, for his tireless enthusiasm and wide look on the research. I appreciate his guidance during the years of my PhD project, his patience and huge help in correction of this thesis. I am very grateful for teaching me how to be tough enough to keep reaching one's aims. It was valuable lesson of life that will profit for my future.

I would like to express my respect and sympathy to Doriano. Without his knowledge, help and availability this thesis could not have its shape.

I want to thank the people from the Protein Structure Group for providing me the FPLC AKTA system and, especially to Corrado, for his valuable hints.

My special thanks to the Rotagirls. To Cathy, whose part of work was included in this dissertation, for her help and tolerance. I am indebted to Michela for her friendship and the fact that she was always ready to serve me with her priceless and full of common sense suggestions, as well as for her tasteful style of dressing that brighten grey colors of the lab. To always cheerful Fra, who never said no to my 'cellular extracts needs', for opening and sharing her world with foreigners, including me.

I would like to thank all the people from the Immunology lab, present during 'my period', for outstanding atmosphere at work. Marco, for never saying 'I don't have time' and for answering my bothering questions. Roby for her energy, Elisa for...well, I don't know, but anyway;) And of course my immerse thanks to Mirza for his friendship and company during coffee-cigarette breaks, often full of discussions. I want to mention, that many of my coworkers at ICGEB became my friends and the time of my PhD studies in beautiful but quite sad Trieste would not be the same without people like Mohamed. Thank you guys!

Finally, a great thanks to all the ICBEB component for creating a fantastic atmosphere, that allowed me to feel there almost like at home.