Struktura proteina srži hepatitis B virusa i njegove interakcije s proteinima virusne ovojnice

Pritišanac, Iva

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO - MATEMATIČKI FAKULTET

BIOLOŠKI ODSJEK

STRUKTURA PROTEINA SRŽI HEPATITIS B VIRUSA I NJEGOVE INTERAKCIJE S PROTEINIMA VIRUSNE OVOJNICE

STRUCTURE OF THE HEPATITIS B VIRUS CORE PROTEIN AND ITS INTERACTIONS WITH THE ENVELOPE PROTEINS

SEMINARSKI RAD

Iva Pritišanac

Preddiplomski studij molekularne biologije

(Undergraduate Study Program of Molecular Biology)

Mentor : Prof. dr. sc. Mirna Ćurković-Perica

Zagreb, 2011.

UNIVERSITY OF ZAGREB

FACULTY OF SCIENCE

DEPARTMENT OF BIOLOGY

STRUCTURE OF THE HEPATITIS B VIRUS CORE PROTEIN AND ITS INTERACTIONS WITH THE ENVELOPE PROTEINS

FINAL SEMINAR

Iva Pritišanac

Undergraduate Study Program of Molecular Biology

Mentor : Prof. dr. sc. Mirna Ćurković-Perica

Zagreb, 2011.

Contents:

1. The introduction:

The hepatitis B virus is a small, human, hepatotropic, enveloped virus, the prototype member of the *Orthohepadnavirus* genus, *Hepadnaviridae* family. It was discovered and first described around forty years ago (Blumberg et al. 1960, Dane et al. 1970) and still represents severe danger to the human health as it is highly infective and in addition increases the risk of liver cancer and liver cirrhosis development in chronic carriers, estimated to be 350 million people worldwide (Perz et al. 2006). Per year, approximately 1 million people worldwide die due to the HBV infection and subsequent liver cancer development (Patient et al. 2009). These findings account for the current research focus set on the understanding of the structure and intracellular morphogenesis of the HBV in order to detect the effective targets for new antiviral strategies.

The HBV infective particle, also known as Dane particle, is spherical virion, 42 nm in diameter, its icosahedral core of either T=4 (34 nm diameter) or T=3 (30 nm diameter) symmetry. The core is composed of the 22kD HBV core protein (HBc), 183-185 amino acid residues long, its exact length determined by the viral genotype, amino acid composition and sequence showing great conservation between virion isolates (Chain and Myers 2005). The primary structure reveals two distinct protein regions; the N-terminal region, essential for the monomer protein fold and capsid assembly and the C-terminal basic region, dispensable for core assembly but involved in the pre-genomic RNA/reverse transcriptase encapsidation. The atomic level crystal structure of the T=4 (C terminally truncated, RNA containing) viral core particle has revealed the structural characteristic of the core protein at 3,3Å resolution (Wynee et al. 1990). The fold the monomer protein takes in the HBV core is unusual largely helical, the structural feature not typical for other icosahedral viral species core particles. The process essential for the assembly of functional HBV core is the cytosolic homodimerization of two monomer core proteins. The core protein dimer is the only stable intermediate in the core assembly and thus its fundamental building unit (Zheng et al. 1992, Zhou and Standring 1992). The details of the further core assembly processes, as the dimer's oligomerization, are still largely unclear (Patient et al. 2009).

Incorporated into the mature core is the HBV genome, the circular, partially ds DNA, approximately 3,2 kb long, covalently linked to the, viral encoded, reverse transcriptase (RT). The partially ds DNA/RT complex and the assembled mature core hosting it, constitute the

HBV nucleocapsid particle. In the infected cell, this particle undergoes the morphogenesis process by which it acquires the envelope composed of the cellular lipid bilayer with incorporated, transmembrane HBV envelope proteins. The HBV genome encodes three envelope proteins by a single open reading frame (the ORF-E) the small(S), the medium (M) and the large (L) envelope protein and these have the major role in the HBV morphogenesis and thus generation of the mature, infective HBV particles.

The molecular details of the events which follow the HBV/ hepatocyte encounter and lead to cell infection have been determined to some extent. However, the exact mechanism of the first step, the HBV entry in the competent cells is still unclear and under investigation. Once inside the cell, the HBV core particle disassembles (the details of the process largely unknown) and the viral partially ds DNA genome released. The later is directed to the cell nucleus (Rabe et al. 2003) where the cellular enzymatic machinery converts it to covalently closed, circular DNA (ccc DNA). The HBV ccc DNA is further transcribed, in the hepatocyte nucleus, in the pre-genomic RNA (pgRNA), 3,5 kb in length. The different other subgenomic viral RNA transcript are also generated at the stage and encode the viral proteins involved in the replication cycle (Beck and Nassal, 2002). The pg RNA and viral RNA transcripts are transported to the cytoplasm were the transcripts are translated and the set of HBV proteins generated.

One pgRNA is in the cytosol covalently linked with translated viral RT and the complex is subsequently incorporated into the assembling core particle. Thus obtained particle is the immature HBV core particle, in which, on yet undetermined molecular trigger, the untypical reverse transcription action of the RT generates the partially ds mature viral DNA (Beck and Nassal, 2002). This event is known to be the trigger for the second crucial point of the HBV viral maturation and that is the generation of the mature core particle, capable to undergo the envelopment process. Namely, the changes in the structural characteristics of the core nucleic acid content influence its interactions with the protein dimers, which constitute the HBV core. Thus reinforced conformational changes in the core dimers lead to the proper-time opening of the hydrophobic pockets on the core dimer outside surfaces. The hydrophobic pocket amino acid chains spatial distribution is perfectly balanced for the high affinity interaction of the HBV core dimers with the envelope proteins, thus enabling the envelopment process.

The goal of this seminar is to highlight the structure of the HBV core monomer protein at several levels- the primary, secondary and tertiary structure (the monomer fold). The

special attention is placed on those structural features with conserved, functional role, such as the acquiring of the quaternary protein structure, observed upon the dimer assembly. Finally, the monomer and dimer structures will be observed in context of their dynamical changes induced by the reverse transcription of viral pg RNA to partially ds DNA, the crucial process for the generation of mature viral genome. The later structural changes in core proteins will be monitored for their special functional role in HBV morphogenesis, which is the enablement of the interactions of the core protein dimers with the pre-S1 region of the S envelope protein, which account for the successful HBV envelopment.

2. The HBV core protein (HBc)-primary structure and functionally important regions

In order to get the proper insight into the way the infective HBV particles are created, the HBV core particle assembly ought to be observed at the precise molecular level. The good start would be the determination of the primary sequence requirements in the HBV core protein which account for its proper intra-molecular fold, and thus its exact structural characteristics. These characteristics are responsible for the intermolecular core protein interactions which lead the formation of compact dimer, the only stable intermediate, and thus the basic structural unit for the HBV core assembly (see Chapter 5).

The HBV core protein is 22 kD, 183-185 amino acids long protein. Its exact amino acid length is determined by the viral genotype. The HBV core protein gene sequence, and hence the translated amino acid sequence, shows relatively good conservation among different genotype variants of the HBV isolates (Chain and Myers 2005) Figure 1 indicates the conservation of the large number of amino acid residues, the feature which provides the basis for the deduction on the underlying structural/functional importance of these conservations. The early experimental works on the HBV core protein used the core protein monomers, as well as the whole core particles, expressed in a broad range of prokaryotic and eukaryotic cell types, in order to determine the HBV core protein exact amino acid content and sequence (Galibert et al. 1979, Pasek et al. 1979, Yaginuma et al. 1987, Gallina et.al. 1989).

An extensive research of the HBV core protein primary structure (Birnbaum and Nassal 1990) highlighted its general and in addition some functionally important properties. There are two main, distinct regions of the HBV core protein primary amino acid sequence- the N-terminal and the C-terminal region. The N-terminal region includes the first 149-151 amino acids (genotype dependent) and is referred to as the assembly domain. The assembly domain is defined as the part of the protein sufficient to direct the self assembly of the core particles, while C-terminal region is dispensable for the function (Birnbaum and Nassal 1990). The later fact has been determined by few studies which designed the series of truncated core protein gene variants (lacking the various number of C-terminal amino acid residues), heterologously expressed them in *Escherichia coli* and examined their capsid assembly ability (Gallina et al. 1989, Zheng et al. 1992, Crowther et al. 1994, Wingfield and Stahl 1995). The results obtained were used to define and map the C-terminal region amino

acid limit for the proper HBV core particle assembly, which was confirmed to reside in the minimum of the first 140 N-terminal amino acids (Birnbaum and Nassal 1990) (Figure 2).

The requirement for this precise length might be the consequence of the structural role imposed by the leucine amino acid residue at the position 140 (Leu-140). Being packed between the phenilalanine (Phe-110) and tyrosine (Tyr-118) in the C-terminal region, Leu-140 potentially forms important part of the hydrophobic core, known to stabilize the fold of the monomer itself (Figure 6, Wynee et al. 1999) which will be further discussed in the Chapter 4. Therefore, the loss of the interactions by which the Leu-140 contributes the monomer fold, could consequently influence the conformation of the monomer loop region (128-139 amino acid residues) (see Chapter 4). This would additionally affect both the folding and the stability of the core protein monomer, as well as its ability to properly interact with other core protein units in the core structure. However, the later hypothesis needs to be further examined. It could be performed by the extensive research on the folding and interactions of the designed protein variant mutants for the 140. residual position.

The C-terminal region consists of the 34 amino acid residues, directly follows the assembly domain, and is extremely rich in Arg (R) residues (Birnbaum and Nassal 1990) (Figure 2). Though dispensable for capsid formation the truncated protein variants, lacking 34-43 C-terminal amino acids, showed to be incapable of providing the proper viral nucleic acid encapsidation (Gallina et al. 1989). Taking into account the later fact and the fact that other related mammalian hepadnavirus core proteins also express the C-terminal region extremely rich in Arg-residues with the nucleic-acid binding function, the same has been postulated for the HBV core protein C- terminus (Pasek et al. 1979, Gallina et al. 1989). The last performed extensive cryo-electron microscopy study of the full-length protein core particles containing RNA and DNA have clearly demonstrated the close interaction of the C-terminal tails with both types of the viral nucleic acid, thus finally confirming the stated hypothesis (Roseman and Berriman 2005.). Studies which followed (Bruss 2007) have demonstrated that the core protein C-terminal tails are in charge for the interactions with viral pregenome/reverse transcriptase complex during virion packing. The close interactions have been detected (see Chapter 5), and the special attention must further on be placed on studying the details of the observed conformational changes in the core proteins (Roseman and Berriman 2005.), which are connected to the reverse transcription of the HBV ssRNA to partially ds DNA and responsible for the viral maturation (Chapter 5).

2.1. The HBV core protein C-terminal region

As the function of the HBV core protein N-terminal region was very early confirmed to be crucial for core protein folding and thus proper capsid assembly (Gallina et al. 1989.), the confirmation of the exact function of the rest of the protein was, for the time, rather challenging. Therefore, the special interest in further HBV core protein research was placed on the extremely basic, Arg rich C-terminal region, also recognized as the prominent feature of the related hepadnavirus core proteins (Pasek et al. 1979.). The C-terminal amino acid sequence, rather unusual in composition with several copies of the serine-proline-arginine-arginine (SPRR) motif (Figure 2), resembles the motifs found in the histone proteins, where it accounts for the rather tight binding of the nucleic acid (Suzuki, 1989.)

In order to precisely detect the C-terminal tail function, the mutant core gene sequences have been constructed in order to avoid any perturbation by the foreign sequences which natively might be fused to the core protein (Birnbaum and Nassal 1990.) The designed mutant core protein (C) gene variants encoded complete length core proteins and variants with different length of the C-terminal region and were all successfully expressed in *E. coli.* These were then compared in their capability of capsid self-assembly, as well as the nucleic acid contents and binding preferences. From the studies the following important conclusions have been made (Birnbaum and Nassal 1990.):

1. The HBV core protein C-terminus from amino acid 144 on, and therefore, the entire Arg-rich region is dispensable in capsid assembly

2. C-terminal truncations beyond the amino acid 140 are not tolerated and do not produce significant amounts of the properly assembled viral core particles, although may still have some limited potential for core protein monomer self- assembly

3. The electrophoretic mobility of the whole length core protein particles and those with truncated C-terminal region is not markedly different

4. The presence of at least part of the highly basic core protein C-terminal region results in the bulk of non-specific nucleic acid binding in *E. coli.* The core particles constructed from the HBV full length core protein non-specifically encapsidated non viral RNA or DNA (Figure 3, Figure 4)

The later was demonstrated in the absence of correct viral genomic RNA and HBV P gene product. In further studies, P gene product was determined as the viral RNase H reverse transcriptase, covalently linked to the viral pre-genomic ss RNA in immature virions (Bruss et al. 2004.)

5. The packed nucleic acid is ssRNA (selectivity over dsDNA), which ranges in length from 100 to 3,000 nt in length

6. The full length core protein exhibits certain selectivity for binding its own mRNA rather than the other cellular RNAs

7. The core protein variants completely lacking the basic C-terminus still contain the small amounts of nucleic acid (level 20-fold lower than that for the full length protein particles) (Figure 3)

8. The presence of the Arg-rich region and therefore the nucleic acid content, accounts for the increased core particles stability

The observations of this early study (Birnbaum and Nassal 1990.) that much more RNA (regarding of the specificity) is encapsidated in the viral core in the C-terminal sequence presence, confirmed the biochemically expected affinity of poly-Arg region for nucleic acid (Pasek et al. 1979., Gallina et al. 1989.) and thus suggested the functional role of the Arg-rich C-terminus to be the nucleic acid binding and encapsidation. Later studies of the C-terminal functions determined it to be, in addition, the functionally most important part of the core protein primary structure for viral maturation (Roseman and Berriman 2005.) (See Chapter 5)

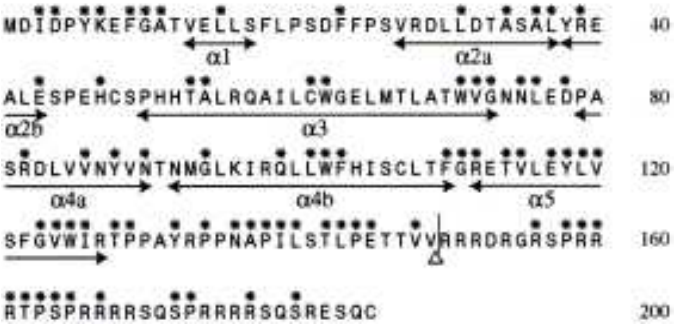

Figure 1: The amino acid sequence of the CW variant of HBV capsid, truncated at amino acid 149 (indicated by Δ). The 94 full-length human HBV capsid sequences in the SWISS-PROT and TrEMBL databases were aligned using CLUSTALW. Fully conserved residues (identical in all 94 sequences) are marked (*). (completely taken from Wynee et al. 1999)

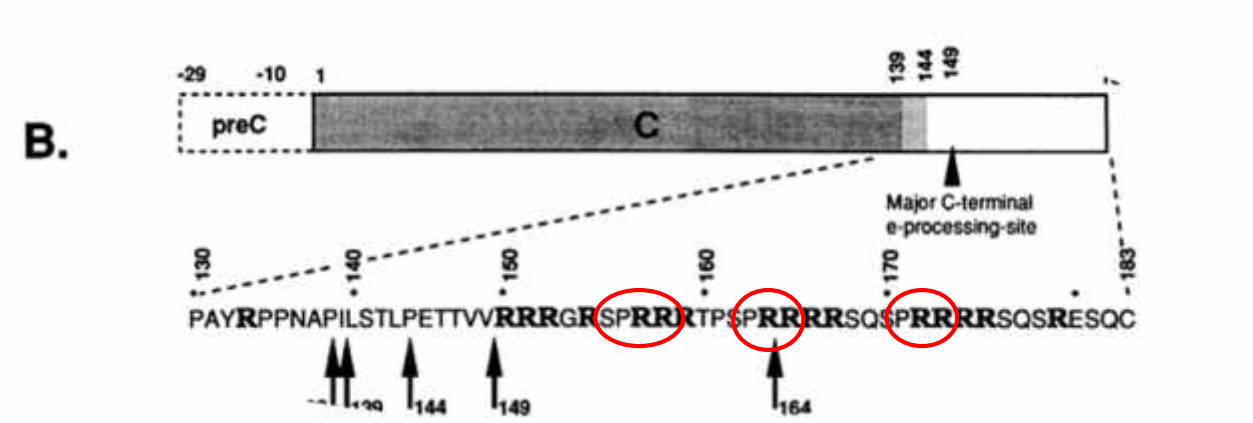

Figure 2: Numbers on top refer to amino acid position in the core protein. The dashed box represents the 29 amino acids encoded by the pre-C region, required for HBcAg secretion. The major C-terminal HBcAg processing site is indicated at amino acid position 149. The shaded area represents the essential assembly domain, the C-terminal border (lighter shading) which has been confined to the sequence between amino acid positions 140-144. The region from amino acid positions 130-183 include the Arg rich tail, shown in detail by using the single –letter amino acid code; Arg residues are in addition bolded and SPRR sequence motif red circled (completely taken from Wynee et al. 1999).

3. The HBV core protein monomer fold

The second chapter introduced the basic primary structure amino-acid sequence of the HBV core protein (Figure 1), defined the N-terminal and C-terminal regions and drove a bit more attention to some experimental observations explaining the function of the C-terminal region (2.1). However, as already introduced there the intra-cytoplasmical HBV core assembly capability completely resides in the first 140-144 amino acids of the N-terminal sequence (Birnbaum and Nassal 1990.). This fact implies that all the structural requirements for the correct monomer fold, as well as the productive intersubunit contacts, are contained in this N-terminal sequence. The C-terminal region, although as already stressed, extremely important for interactions with, and encapsidation of, the pre-genome/RT complex (Bruss et al. 2004), does not seem to play any essential role in the HBV core protein monomer fold.

The first hypothesis on the HBV core protein fold was introduced after its primary structure and the minimum length of the protein required for the proper core assembly were determined (Birnbaum and Nassal, 1990). The fact that the length of minimum 140 N-terminal amino acids provided proper fold of the viral core (thus also the proper fold of all its constituting monomers) it was suggested that the protein folds into the eight stranded β-barrel structure. The basis for the conclusion was that the β barrel folding structure of the monomer core proteins with similar length was already determined for some RNA viruses with icosahedral core. The minimal length requirement for correct folding in such structure is estimated to be 150 amino acids (Rosemann and Johnson, 1989).

However, the model of the HBV core protein fold, set after the first early cryo-electron microscopy studies of the core particles, introduced completely unexpected structural features. The early cryo-electron microscopy experiment was performed on the series of the empty HBV core particles constituted from the C-terminally truncated core protein (140 amino acids in length) heterologously expressed in *E. coli* (Böttcher et al. 1997). The results obtained described the unexpected largely helical fold of the monomer core protein which formed T=4 icosahedral HBV core particles. At the time the structure of the HIV-1 core protein (Momany et al. 1996, Gamble et al. 1997) was also determined to be largely helical, the HBV and HIV-1 core proteins were among the first proteins of icosahedral viral core described to have the dominant helical fold instead of mostly observed β-barrel structure. The conserved amino acid residues in monomer in primary amino acid non-continuing chain distribution give rise to the conserved fold motif called hydrophobic core (Figure 6). The

hydrophobic interactions which mediate the formation of such hydrophobic core constitute the major driving force in the monomer HBV core protein folding.

3.1. The structure of the HBV core protein (HBc) fold determined by the protein X-ray protein crystallography

Even more detailed structural information came with the atomic level X-ray microscopy structural analyses performed after the HBV empty core particles had been successfully crystallized (Wynee et al. 1999). It not only confirmed the excellent agreement with cryo-electron microscopic-suggested helical monomer fold, but in addition proved assignment of the core protein amino acid residues as remarkably accurate (matching the X-ray atomic level structure by 97%) (Wynee et al. 1999). The initial cryo-EM map model of the helical monomer fold was used to generate the initial resolution phases to 8 Å and then extended to better X-ray resolution of 3.3 Å. The X-ray determined 3.3 Å resolution provides remarkable atomic-level details of all parts of the HBV core protein secondary and tertiary structure, except for the C-terminal tail region (amino acids 149-181/3) (as the crystallized particles were empty core particles of C-terminally truncated core protein). In order to fulfill the deficiencies in the information on the contribution of the C-terminal tail region to the monomer fold, the region was observed by the later taken cryo-electron microscopy study of the mature, partially ds DNA containing, core particles (Roseman and Berriman 2005).

The N-terminal region (1-12 amino acid residues) adopts an irregular structure of extended amino acid chain, which is followed by the short α1 helix (residues 13-17). The next is 10 amino acid long extended chain after which the longer α2 helix follows (residues 27-43). The α2 helix is kinked after the residue 37 - the conserved leucine (L) residue (Figure 1). The chain of amino acids 50-73 constitutes the helix α3, which is joined to helix α4 (residues 79-110) by a very short amino acid chain (residues 73-79). The α4 helix is also kinked between the residues 90-92 (the kink divides it to α4 a and α4 b part), the feature which is very significant in functional aspect, and will therefore be later described (see Chapters 4 and 5). The C-terminal region follows immediately after α4 helix and folds into the helix α5 (residues 112-127). The helices α4 and α5 are thus divided solely by one amino acid residue, the conserved glycine residue (G-111) which, as well as the kink in α4 helix, has conserved functional role (see Chapter 5). After helix α5 an irregular loop (residues 128-136) rich in conserved proline (P) amino acids follows. The protein sequence finishes with an extended

strand (residues 137-149) with the rest of the C-terminal tail not observable due to the truncation at the residue 149 for the study (Figure 3).

The tertiary structure of the HBV core protein is constituted from distinct secondary structural parts (the α helices and an irregular loop). The monomer fold is dominated by the long α helical hairpin, formed by the interactions of the amino acid side chains from the helices α3 and α4. However, due to the kink in between residues 90-92 in α4 helix, the lower part of the α4 helix (α4 b) is splayed away from the α3 helix and therefore contributes less in the formation of the hairpin. Nevertheless this feature accounts for the functionally significant conformational properties of the hairpin important in viral maturation (see Chapter 5). The irregular N-terminal region, helices α1 and α2, together with the connection loops between them, surround three sides of the dominant α helical hairpin. The results obtained by the observations of the full length core protein from (partially ds DNA containing) mature viral core particles (Roseman and Berriman 2005) suggest that, as biochemical and structural *in vitro* analyses indicated- rather flexible C-terminal tail- forms observable interactions with the α5 helix and thus additionally contributes to the stabilization of the monomer fold.

Although the secondary and tertiary structures of the HBV core protein have been elucidated to great detail, there is still lack of information on how exactly is the observed structure generated. The folding of the HBV core protein monomer has been under investigated, as the assembly is mostly observed solely in the context of the dimer and capsid formation (Chapters 4 and 5). The extensive researches on the details of capsid assembly are result of the attempt ions to detect the targets for its inhibition and thus provide the effective antiviral strategy. The details of the folding of the HBV core protein monomer itself, its rate, dynamics and even general details of the conformations the protein can adopt in solution, unfortunately, seem to be less attractive. These dynamical properties of the HBV core protein in the solution could be highlighted by converting the protein to the protein nuclear magnetic resonance technique (e.g. the advanced protein NMR).

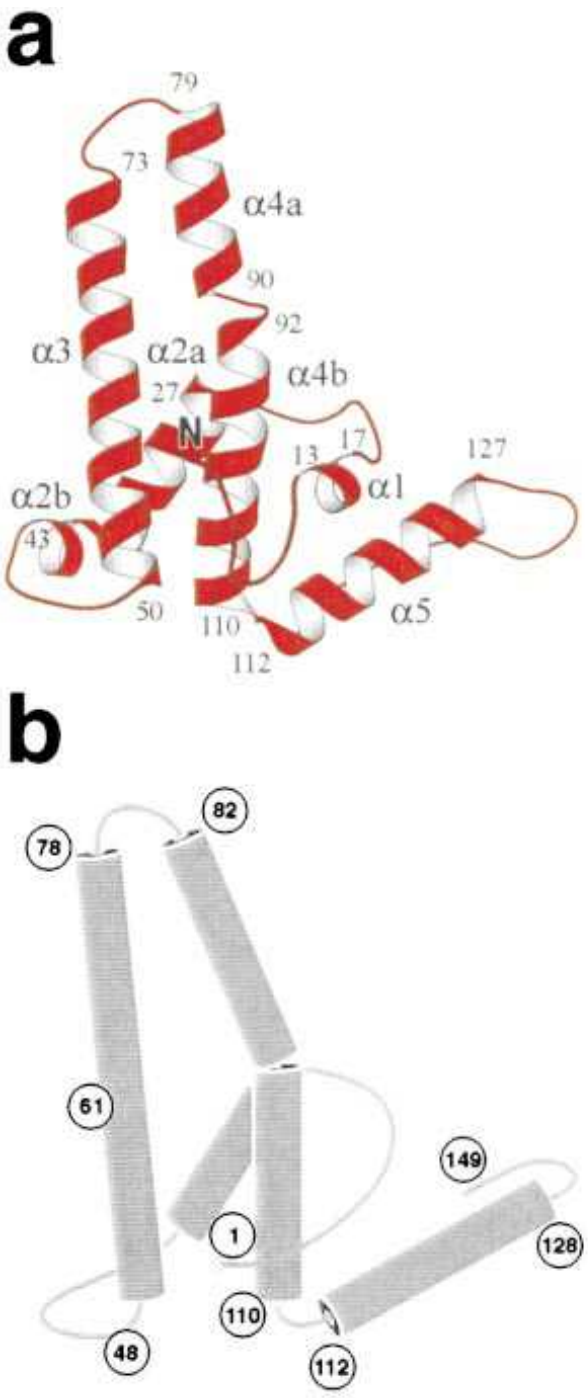

Figure 3: The fold of the HBV core protein monomer determined by the combination of computed maps constructed upon the cryo-EM and X-ray crystallography studies (completely taken from Wynee et al. 1999)

4. The dimer of the two HBV core protein monomers-the main core structural unit

The early biochemical *in vitro* studies of the HBV core monomer protein, defined the dimer association, acquired by the cytoplasmic homodimerization of the two core monomers, as the only stable intermediate in the HBV core assembly and thus its fundamental structural unit (Zheng et al. 1992, Zhou and Standring 1992). The fold of the HBV core protein, described in the last chapter, has introduced some structural characteristics, which are here going to be further explained as these are very important for the proteins functional role. The chapter 3 has given certain details of the HBV core protein monomer fold, defining α helical hairpin as its major tertiary structural feature. Formed by the interactions of the helices α3 and α4, the hairpin presents the central protein fold, around which the rest of the protein's secondary structural components (N-terminal region, helices α1 and α2, and the connection loops between them) are folded.

Early cryo-electron microscopy observations (Böttcher et al., 1997) of the HBV core particles detected the structure of the compact dimer as the quaternary protein structure formed by the adjacent core protein monomers. This dimer association accounted for the observed characteristic spikes decorating the HBV core surface (Figure 4) (Ceres et al, 2004). The observed spikes were defined as the four-helix bundle, the association of the two α helical hairpins, one from each of the two associated monomers. Each monomer contributed the observed structure with its two helices (α3 and α4) (Böttcher et al. 1997). The later taken X-ray crystallography diffraction analyses (Wynee et al.1999) have confirmed the existence of the four-helix bundle structure and in addition provided more detailed insight into its structural characteristics. Nevertheless, the four helix bundle presents only the protruding part (the spike) of the dimer formed (Figure 4). The whole dimer resembles the "pickaxe", four helix bundle corresponding the shaft and the basal region corresponding the head (Wynee et al.1999) (Figure 5).

4.1. The HBV core protein monomer structural features responsible for the compact dimer formation

As already introduced in the Chapter 2, the structural characteristics of the monomer fold and its amino acid side chain distribution is responsible for monomer's interactions which give rise to compact, stable dimeric quaternary structure. The major structural feature of the monomer fold, the helical hairpin is the first to be considered in this aspect. The amino acid content of the helices α3 and α4, (Figure 1) is diverse. There is an approximately equal

amount of the charged and hydrophilic vs. hydrophobic amino acids, which accounts for the largely amphipatic helical hairpin formed upon their folding into the α3 and α4 helices. The hydrophobic faces of the two monomer hairpins associate on dimer formation, resulting in the 63% hydrophobic surface (Wynee et al. 1999). Thus formed largely hydrophobic dimer interface corresponds to the 2000 $Å^2$ surface area. These hydrophobic interactions between the two monomers present the major driving force for the dimer formation (Ceres and Zlotnick, 2002) (see the subchapter 4.2). Although referred to as the four-helix bundle, an important feature must be taken into account. The four helices do not contribute equally to the contacts which give rise to dimer stability. Only two of the four helices, the α3 helices of each monomer form the extensive interactions, while the α4 helices demonstrate almost no contact interactions except on the tip of the bundle (Wynee et al. 1999).

Furthermore, the interactions between the two core protein monomers are not constricted only to the interface between the two amphipatic helical hairpins. The rest of the monomer fold also accounts for the extensive monomer-monomer interactions, which give rise to the head of the observed dimer "pickaxe". The rest of the monomer fold, surrounding the dominant helical hairpin, constitutes the other, for dimer formation important, structural element, defined as the hydrophobic core. The core includes the following residues: Tyr-6, Phe-9, Leu-15, Leu-16, Phe-18, Leu-19, Phe-23, Phe-24, Trp-102, Phe-103, Phe-110, Val-115, Tyr-118, Leu-119, Phe-122, Trp-125 and Leu-140 (Wynee et al.1999) (Figure 6).

The other important monomer residues responsible for forming the interactions in the dimer interface are the conserved cysteine residues (Cys-61) in the α3 helices of the two monomers folds (Figure 7). Although between these two residues there is always observable disulfide bond at the orthogonal view of the dimer interface (Figure 7), it has been proved that this interaction does not constitute an essential component of the dimer assembly. Mutations on this residue, leading to its change to either serine (Ser) or alanine (Ala), did not influence the stabile dimer and capsid assembly (Nassal et al., 1992; Zheng et al. 1992; Zlotnick et al. 1996). This fact confirms the hypothesis of the dominance that hydrophobic interactions have in driving the dimer assembly. Although several *in vitro* studies of the core assembly from dimers of purified, *E.coli* expressed, HBV core proteins were taken, the details of the oligomerization core assembly processes are, unfortunately, still largely unclear(Patient et al. 2009).

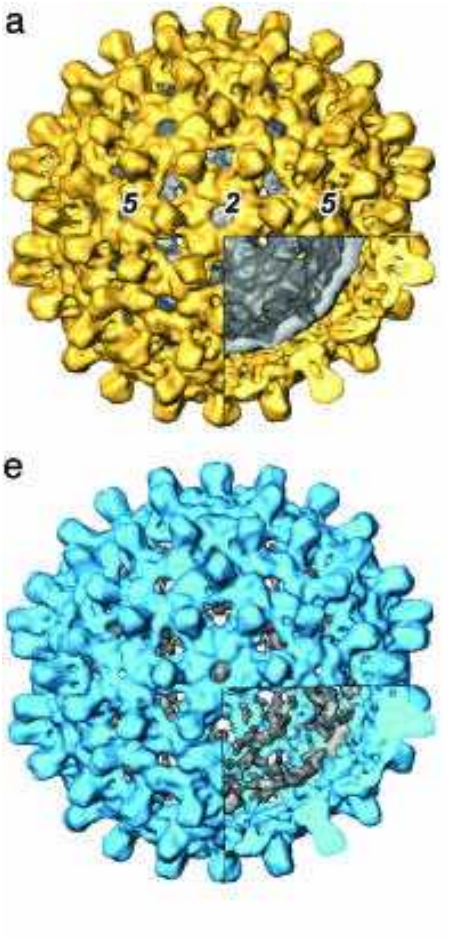

Figure 4: The spikes observed on the maps of the RNA (yellow) and DNA (blue) core containing particles computed from the X-ray diffraction image of the HBV core particles (completely taken from Roseman and Berriman 2005)

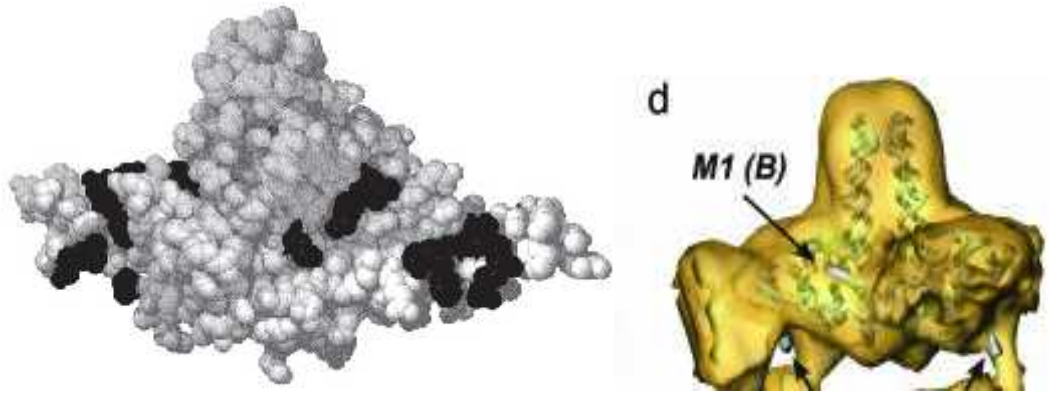

Figure 5: The HBV core protein dimer formed by the monomer proteins truncated at C-terminal region (left) and full-lenght monomer proteins (right) (completely taken from Bruss 2007 (left) and Roseman and Berriman 2005 (right))

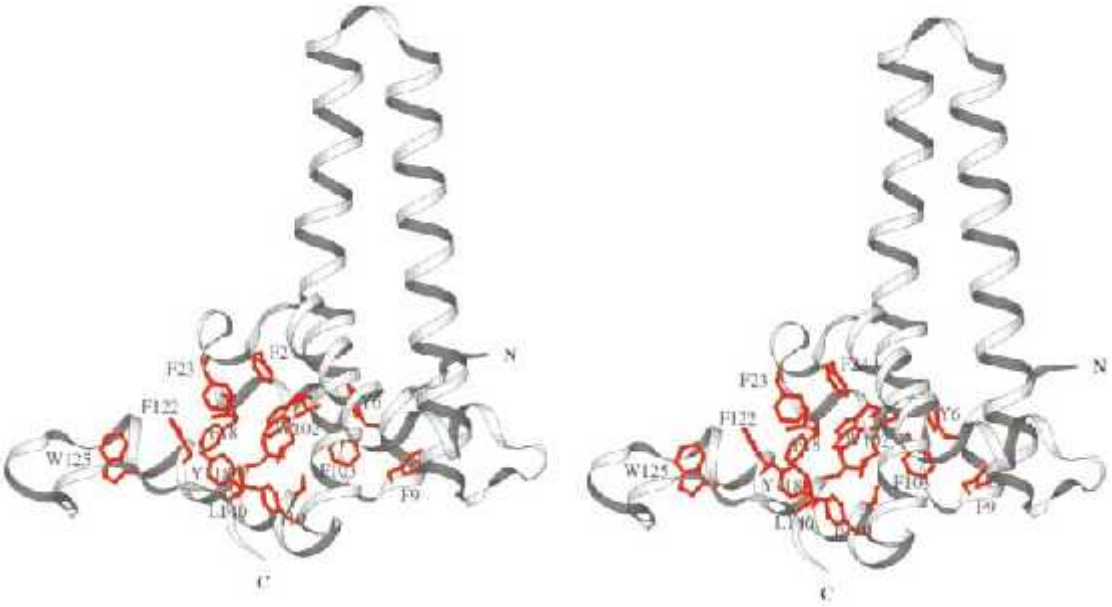

Figure 6: The monomer fold of the C-terminally truncated HBV core protein with some residues of the hydrophobic core indicated in red colour (completely taken from Wynee et al. 1999).

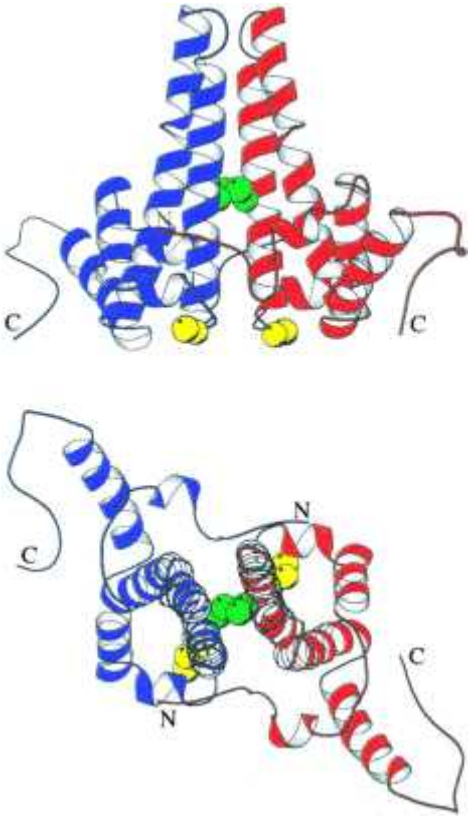

Figure 7: The two orthogonal views on the HBV capsid protein dimer, two monomer Cys-61 residues forming the disulfide bridge at the interface are shown in green (completely taken from Wynee et al.1999).

5. The interactions of the HBV core protein with the viral nucleic acid and the pre-S1 region of the S envelope protein

In the previous chapters the general structural characteristics of the HBV core protein have been introduced and the special stress placed on its fold (Chapter 3) and those features of it which allow special functional roles, such as the features accounting for the dimer formation (Chapters 2 and 4). However, although the HBV core protein interaction with other core protein monomers is essential for the proper dimer formation and thus core assembly, it is not the only interaction this protein is capable of demonstrating. The interactions of the HBV core protein monomer with the viral nucleic acid and pre-S1 region of the S envelope protein are also as crucial for the successful HBV life cycle as are its inter-monomer and inter-dimer interactions previously described.

The fact that the HBV core protein interacts with the nucleic acids was suggested early after the determination of its primary structure with the Arg-rich, positively charged C-terminal region (Gallina et al. 1989).The later was further confirmed by the cryo-electron microscopy and X-ray diffraction crystallography studies of the full length HBV core protein particles enclosing the viral and variety of non-viral nucleic acids (Birnbaum and Nassal, 1990). However, the precise details of the differences in structure of the core protein interacting with different kinds of nucleic acid were mostly neglected, probably because for the time, available techniques were not able to detect them.

The reason for the extensive research on these subject (Roseman and Berriman, 2005) came from already earlier observed, interesting fact that the envelopment of the composed viral core particles does not occur unless these contain the partially ds DNA genome in the HBV protein core (Summers and Mason, 1982). It was the basis for setting the hypothesis of the „envelopmental signal"(Summers and Mason, 1982) which must be transported from the core interior, where the newly synthesized DNA is located, to the core surface, which is in charge for the interactions with the membrane enclosed surface proteins. The remarkable research performed by Roseman et al. 2005, highlighted the structural differences in the HBV core proteins interacting with RNA vs. partially ds DNA. These observations suggest the attractive mechanism of the „envelopment signal", which unites the idea that the proper-space and proper-time interactions of three HBV molecular species (the viral nucleic acid, the core protein and the viral envelope protein) account for the successful viral maturation.

5.1. The structural differences between the HBV RNA containing core particles and HBV DNA containing core particles

The differences in the structure of the HBV core protein interacting with RNA or partially ds DNA were observed by cryo-electron microscopy studies (Roseman and Berriman 2005). The obtained computed maps of the RNA containing core particles (full-length core protein, expressed in *Escherichia coli*) and viral partially ds DNA containing core particle (authentic DNA cores from isolated virions) were compared. The most obvious large differences in structure of the cores were the shapes of the core protruding spikes. The RNA core spikes were rather bulbous or knob shaped, wider toward the outer end (observed on the

core profile map) while the DNA core profile map showed more uniform and regular spikes (Figure 8).

As described in more detail in the Chapter 4, the spikes protruding from the viral core particles are the four-helix bundles of the interacting core monomer hairpins (constituted from the monomer helices α3 and α4) and joined to form the core dimer. Therefore, the observed large structural differences between RNA core and DNA core spikes actually result from the differences in the RNA core and DNA core dimer structures, and consequently the precise RNA core and DNA core monomer fold level, respectively. The observed differences in the monomer folds of the DNA core dimer from the features observed for the folds of the monomers in the RNA core dimers were as follows:

The α4 helix in monomer proteins of the RNA containing cores has got a kink in between the residues 90-92 (Wynee et al. 1999) while:

1. In the observed DNA core dimer, one of the monomers (monomer A) showed the straighter (un-kinked) α4 helix, which has, due to the direct interaction with the α3 helix of the other monomer (monomer B) at the hairpins interface, induced its conformational change. This resulted in the change of the α3 helical orientation (the helical turn)

2. As all four α helices at the spike (the hairpins interface-four helix bundle) are in close contact, the conformational changes in either helix, influences the conformations of all the others. As the result the conformational changes of the helices α3 of the A monomer and α4 of the B monomer are also observable

3. The latter changes differed from those in the monomer partner helices, and thus the helix α4 of the monomer B is un-kinked and straightened more apparent toward the cores inside than at the four-bundle tip, which is opposite of the observed straightening of the helix α4 of monomer A (more apparent toward the spike tip)

5.2. The model for the induction of the conformational changes in the HBV core protein upon the reverse transcription of the viral RNA pre-genome to viral partially ds DNA genome

The described comparisons of monomer folds in DNA and RNA core dimers have confirmed that observed spike differences origin from the local differences in the protein chains conformation, namely the un-kinked and thus straighten α4 helix in one monomer of the DNA containing core particles. This has provided the basis for the construction of the model which would describe the switch from the immature HBV pre-genome RNA containing viral core particles to mature HBV DNA containing core particles, the dynamical process that is still impossible to observe experimentally.

Such model (Roseman and Berriman 2005) is based on the observed differences in structural characteristics of the HBV RNA and DNA containing core particles and attempts to describe the processes which mediate the switch from the immature to mature HBV particles. The model implies that the C-terminal tails of the HBV core proteins are the first protein regions to be influenced by the introduction of the second nucleic acid strand. Namely, the reverse transcription which occurs in the assembled immature HBV core particles leads to the synthesis of the partially ds DNA from originally incorporated ss RNA. Such change in the structure of the nucleic acid with which it closely interacts could induce the conformational change of the C-terminal protein tails in the inside of the core particles. As the C terminal tail stabilizes the monomer fold by the interaction with the helix α5 (see Chapter 3, Figure 3) the influence of this change could propagate to the outside of the core through the subsequent conformational change in the helix α5. As introduced in the chapter 3, the close link between the basis of the helix α4 and helix α5 (the solely highly conserved Gly-111 residue) could account for these “conformational change wave” to be readily transferred to the tip of the four helix-bundle through the helix α4 which run out to this tip. The conformational change in the helix 4 could account for its un-kinked conformation, the structural feature observed for the monomer proteins of the HBV DNA core particles. Furthermore, the changes in position of α4 helix would necessarily changed the turn of the helix 3 from the adjacent monomer with which it is associated in dimer (see above). The following would again influence the α3 and α4 helices of the partner monomer (in the dimer association in the viral core) and thus the combined effects of the movements of these helices in the four helix-bundle could account for the change of the spike shape. The stated hypotheses provide the attractive model for explanation of how the DNA genome synthesis accounts for the changes of the structural features of the viral protein core and thus guides the viral maturation.

5.3. Interactions of the mature HBV core protein with the pre-S1 of the S envelope protein

However, one important fact has not yet been explained. How are all these conformational changes in the HBV core particle, connected to the final stage of virus maturation, namely its envelopment and budding?

The changes in the HBV core protein dimer spikes described in the last chapter, account not just for the observed morphological changes of the viral core spikes, but more importantly influence the conformation of the entire four helix-bundle region. The region thus obtains distinct structural features which are essential for its function. Namely, the conformational changes accompanying the α3 and α4 helical movement in the four-helix bundle, upon the maturing of the core particle, influence the other regions of the core protein dimer. On the either side of the dimer spikes, at the interface of the dimer subunits, the new specific amino acid positional arrangement accounts for the formation of the hydrophobic pocket in the spike (Figure 9). This pocket contains the distinct spatial arrangement of the following amino acid residues: Proline-5, Leucine-60, Leucine-95, Lysine-96 and Proline-130 (P5, L60, K96, and P130) all of which have proved to have the important functional role in the HBV mature particle envelopment. Namely, the mutant protein variants with these amino acid residues changed, have demonstrated various defective secretion phenotypes (Ning and Shih, 2004). The reason for these observations is the fact that each of these amino acid residues has an important role in the proper folding of the hydrophobic pocket and thus its affinity for binding the pre-S1 region of the S surface protein.

These mutational studies, together with *in vitro* biochemical analyses, have determined the hydrophobic pocket region of the mature core dimer spike to be the region which allows for the productive interaction of the HBV core protein and pre-S1 region of the HBV S envelope protein (HBs). This suggests a model by which the conformational changes in the HBV core protein, associated with viral ds DNA synthesis, account for the formation of the hydrophobic pocket which then provides the productive interaction with the pre-S1 region of the S envelope protein thus enabling the envelopment of only those HBV core particles which contain the mature partially ds DNA viral genome (Figure 10).

Figure 8: The superposition of the central sections of the HBV DNA core computed map (blue) and RNA core computed map (yellow). Pointer shows the differences in the spike size and shape (completely taken from Roseman and Berriman 2005)

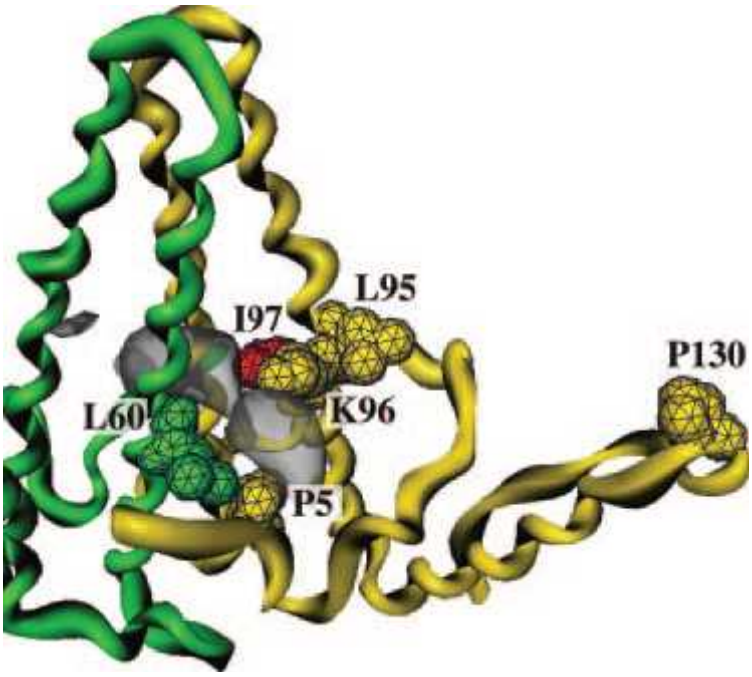

Figure 9: The hydrophobic pocket in the mature HBV core protein dimer with its functionally important amino acid residues indicated (completely taken from Roseman and Berriman 2005)

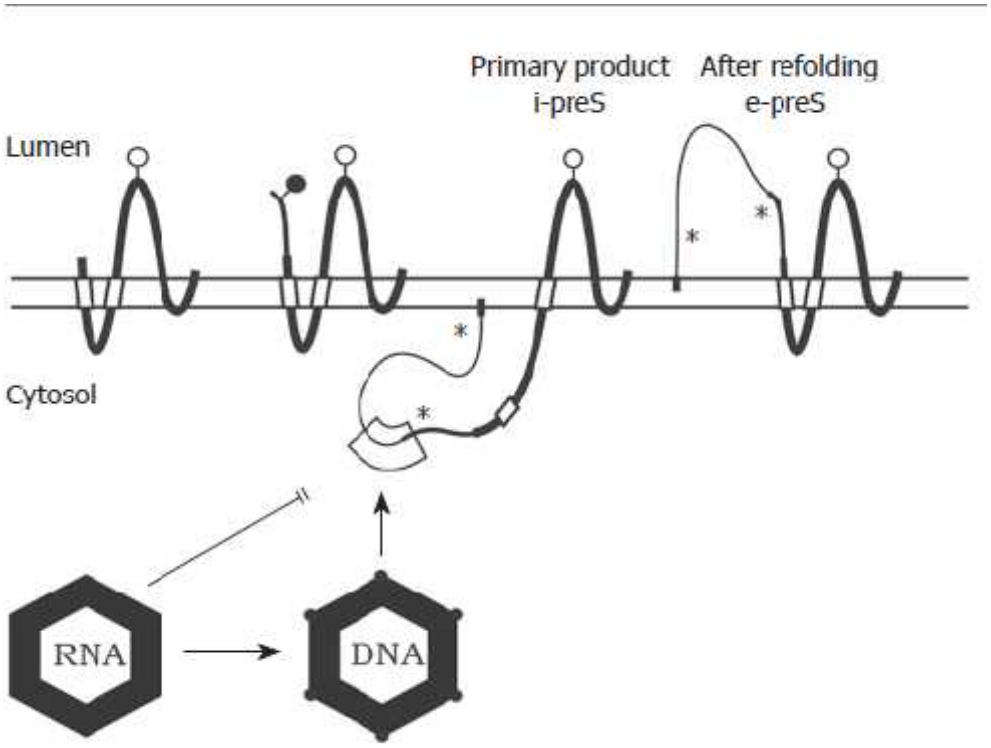

Figure 10: The model of the mature HBV core particle envelopment (completely taken from Bruss 2007)

6. Conclusion:

The structure of the HBV core, and thus the core monomer protein fold, described in this seminar was obtained by the cryo-EM and protein X-ray crystallography studies on purified, C-terminally truncated protein variants heterologously expressed in *E.coli*. It is therefore questionable to which extent obtained structural information are comparable to the structure of the HBV core protein in immature and mature core particles assembled *in vivo* in infected hepatocyte. However, the extensive research done by Roseman and Berriman 2005, compared the structures of the several HBV core variants; the bacterial RNA containing (*E.coli* expressed and assembled HBV core protein), the viral RNA containing and the viral mature (partially ds DNA containing) HBV core particles. The structural comparison of the general core morphology, core dimer and monomer folds, concluded that there are no detectable differences between the two RNA containing core types. The later implies that the monomer protein fold is largely determined by the in primary structure contained information and thus the possible influences of the cellular factors in HBV core monomer folding could be to some extent neglected. The precise information on the dynamics of the core monomer folding, the subsequent dimer and core assembly are still largely unknown, as the features have been under investigated.

The differences between the viral RNA and mature DNA containing particles showed to be the differences in the conformation of the dimer core subunits with the stressed difference in the hydrophobic pocket accessibility. The later structural information was used for the construction of the model for the core protein interactions with pre-S1 region of the S envelope protein and consequent HBV maturation. However, the information obtained must be considered with caution for several reasons. First of all, the viral RNA containing core particles, although hosting the viral encoded RNA, did not contain the exact viral pgRNA/RT complex found in the immature core particles assembled in infected hepatocyte. Therefore, it is questionable to which extent the core, dimer and monomer structural features of such particles are relevant for the comparison with the mature viral isolated cores. The lack of the ways to experimentally track the dynamical structural changes in HBV core dimers in the exact moment when the process of reverse transcription and consequent pgRNA-> partially ds DNA switch occurs, unfortunately, leaves only the possibility to observe the fragments of structural information (with questionable relevance) to deduce on these events.

7. References:

Birnbaum, F., Nassal, M. (1990) Hepatitis B virus nucleocapsid assembly; primary structure requirements in the core protein. *J Virol*; **64**: 3319-3330

Beck, J., Nassal, M. (2007) Hepatitis B virus replication. *World J Gastroenterol*; **13**: 48-64

Böttcher, B., Wynee, S.A., Crowther, R.A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature;* **386**: 88-91

Bruss, V. (2007) Hepatitis B virus morphogenesis. *World J Gastroenterol*; **13(1)**: 65-73

Ceres, P., Zlotnick, A. (2002) Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids. *Biochemistry*; **41**: 11525-11531

Chain, B.M., Myers, R. (2005) Variability and conservation in hepatitis B virus core protein. *BMC Microbiol*; **5**:33

Conway, J.F., Cheng, N., Zlotnick, A., Wingfield, P.T., Stahl, S.J., Steven, A.C. (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature*; **386**: 91-94

Crowther, R.A., Kiselev, N.A., Böttcher, B., Berriman, J.A., Borisova, G.P., Ose V., Pumpens, P. (1994) Three-dimensional structure of hepatitis B virus core particles determined by electron cryomicroscopy. *Cell*; **77**: 943-950

Dane, DS., Cameron, CH., Briggs M. (1970) Virus like particles in serum of patients with Australia-antigen-associated hepatitis *Lancet*; **1**: 695-698

Gallina, A., Bonelli, F., Zentilin, L., Rindi, G., Muttini, M., Milanesi, G. (1989) A recombinant hepatitis B core antigen polypeptide with the protamine-like domain deleted self-assembles into capsid particles but fails to bind nucleic acids. *J Virol*; **63**: 4645-4652

Ganem, D., Varmus, H.E. (1987) The molecular biology of hepatitis B viruses. *Annu Rev Biochem*; **56**:651-693

König, S., Beterams, G., Nassal, M. (1998) Mapping of homogous interaction sites in the hepatitis B virus core protein. *J Virol*; **72**: 4997-5005

Newman, M., Suk, F.M., Cajimat, M., Chua, P.K., Shih, C. (2003) Stability and morphology comparison of self-assembled virus like particles from wild-type and mutant human hepatitis B virus capsid proteins. *J Virol*; **77**: 12950-12960

Pasek, M., Goto T., Gilbert, W., Zink, B., Schaller, H., MacKay, P., Lead-better, G., Murray, K. (1979) Hepatitis B virus genes and their expression in *Escherichia coli*. *Nature*; **282**: 575-579

Patient, R., Hourioux, C., Roingeard, P. (2009) Morphogenesis of hepatitis B virus and its subviral envelope particles. *Cell Microbiol*; **11(11)**: 1561-1570

Perz, JF., Armstrong, GL., Farrington, LA., Hutin, YJ., Bell, BP. (2006) The contribution of hepatitis B and hepatitis C virus infections to cirrhosis and primary liver cancer worldwide. *J Hepatol*; **45:** 529-538

Roseman, A.M., Berriman, J.A., Wynee, S.A., Butler, P.J.G., Crowther, R.A.(2005) Structural model for maturation of the hepatitis B virus core. *PNAS;* **102**: 15821-15826

Wingfield, P.T., Stahl, S.J., Williams, R.W., Steven, A.C.(1995) Hepatitis core antigen produced in *Escherichia coli*: subunit composition, conformational analysis, and in vitro capsid assembly. *Biochemistry*; **34**: 4919-4932

Wynee, S.A., Crowther, R.A., Leslie, A.G. (1999) The crystal structure of the human hepatitis B virus capsid. *Mol Cell*; **3**:771-780

Zheng, J., Schödel, F., Peterson, D.L. (1992) The structure of hepadnaviral core antigens. *J Biol Chem* **267**: 9422-9429

Zlotnick, A., Cheng, N., Stahl, S.J., Conway, J.F., Steven, A.C., Wingfield, P.T. (1997) Localization of the C terminus of the assembly domain of hepatitis B virus capsid protein:implications for morphogenesis and organization of encapsidated RNA. *Proc Natl Acad Sci USA*; **94**: 9556-9561

8. Summary:

The hepatitis B virus (HBV) is small, enveloped virus, member of the *Hepadna* viral family with unusual, partially double-stranded DNA genome. The predominant viral core form *in vivo* is 34 nm particle with T=4 icosahedral symmetry, constructed from 240 copies of the HBV core protein (HBc). The HBV core protein is 22 kD, 183-185 amino acids long protein with primary structure composed of two structurally and functionally distinct regions. The N-terminal region is 149-151 amino acids long assembly domain, capable of directing the correct fold of the core monomer. The 34 residues long C-terminal domain, completely dispensable for core assembly, is extremely rich in conserved arginine (R) residues and takes part in the viral pre-genome/RT binding and encapsidation. The cryo-electron microscopy and X-ray crystallography studies on the HBV core particles revealed an unexpectedly large helical protein structure, with the long α helical hairpin dominating the entire monomer fold. The amphipatic helical hairpins of two core monomers interact forming the compact dimer, whose interface covers the large hydrophobic surfaces of the individual monomers. The hydrophobic interactions thus constitute the major driving force for the dimer assembly and account for the assembled core stability. The four helix bundle of the dimer is a distinct core structural feature, with important functional role. This is well observed upon the maturation of the immature pre-genomic RNA/RT complex-containing HBV core particle. The unusual reverse transcription of the pre-genomic RNA generates the partially ds DNA of mature virion, the event which induces the sequence of the conformational changes in the dimer core protein with which the nucleic acid interacts. The changes are best observable on the four-helix bundle dimer structure and insure the generating of the dimer hydrophobic pocket on core surface which subsequently interacts with the pre-S1 region of the HBV S envelope protein. This mechanism ensures that only the mature viral core particles (with encapsidated dsDNA) are capable of the successful envelopment and budding at an intracellular hepatocyte membrane.

9. Sažetak:

Virus hepatitisa B (HBV) je član *Hepadna* virusne obitelje s ovojnicom i neobičnim, djelomično dvostrukim DNA genomom. Prevladavajući oblik virusne srži, *in vivo*, je sferična proteinska čestica, 34 nm u promjeru, T = 4 ikosaedarne simetrije, građena od 240 monomera proteina srži HBV(HBc). Protein srži ima molekulsku težinu 22 kD te 183-185 duge sekvence aminokiselina koja se sastoji od dvije strukturno i funkcionalno različite regije. N-terminalnu regiju, nazvanu i domena slaganja, čine prvih 149-151 aminokiselina. Ova je regija posve samostalna u usmjeravanju funkcionalnog smatanje monomera proteina srži i posljedično sklapanje same srži. 34 aminokiseline duga C-terminalna domena, posve je neophodna za smatanje proteina i sklapanje srži te budući da je izuzetno bogata konzerviranim argininskim (R) amino kiselinama ima ulogu u slaganju kompleksa pre-genomske RNA/reverzna transkriptaze u sklapajuću proteinsku srž. Krio-elektronsko mikroskopske analize i proteinska X-ray kristalografija na srži HBV ukazala su na neočekivanu dominaciju α zavojnica u tercijarnoj strukturi monomera, s dugom α-helikalnom ukosnicom dviju zavojnica (α3 i α4) kao njezinom osnovom. Amfipatske α-helikalne ukosnice dvaju monomera, interakcijom tvore kompaktan dimer, čije sučelje pokriva veliku površinu hidrofobnih bočnih lanaca individualnih monomera. Hidrofobne interakcije, predstavljaju glavnu pokretaču snagu za spontano sklapanje dimera te ujedno čine osnovu stabilnosti čitave srži. Dvije α-helikalne ukosnice okupljaju u dimeru četiri α uzvojnice, koje tako čine četvero-helikalni snop, osobitu strukturu srži s važnom funkcionalnom ulogom. Uloga strukture četvero-helikalnog snopa najbolje se očituje u sazrijevanju nezrelih čestica srži HBV, koje sadrže kompleks pre-genomske RNA/RT. Neobičajenom reverznom transkripcijom pre-genomske RNA generira se djelomično dvolančana DNA, što je okidač za slijed konformacijskih promjena u dimerima srži. Promjene se najbolje očituju na strukturi četvero-helikalnog snopa dimera gdje osiguravaju izbočivanje hidrofobnog džepa sučelja dimera. Hidrofobni džep tako biva izložen na površini sada zrele srži te osigurava njezinu interakciju s pre-S1 regijom S proteina HBV ovojnice. Ovaj mehanizam osigurava da se samo zrelim česticama srži (s uklopljenom djelomičnom dvolančanom DNA) HBV-a omogućuje uspješna interakcija s proteinima ovojnice, što rezultira virusnim omatanjem i pupanjam s unutarstanične membrane hepatocita.