Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Characterization of the Hepatitis Delta Virus Small Antigen: Intracellular Localization, Structure, Multimerization and RNA Binding Ability

Carolina Alpalhão Mantero de Mendonça Alves

DISSERTAÇÃO PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS BIOMÉDICAS,
ESPECIALIDADE DE BIOLOGIA MOLECULAR E CELULAR

MARÇO 2013

Universidade Nova de Lisboa
Instituto de Higiene e Medicina Tropical

Characterization of the Hepatitis Delta Virus Small Antigen: Intracellular Localization, Structure, Multimerization and RNA Binding Ability

Carolina Alpalhão Mantero de Mendonça Alves

Supervisor: Professor Celso Cunha (IHMT, UNL)
Co-supervisor: Professor Emeritus John Taylor (Fox Chase Cancer Center)

Tutorial Comittee: Professor Celso Cunha (IHMT, UNL)
Professor Emeritus John Taylor (Fox Chase Cancer Center)
Professor Aida Esteves (IHMT, UNL)

Dissertation to obtain a Doctoral Degree in Biomedical Sciences, Specialization in Molecular and Cellular Biology

The financial support by the Fundação para a Ciência e a Tecnologia was granted with a PhD fellowship SFRH / BD / 41572 / 2007.

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Acknowledgements

I want to gratefully acknowledge those that directly or indirectly contributed to the work presented here.

To Professor Celso Cunha for everything – for his guidance and encouragement throughout all these years, for the knowledge he shared and his support and sense of humor!

To Professor John Taylor, for his guidance and support throughout this project, and for receiving me in his lab at Fox Chase Cancer Center.

To everybody who contributed to my research at the Fox Chase Cancer Center. Particularly, to Severin Gudima for his support and shared knowledge. To Ziying Han and Ning Chai, for their help and support. And to William Mason for his constructive criticism throughout the project.

To all my colleagues and former colleagues at Instituto de Higiene e Medicina Tropical. Particularly Ana Casaca, Marta Mendes and Cristina Branco, who "suffered" the most!

To my parents, my brother and Eduardo for their unconditional support and especially for their infinite patience.

Last, but definitely not least, to my American family, specially my "mom" Karen: we went through a lot in two years – I could not have done it without your support!

Caracterização do Antigénio Pequeno do Vírus da Hepatite Delta: Localização Intracelular, Estrutura, Multimerização e Capacidade de Ligação ao RNA

Carolina Alpalhão Mantero de Mendonça Alves

Resumo

O vírus da hepatite delta (HDV) é o agente patogénico responsável por uma das formas mais severas de hepatite viral. O genoma consiste numa molécula circular de RNA de cadeia simples de polaridade negativa e apresenta uma única proteína viral, o antigénio delta pequeno (S-HDAg). Na sequência de um mecanismo de *editing* outra proteína viral é traduzida, o antigénio delta grande (L-HDAg). Apesar de partilharem grande parte da sua sequência, as duas proteínas desempenham funções distintas. O S-HDAg é essencial para a acumulação de RNAs virais enquanto o L-HDAg inibe a replicação viral e é necessário para o empacotamento. O HDV depende extensivamente de factores do hospedeiro para completar o seu ciclo de replicação. Pensa-se que a polymerase II (pol II) do hospedeiro é redireccionada para transcrever o RNA viral.

No presente trabalho procurou-se caracterizar o S-HDAg e clarificar o seu papel no ciclo de replicação do HDV.

Observamos que quando o S-HDAg é expresso na presença de replicação do RNA viral, o antigénio co-localiza com a pol II. Contudo, a co-localização com pol II verifica-se mesmo na presença de RNA viral incapaz de ser replicado e na ocorrência de inibição da replicação viral, sugerindo que o S-HDAg não participa directamente na transcrição do RNA viral. Assim, propomos que o S-HDAg é essencial para acumulação de RNAs virais protegendo ou estabilizando os RNAs. Observamos ainda que na ausência de RNA viral o S-HDAg co-localiza com a nucleolina nos nucléolos. Contudo, na presença de RNA incapaz de ser replicado, o antigénio desloca-se para o nucleoplasma mantendo-se a nucleolina nos nucléolos, sugerindo que o S-HDAg não interage directamente com a nucleolina.

Ao estudarmos as características estruturais do S-HDAg verificámos, utilizando um preditor de desordem intrínseca, que apresenta um elevado grau de desordem. A previsão foi confirmada *in vitro* por dicroísmo circular observando-se que apenas 30% dos amino ácidos adoptam uma conformação de hélice α. A ausência de uma estrutura rígida pode conferir ao antigénio flexibilidade para se adaptar a diferentes parceiros e participar em vários passos do ciclo de replicação viral.

A multimerização do S-HDAg foi analisada por dispersão de luz dinâmica. Os resultados indicam que o antigénio recombinante purificado é capaz de formar multímeros de 12 moléculas. Adicionalmente, foram observados multímeros de seis a oito moléculas em gel de poliacrilamida desnaturante, após *cross-linking*. Os mesmos multímeros foram observados para S-HDAg presente em partículas virais sugerindo que a multimerização do antigénio ocorre *in vivo*.

Finalmente, estudamos a capacidade do S-HDAg interagir com ácidos nucleicos. Verificamos que multímeros e monómeros de S-HDAg são capazes de interagir *in vitro*

com RNA e DNA. A falta de especificidade observada pode dever-se apenas a interacções electrostáticas entre o S-HDAg de carga positiva (+12) e ácidos nucleicos de carga negativa. Propomos que, *in vivo*, a fosforilação extensiva do S-HDAg reduza a carga positiva contribuindo para que a interacção seja específica para os RNAs virais.

Palavras-chave: vírus da hepatite delta, antigénio delta pequeno, desordem intrínseca de proteínas, multimerização de proteínas, ligação proteína-ácidos nucleicos

Characterization of the Hepatitis Delta Virus Small Antigen: Intracellular Localization, Structure, Multimerization and RNA Binding Ability

Carolina Alpalhão Mantero de Mendonça Alves

Abstract

Hepatitis delta virus (HDV) is the causative agent of one of the most severe forms of viral hepatitis. It has a small single-stranded circular RNA genome of negative polarity and only one viral protein, the small delta antigen (S-HDAg). Following site-specific RNA editing, a second longer protein is translated, the large delta antigen (L-HDAg). Although these viral proteins share most of their sequence they play distinct roles. S-HDAg is essential for the accumulation of HDV RNAs whereas L-HDAg inhibits HDV replication and is necessary for viral assembly. With such a limited coding capacity HDV must rely extensively on host cell components to complete its replication cycle. The host DNA-directed RNA polymerase II (pol II) is thought to be re-directed to transcribe HDV RNAs.

The objective of this study was to further characterize S-HDAg and clarify its role(s) during the HDV replication cycle.

We observed that when S-HDAg was expressed *in vivo* along with replicating HDV RNA it co-located with host pol II. However, such co-localization was also observed in the presence of non-replicating HDV RNAs or when replication was inhibited by specific doses of α-amanitin. Thus, we propose that S-HDAg is essential for HDV RNA accumulation by stabilizing or protecting the viral RNAs rather than acting as a direct player in HDV RNA transcription. Additionally, we observed that S-HDAg located in nucleolus when expressed in the absence of HDV RNA, and co-located with host nucleolin. However, in the presence of non-replicating HDV RNAs, S-HDAg moved to the nucleoplasm whereas nucleolin was unchanged. This suggests that S-HDAg is not interacting directly with nucleolin.

In our examination of S-HDAg's structural features we applied a meta-predictor of intrinsic disorder, PONDR-FIT. It predicted that full-length S-HDAg has extensive intrinsic disorder. This result was confirmed *in vitro* by circular dichroism measurements that indicated no more than 30% of S-HDAg amino acids adopted an α-helical structure. Such a lack of a well-defined rigid structure is expected to grant flexibility to the antigen allowing it to interact with several partners and perform distinct roles during the HDV replication cycle.

Protein multimerization was studied by dynamic light scattering. Data analysis indicated that purified recombinant S-HDAg was able to assemble into homomultimers as high as dodecamers. Similarly, denaturing polyacrylamide gel electrophoresis with prior cross-linking indicated formation of at least hexamers and octamers. Similar multimers were observed for S-HDAg present in virus-like particles indicating that S-HDAg multimerization also occurs *in vivo*.

Abstract

Finally we examined the ability of S-HDAg to bind nucleic acids *in vitro*. Both multimers and monomers bound to conformations of both RNA and DNA. Such a lack of specificity was probably due to electrostatic interactions between the positively-charged S-HDAg (+12) and negatively-charged nucleic acids. We propose that *in vivo*, extensive post-translational phosphorylation of S-HDAg reduces the positive charge, thereby contributing to interactions more specific for HDV RNAs and possibly dependent upon protein multimerization.

Despite our observations presented here, some issues relating to our aims remain unresolved.

Key words: Hepatitis delta virus, small delta antigen, intrinsic disorder, protein multimerization, protein-nucleic acid binding.

Publications

Han, Z., Alves, C., Gudima, S. and Taylor, J. (2009) Intracellular localization of hepatitis delta virus proteins in the presence and absence of viral RNA accumulation. *J. Virol.,* 83, 6457-6463.

Alves, C. Cheng, H., Roder, H. and Taylor, J. 2010. Intrinsic disorder and oligomerization of the hepatitis delta virus antigen. *Virology*, 407, 333-340

Alves, C., Casaca, A. and Cunha, C. Multimerization of the delta antigen is not essential for in vitro nucleic acid binding. Submitted to Virology Journal

Table of contents

Table of contents

Abbreviations

3-D	3-dimensional
ADAR1	Adenosine deaminase acting on RNA
Ala	Alanine
Arg	Arginine
Asn	Asparagine
CCD	Coiled-coil domain
CD	Circular dichroism
$CPEB_3$	Cytoplasmic polyadenylation element-binding protein 3
Cys	Cysteine
DIPA	Delta interacting protein A
DNA	Deoxyribonucleic acid
ERK1/2	Extracellular signal-related kinases 1 and 2
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
HBV	Hepatitis B virus
HBsAgs	Hepatitis B virus surface antigens
HDAg	Hepatitis delta antigen
HDV	Hepatitis delta virus
IC	Internal control
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
IFN-α	Interferon α
IgG	Immunoglobulin G
IgM	Immunoglobulin M
Kb	Kilo-base pair

Abbreviations

Leu	Leucine
L-HDAg	Large hepatitis delta antigen
Lys	Lysine
mRNA	messenger RNA
NELF-A	Negative elongation factor, subunit A
NES	Nuclear export signal
NLS	Nuclear localization signal
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PEG-IFN-α	Pegylated interferon α
Phe	Phenylalanine
Pol II	DNA-dependent RNA polymerase II
Poly(A)	Polyadenylated
Pro	Proline
PTMs	Post-translational modifications
RNA	Ribonucleic acid
RNPs	Ribonucleoproteins
RT-qPCR	Real-time quantitative polymerase chain reaction
Ser	Serine
S-HDAg	Small hepatitis delta antigen
SUMO1	Small ubiquitin-related modifier isoform 1
Thr	Threonine
Trp	Tryptophan
Tyr	Tyrosine
Val	Valine

1. Introduction

1.1. Hepatitis Delta Virus

In 1977, a novel antigen was found in the nucleus of hepatocytes from patients with a more severe form of hepatitis B (Rizzetto et al., 1977). It was first thought to be a previously unknown marker of hepatitis B virus (HBV; Rizzetto et al., 1979). Only later, it was found that the delta antigen was not part of HBV but of a separate defective virus that requires the presence of HBV for infection (Rizzetto et al., 1980). The newfound virus was designated hepatitis delta virus (HDV) and, by 1986, its RNA genome was cloned and sequenced (Wang et al., 1986). This peculiar virus has been classified as the only member of the genus *Deltavirus* due to its uniqueness (Murphy, 1996). The HDV virion is a hybrid particle, composed of the delta antigen and HDV RNA enclosed by the surface antigens of HBV (HBsAgs). HDV has the smallest RNA genome of all known animal viruses. However, it is comparable, although larger, to viroid RNAs, pathogenic agents of higher plants (Rizzetto et al., 1980).

1.1.1. Epidemiology

HDV infection is distributed worldwide, although not uniformly, and it is estimated that 5% of HBsAgs carriers are also infected with HDV, which signifies that there might be between 15 and 20 million HDV infected individuals (Rizzetto and Ciancio, 2012). This is a very rough number because it lacks data from areas where HBV is highly prevalent and HDV is poorly studied.

HDV is highly endemic in Mediterranean countries, the Middle East, Northern parts of South America, and Central Africa (Radjef et al., 2004). HDV also has high prevalence in Turkey (Bahcecioglu et al., 2011; Değertekin et al., 2008), Western Pacific populations (Dimitrakakis and Gust, 1991), Central Asia (Tsatsralt-Od et al., 2005), and the Amazonian region of Western Brazil (Paraná et al., 2006).

In Southern Europe, HDV infection has been highly prevalent, with studies from the 1980s and 1990s showing that the incidence of HDV in HBsAgs positive individuals was higher than 20% (Farci, 2003). With the implementation of HBV vaccination programs in the 1980s, HDV prevalence considerably decreased to 5-10% by the late 1990s (Gaeta et al., 2007). However, in the beginning of the XXI century,

the number of HDV infected HBsAgs carriers in Europe increased to 8-12% (Gaeta et al., 2007; Wedemeyer et al., 2007a). This increase has been attributed to immigration of individuals from highly endemic regions (Wedemeyer et al., 2007a). Another report claims that the increase in HDV incidence is not only due to immigration but also to other factors associated with HDV modes of transmission (Gaeta et al., 2007). Drug addiction and other risk behaviors, such as multiple sexual partners, tattooing and piercing or uncontrolled medical procedures, have been shown to contribute to the spread of hepatitis D in Italy (Gaeta et al., 2007). In fact, in Western countries the virus is highly prevalent in intravenous drug addicts with chronic HBV infection (Gaeta et al., 2007; Wedemeyer et al., 2007a).

More recent and reliable data are needed, especially from poorly studied high endemic regions. Only in the last couple of years are we getting to know numbers from certain areas of the world. In the last monothematic conference on HDV, held in Istanbul in 2010, new HDV epidemiologic data was presented for several countries, some of which had no previous data. In countries such as Albania (Sadiku and Basho, 2010), Libya (Elhaasi et al., 2010) and Mauritania (Mansour et al., 2010) high HDV prevalence was reported, with 10-19% of HBsAgs carriers being anti-HDV antibody positive. In Cameroon, 7.9% of pregnant women HBsAgs positive were anti-HDV antibody positive (Abgueguen et al., 2010), and a study in Tyva Republic, in the Russian Federation, reported that 2.5% of healthy individuals present in a trial were anti-HDV antibody positive (Kozhanova et al., 2010). An epidemiological study in the Republic of Korea has reported that only 1 patient in 226 HBsAgs carriers was positive for anti-HDV antibody, showing very low prevalence in that part of Asia (Jung et al., 2010). In Portugal, the only epidemiological study on HDV, dating back to 1987, reported that 8.4% HBsAgs carriers had chronic HDV infection (Ramalho et al., 1987).

Based on nucleotide sequence analysis, eight HDV genotypes have been defined, some of which are distributed by distinct geographic regions (Le Gal et al., 2006; Radjef et al., 2004). The divergence in nucleotide sequence between isolates of the same genotypes is less than 15% and between different genotypes it can be as high as 40% (Radjef et al., 2004). HDV genotype 1 is the most common and prevalent

worldwide, present mainly in Europe, Middle East, North America and Northern Africa. It is associated with both severe and mild forms of the disease (Su et al., 2006). Genotype 2 is more common in the Far East, being present in Japan, Taiwan and parts of Russia (Hughes et al., 2011). Genotype 2 is associated with a milder disease course (Su et al., 2006). HDV genotype 3 is exclusively found in the Amazon Basin (Paraná et al., 2006) and genotype 4 is present in Japan and Taiwan (Wu et al., 1998). Genotypes 5 to 8 were found in African patients that had migrated to Northern Europe (Le Gal et al., 2006; Radjef et al., 2004). Phylogenetic reconstructions based on the delta antigen coding sequence have shown a probable ancient radiation of African lineages (Radjef et al., 2004).

1.1.2. Clinical Expression

Hepatitis delta virus usually induces a severe form of hepatitis but, as will be discussed in this Section, the range of clinical manifestations is very wide going from asymptomatic cases to fulminant hepatitis.

Regarding HDV transmission, like its helper virus HBV, it is parenterally transmitted through exposure to infected blood or body fluids. Intrafamilial spread is naturally common in highly endemic regions. Tests made in chimpanzees have shown that very small inocula are sufficient to transmit infection (Ponzetto et al., 1987). Hence, transmission rates are high amongst intravenous drug users. Also, people with high-risk sexual activity have an increased risk of infection (Gaeta et al., 2007). However, blood transfusion recipients or patients subject to haemodialysis are no longer at risk of infection in developed countries because prior screening of blood products is performed.

HDV requires the presence of HBsAgs to form new infectious virions and propagate HDV infection. Thus, hepatitis D only occurs in individuals infected with HBV. Consequently, there are two major patterns of infection: “co-infection” with HBV and HDV or “super-infection” of patients already infected with HBV. A rare third pattern has been reported; it can occur after liver transplantation for an HDV-infected individual, and is designated as “helper-independent latent infection” (Ottobrelli et al., 1991). In this scenario an initial HDV infection of the new liver occurs without any

apparent help from HBV. Such an infection remains asymptomatic unless reactivated by HBV appearance (Ottobrelli et al., 1991).

For an HBV and HDV acute co-infection the most common outcome (95%) is viral clearance (Hughes et al., 2011). However, such a co-infection can be more severe than an acute HBV mono-infection, resulting in some cases in acute liver failure (Govindarajan et al., 1984). Acute hepatitis strikes after an incubation period of 3-7 weeks, beginning with a period of non-specific symptoms such as fatigue, lethargy or nausea (Farci and Niro, 2012). This pre-icteric phase is also characterized by elevated serum alanine aminotransferase. The subsequent icteric phase is not always observed but when it occurs it is characterized by high levels of serum bilirubin (Farci and Niro, 2012).

HDV super-infection of chronic HBV patients also causes severe acute hepatitis but in this case, for up to 80% of patients, it progresses to chronicity (Smedile et al., 1982). The processes, which determine whether a patient clears HDV spontaneously or becomes chronically infected, remain unclear. When chronic HDV infection is established, the pre-existing liver disease caused by HBV is usually aggravated (Smedile et al., 1981). It has been claimed that during the acute phase of HDV infection, HBV replication is suppressed to very low levels and that this suppression can persist once a chronic HDV infection is established (Farci et al., 1988). Patients with HDV super-infection suffer a more rapid progression to cirrhosis (Fattovich et al., 1987; Saracco et al., 1987), increased liver decompensation and eventually death (Fattovich et al., 2000; Romeo et al., 2009), when compared with patients with HBV mono-infection. Despite the higher rates of progression to cirrhosis not all published studies refer to an increased rate of hepatocellular carcinoma (Cross et al., 2008). One explanation of this may be the abovementioned suppression of HBV replication by HDV, since other studies assert that higher HBV DNA serum levels correlates with a greater risk of carcinoma (Chen et al., 2006).

HDV super-infection can also be associated with an increased risk of acute or fulminating hepatitis. This may be due to a sudden loss of hepatocytes, leading to

multiorgan failure, so that a patient can go from a healthy status to near death within just 2 to 10 days (Farci and Niro, 2012).

1.1.3. Diagnosis and Treatment

There are three serological markers specific for an HDV infection: HDV RNA, delta antigen (HDAg), and anti-HDV antibodies. The presence of HDAg in serum marks an acute infection, while anti-HDV IgG antibody reflects a past or chronic infection, and anti-HDV IgM antibody is characteristic of the period between the appearance of HDAg and the development of IgG anti-HDV in chronic infections (Shattock and Morris, 1991).

Since HDV is a satellite virus of HBV, every HBsAgs positive patient should be screened for co-infection with HDV; that is, patients should be tested, at least once, for anti-HDV antibodies. A negative result does not justify testing for HDV RNA as, so far, it seems that every individual infected with HDV develops anti-HDV antibodies (Wedemeyer and Manns, 2010). In contrast, a positive result for anti-HDV antibodies requires confirmation of continued HDV infection, through detection of HDV RNA in serum. Anti-HDV antibodies may be present even after HDV RNA has disappeared during recovery from the infection (Wedemeyer and Manns, 2010).

Currently, there is no need for quantification of the HDV RNA levels in serum during the diagnosis step. There is no evidence that a correlation exists between the stage of liver disease and the levels of HDV RNA (Zachou et al., 2006). Thus, a liver biopsy is still the major tool for evaluating the stage of delta hepatitis in patients (Wedemeyer and Manns, 2010). However, a quantitative assay of HDV RNA is useful during the therapy stage to monitor the treatment response of patients undergoing therapy. Unfortunately, very few data are available on the levels of HDV RNA during the different stages of the disease. Thus, there is no accepted threshold level at which one might recommend treatment.

For some time quantification of HDV RNA levels in clinical samples has suffered from the lack of a standardized test. Quantification of HDV RNA was done in

specialized laboratories using in-house protocols, which unfortunately become irrelevant outside the laboratory of origin. Such assays typically lacked an internal control (IC) and were limited to only one genotype. Furthermore, there was no international reference standard to make results from different laboratories comparable (Chudy, 2010; Pawlotsky, 2010). As proposed elsewhere, an HDV RNA reference preparation should be defined by the World Health Organization to be used as an international standard (Olivero and Smedile, 2012).

In 2012, two standard protocols were proposed to detect and quantify HDV RNA from clinical samples (Ferns et al., 2012; Scholtes et al., 2012). The method proposed by Caroline Scholtes *et al.* is described as able to be automated to accurately quantify the major HDV genotypes present in Europe (genotype 1 and the migrant African strains 5-8; Scholtes et al., 2012). It uses a commercial kit to extract nucleic acids from samples, and includes an internal control. The IC is added to the samples before extraction to enable monitoring of the overall performances of the assay. The one-step RT-qPCR makes use of another commercial kit. To detect HDV nucleic acids two forward primers and one reverse primer were designed to bind conserved parts of the delta antigen coding sequence. In addition, *in vitro*-transcribed HDV RNAs were consecutively diluted and used as standards.

The other standardized test, as proposed by R. B. Ferns and colleagues, uses a Brome Mosaic virus RNA as an internal control and also requires a one-step RT-qPCR, using a commercial kit (Ferns et al., 2012). They describe the protocol as being able to detect and quantify all HDV genotypes (Ferns et al., 2012).

Application of standardized procedures is crucial to improve our understanding of HDV RNA kinetics during the course of disease. It will improve patient management, as data can be gathered that will help in the decision to start treatment, as well as monitoring the response to therapy in chronic patients. Also it will contribute to the screening of HDV infections in the endemic areas, providing more reliable epidemiological data. Overall, acceptance of standardization will help clarify the pathophysiology of HDV infections.

Ideally, a successful treatment of an HDV infection eradicates HDV and its helper virus HBV. Clearance of HDV is obtained when both HDV RNA and HDAg in

the liver become persistently undetectable and a complete resolution is achieved when HBsAgs clearance is also obtained.

However, at this time there is no efficient therapy. Prolonged treatment with recombinant interferons is the only therapy that has shown antiviral activity against HDV. Such therapies, which last up to 2 years, have been reported as only 20-40% efficient (Wedemeyer et al., 2011).

In general, when searching for a treatment for viral disorders the first and preferred targets analyzed are the viral components, such as enzymes involved in the virus replication cycle. But HDV lacks any specific enzymatic function to target. Since the only known enzymatic activity the virus possesses is a ribozyme, the virus relies on the host cell to provide for all other enzymatic activities needed for its life cycle. This represents a serious challenge in finding an HDV-specific therapeutic target.

Puzzlingly, the nucleoside and nucleotide analogues used for treatment of HBV infection are inefficient against HDV. Although they block HBV DNA synthesis in chronic patients, they have little impact on HDV and do not even enhance interferon treatments (Wedemeyer et al., 2011). Famciclovir, lamivudine and adefovir, all used in HBV treatment, have been shown to lack any significant antiviral activity against HDV (Niro et al., 2005; Wedemeyer et al., 2007b; Yurdaydin et al., 2002). Ribavirin, a nucleotide analogue, which inhibits HDV replication in cell culture, when administered alone or in combination with interferon also failed to increase rates of HDV RNA clearance (Garripoli et al., 1994; Niro et al., 2006).

Interferon-α (IFN-α) has been used for treatment of HDV infections since the mid-1980s (Rizzetto et al., 1986). Several trials were carried out exploring different doses and durations. Responses to treatment varied and clearance occurred at different times from the beginning of treatment, occurring even after discontinuation of treatment (Niro et al., 2005). Although, such data are difficult to compare and very few studies have measured HDV RNA levels during treatment, some conclusions arose from these trials (Niro et al., 2005). Researchers have yet to identify pre-treatment characteristics that determine responders and non-responders to IFN-α therapy. It seems that 2 years of treatment with IFN-α is superior to shorter treatment durations to obtain HDV RNA clearance (Niro et al., 2005). It has been reported that in a 1-year treatment there is only

a 10 to 20% chance of HDV clearance and in a 2-years treatment trial, 20% of patients were cleared (Farci et al., 2007). The rate of response is proportional to the dose of IFN-α; patients treated with doses of 9 million units responded better than those treated with only 3 million units, and relapse was common when the IFN-α dose was reduced (Niro et al., 2005). Unfortunately, a prolonged treatment with high doses of IFN-α is tolerated by only a minority of patients (Wedemeyer and Manns, 2010). IFN-α side effects include flu-like symptoms, fatigue and weight loss as well as severe psychiatric disturbances. Patients have a tendency to become deeply depressed; suicides and attempted suicides have been reported (Niro et al., 2005). The severity of reactions tends to be proportional to IFN-α dose and intermittent use of IFN-α, observed in drug abusers, increased incidence and severity of side effects (Niro et al., 2005).

By 2006, IFN-α was largely replaced by longer-lasting pegylated IFN-α (PEG-IFN-α; Castelnau et al., 2006; Erhardt et al., 2006; Niro et al., 2006). Clearance of HDV RNA was obtained for 6 out of 14 patients in a 1-year treatment plan (Castelnau et al., 2006). However, in a similar study only 2 patients in 12 were cured (Erhardt et al., 2006). In a third study, 8 patients out of 38 became HDV RNA-negative after 72 weeks of treatment (Niro et al., 2006). Ribavirin was also used in this trial but without any apparent beneficial effect (Niro et al., 2006).

The Hep-Net International hepatitis D intervention trial, which included 90 patients from Germany, Greece and Turkey, tested PEG-IFN-α2a alone or with adefovir and adefovir alone (Wedemeyer et al., 2007b). HDV RNA clearance was only observed in patients who had received treatment including PEG-IFN-α2a, showing an antiviral efficacy in more than 40% of patients, and 25% became HDV RNA negative (Wedemeyer et al., 2007b). Adefovir showed little efficacy in reducing HDV RNA levels but a PEG-IFN-α2a plus adefovir therapy was superior in reducing HBsAgs serum levels (Wedemeyer et al., 2007b).

Currently, it is usually recommended to treat chronic hepatitis D with PEG-IFN-α for one year or longer, if the patient can tolerate the adverse effects of such therapy (Hughes et al., 2011). For patients with advanced liver disease, liver transplantation is the only therapy available (Farci et al., 2007).

An optimization of the available treatment strategies is clearly needed, either regarding doses and duration and also, possible combinations such as PEG-IFN-α2a with adefovir to also tackle HBsAgs, crucial for HDV propagation. Most importantly, alternative treatments need to be explored, as the efficacy of the current therapies is clearly unsatisfactory. One of the most promising alternatives are prenylation inhibitors since, as will be discussed subsequently, prenylation of HDAg is essential for interaction with HBsAgs. Furthermore, prenylation inhibitors have already been developed to treat malignancies and were shown to be safe (Bordier et al., 2003).

1.2. HDV Biology

1.2.1. HDV Virions and Putative Host Cell Receptors

An infectious HDV virion is an enveloped, roughly spherical particle, of around 36 nm in diameter (He et al., 1989). The outer coat of the virion contains host lipids and the HBsAgs. As illustrated in Figure 1, there are three size classes of HBsAg, referred to as large (L-), medium (M-) and small (S-). The envelope surrounds an inner nucleocapsid consisting of viral ribonucleoproteins (RNPs) with the genomic RNA and about 200 molecules of HDAg per genome (Gudima et al., 2002).

Since HDV and HBV share the same envelope proteins it is often assumed that attachment and cell entry occur via similar mechanisms. Attachment of HBV to cultured cells has been reported to require glycosaminoglycans but it is not known if this is also true for HDV or for hepatocyte attachment (Leistner et al., 2008; Schulze et al., 2007).

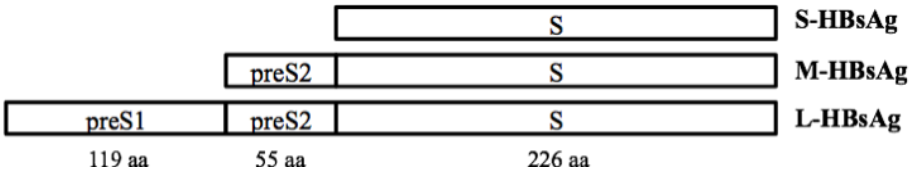


Figure 1: HBV surface antigens. Schematic representation of the three HBsAgs: S-HBsAg, M-HBsAg, and L-HBsAg. The C-terminus region, common to the three antigens is represented as S. preS2 represents the domain unique to M-HBsAg and L-HBsAg; preS1 represents the domain exclusive to L-HBsAg. The average size of each region is referred to by the number of amino acids (aa).

Several studies have attempted to identify the regions of the HBsAgs required for HDV and HBV entry. The three HBsAgs share a common C-terminus. M-HBsAg and L-HBsAg also share a preS2 domain whereas L-HBsAgs has a unique domain designated preS1 (Figure 1). The preS1 region is myristoylated at the N-terminus. This post-translational modification and about 48 adjacent amino acids are essential for HBV and HDV entry into hepatocytes. Synthetic peptides that mimic this region are potent inhibitors of virus entry (Engelke et al., 2006).

Many studies have aimed to discover the host receptors for HBV (and maybe HDV). Many candidates have been proposed but not confirmed (Glebe and Urban, 2007).

As a recent example, it has been suggested that functional purinergic receptors are required for HDV entry as compounds that block the activation of such receptors inhibited HDV and HBV infection of primary human hepatocytes (Taylor and Han, 2010).

In contrast to all previous studies an important new report by Yan and colleagues demonstrates that a necessary and sufficient receptor for HBV and HDV is the sodium taurocholate co-transporting polypeptide (Yan et al., 2012). This protein is a multiple transmembrane transporter expressed in the liver. Silencing expression of this protein in primary hepatocytes using small interfering RNAs inhibited HBV and HDV infection. Expression of this protein in human liver cell lines rendered them susceptible to infection by HBV and HDV. Therefore, it is now possible for the first time to study the infection processes for these viruses *in vitro*, using established human liver cell lines, which are much more convenient and reproducible than primary hepatocyte cultures.

1.2.2. HDV RNAs

Since its discovery as an HBV satellite virus, HDV has puzzled scientists. HDV has a small circular RNA genome with only ~1700 nucleotides; this sequence length varies by no more than 30 nucleotides among HDV isolates (Dény, 2006). In native conditions the RNA folds into an unbranched rod-like structure due to intramolecular base pairing involving around 74% of its nucleotides (Kuo et al., 1988).

Largely due to its small length and circular conformation, the HDV RNA genome is very different from all known viruses of animals. However, it does have some similarity to that of the plant viroids, whose genomes are also circular in conformation and even smaller in length, lacking any known coding capacity (Tsagris et al., 2008).

HDV contains one functional open reading frame (ORF), encoding the delta antigen (Chen et al., 1986). This ORF is not encoded by the genomic RNA but by another RNA species that arises during replication, the HDV antigenome, an exact complement of the genome.

The delta antigen is transcribed from a third RNA species, a linear 0.8 Kb messenger RNA (mRNA) of antigenomic polarity and a 5'-cap and 3'-polyadenylated tail (Hsieh and Taylor, 1991). The different HDV RNA species are represented in Figure 2. In an infected cell, the three HDV RNA species accumulate in very different amounts, although genomic RNA is the only species assembled into HDV virions. HDV genomic RNA is the most abundant, around 300,000 copies accumulate in an infected cell whereas 100,000 copies of the antigenome are present (Chen et al., 1986). The HDV mRNA is considerably less abundant with approximately 500 copies per cell (Gudima et al., 2000).

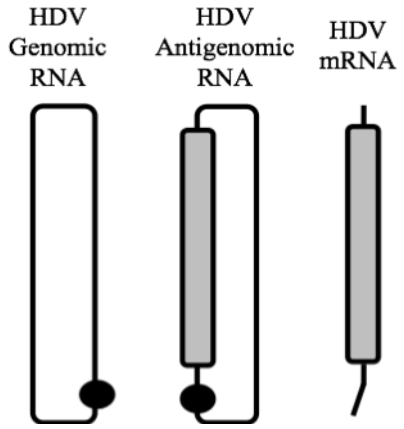


Figure 2: Schematic representation of three HDV RNA species. The HDV genomic RNA is a single-stranded circular RNA with ~1700 nucleotides. It forms an unbranched rod-like structure due to intramolecular base pairing. The HDV antigenomic RNA is the exact complement of the genomic RNA, both RNAs have site-specific ribozymes indicated by the black circle. The HDV antigenomic RNA contains the open reading frame for the delta antigen, represented in grey, but the antigen is translated from another RNA species, the mRNA. The mRNA is ~800 nucleotides long with a 5'-cap and a 3'-polyadenylated tail.

Another peculiar characteristic was found in 1988 when site-specific self-cleavage and ligation were reported on antigenomic HDV RNA, showing that HDV RNA possesses ribozyme activity, just like plant viroids (Sharmeen et al., 1988). Both genomic and antigenomic RNAs display this ribozyme activity, which is comprised within a contiguous sequence of less than 100 nucleotides (Branch et al., 1989). Both

ribozyme RNAs have been crystallized and an atomic structure solved (Ferré-D'Amaré et al., 1998). They enhance HDV RNA self-cleavage by a 10^6- to 10^7-fold when compared with uncatalyzed cleavage (Been, 2006; Ferré-D'Amaré et al., 1998). Although ribozymes are characteristic of viroids, their structures are different from HDV ribozymes, which are actually more related to the cytoplasmic polyadenylation element-binding protein 3 ($CPEB_3$) ribozyme, a conserved mammalian sequence within an intron of the $CPEB_3$ gene (Salehi-Ashtiani, 2006). In fact, numerous HDV-like ribozymes have since been found in several eukaryotic species (Webb et al., 2009).

1.2.3. HDV RNA Replication

HDV RNAs are transcribed in the nucleus of infected cells, but the details of this process remain poorly defined. As mentioned before, during HDV replication three RNA species accumulate in infected cells: the genome, antigenome and mRNA (Figure 2). Each of these is the product of post-transcriptional processing.

The precursors, from which these species arise, are thought to be transcribed by a double-rolling circle mechanism, exemplified in Figure 3. In this model the circular genome RNA is used as a template to produce multimeric species of opposite polarity (Taylor, 1990). These greater than unit-length RNAs are subsequently self-cleaved by the HDV ribozymes and re-ligated, producing unit-length circular antigenomic RNAs. The re-ligation step is thought to involve a host ligase (Reid and Lazinski, 2000) although it has been shown that the HDV ribozyme can self-ligate *in vitro* (Sharmeen et al., 1989). Through a similar mechanism the unit-length circular antigenomic RNA acts as a template for the transcription of multimeric species, which are processed to produce genomic RNA. The genomic RNA also acts as a template for transcripts that are processed into mRNA.

Even though such a rolling-circle mechanism has been widely accepted as a model for HDV replication, critical details remain to be confirmed and/or clarified such as the host cell components involved.

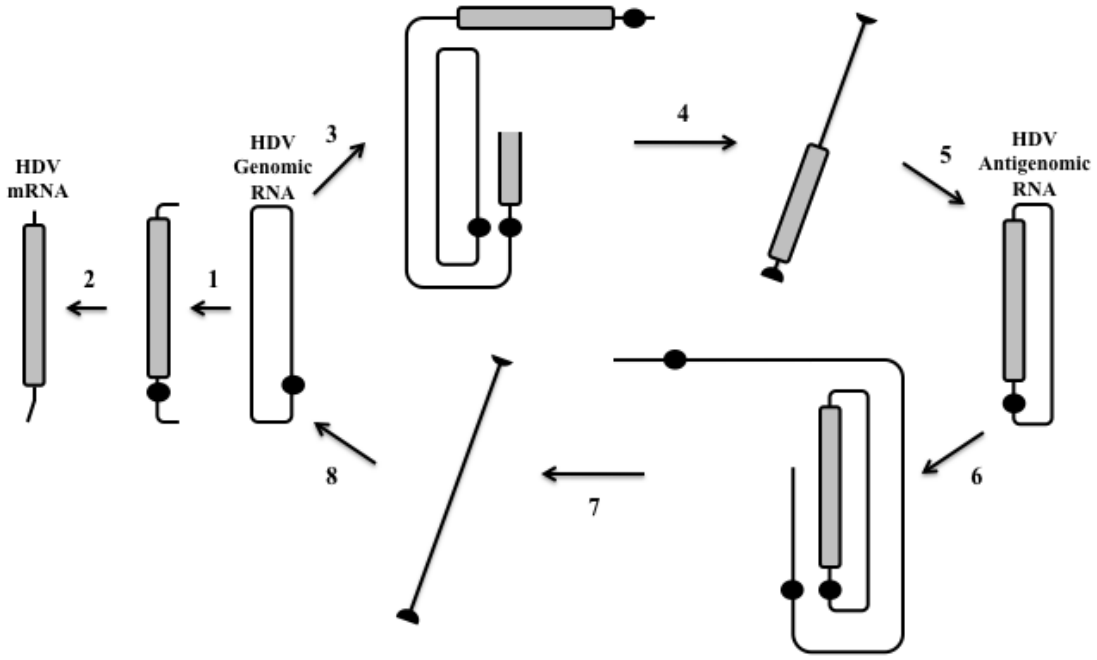


Figure 3: Model of HDV replication through a rolling-circle mechanism. The HDV genomic RNA is used as a template for the precursors of HDV mRNA (Steps 1-2) and also acts as a template for multimeric RNAs of antigenomic polarity (Step 3). These multimeric RNAs contain at least two copies of the HDV ribozyme and are thus self-cleaved to produce linear unit-length HDV antigenomes (Step 4), which are then ligated to produce circular antigenomic RNA (Step 5). In turn, the new antigenomic RNA is a template for multimeric RNAs of genomic polarity (Step 6) that are similarly self-cleaved and subsequently ligated to produce new circular genomic RNA (Steps 7-8).

Most RNA viruses use a virus-encoded RNA-directed RNA polymerase for replication. This is consistent with the dogma that host cell RNA polymerases are DNA-dependent and do not accept RNA templates. An important exception are retroviruses which first synthesize a DNA intermediate using a virus-encoded reverse transcriptase, to produce a DNA intermediate, which becomes integrated into the host genome and is then transcribed into RNA, using a host RNA polymerase. However, HDV has no known DNA intermediate (Taylor, 2009), and the only HDV protein, the delta antigen, is too small to be a polymerase. This means that HDV RNA must somehow redirect host DNA-dependent RNA polymerases to use HDV RNAs as templates. How this is achieved and which host polymerase(s) is (are) involved has been extensively studied but the results remain somewhat controversial.

The host RNA polymerase II (pol II) seems to be required for genomic HDV RNA transcription. Nuclear run-on experiments on an endogenous HDV RNA template

have shown that inhibition of pol II by low concentrations of the specific inhibitor α-amanitin blocks HDV RNA synthesis of both the genomic and antigenomic strands (Chang et al., 2008). One possible explanation is that the rod-like conformation of HDV RNAs may trick pol II into accepting the RNA as a double-stranded DNA template. It has been shown, through immunoprecipitation assays, that pol II binds the terminal stem-loop regions of HDV genome (Greco-Stewart et al., 2007). It has also been reported that, after binding to the stem-loop, pol II is able to elongate multimeric RNA species, carrying out transcription (Filipovska and Konarska, 2000). Such elongation was observed on a partial antigenomic RNA stem-loop and originated a chimeric molecule of newly synthesized transcript covalently bound to the 5'-end of the template. Thus, it is not clear if such elongation is biologically relevant.

Despite being shown that pol II interacts with genomic HDV RNA it has been suggested that a different host polymerase is responsible for the synthesis of antigenomic HDV RNA (Macnaughton et al., 2002; Modahl et al., 2000). The idea that at least two different host polymerases are involved in the HDV replication cycle is based on the observation, that in transfected cells, the synthesis of new HDV antigenomic RNA was not inhibited by concentrations of α-amanitin that would inhibit pol II activity (Modahl et al., 2000). This has led to the speculation that pol I copies genomic HDV RNA to produce new antigenomic RNA. This is contrary to the afore mentioned nuclear run-on assays, which have shown that both genomic and antigenomic RNA synthesis are sensitive to low doses of α-amanitin, consistent with pol II (Chang et al., 2008).

An additional complication arises from *in vitro* studies, which indicate that fragments of the HDV RNA genome interact not only with pol II but also with pol I and pol III (Greco-Stewart et al., 2009). However, such *in vitro* interactions may not have biological relevance, especially since they do not lead to RNA-directed transcription.

The HDV mRNA possesses characteristics of a pol II transcript that is processed to a mRNA: namely a 5'-cap structure and a 3'-poly(A) tail. In fact, the role of pol II in HDV mRNA transcription has been generally accepted (Chang and Taylor, 2002; Chang et al., 2008; Fu and Taylor, 1993; Macnaughton et al., 2002; Modahl et al., 2000).

The controversy regarding the transcription process is thus limited to whether genomic RNA is transcribed by pol II or another polymerase, possibly pol I (Taylor, 2012).

In addition to the post-transcriptional processing to make the abovementioned three HDV RNAs, there is an important RNA-editing event. During the virus replication cycle some of the antigenomes are edited at a specific site by a host adenosine deaminase (ADAR1). This changes the adenosine in the amber codon to inosine. After subsequent RNA-directed RNA synthesis, it leads to the replacement of inosine with guanosine (Polson et al., 1996). That is, the UAG stop codon is changed to a UGG tryptophan codon. In this way the delta antigen ORF is extended by 19 amino acids; that is, to the next stop codon. The specificity of the editing site is in part directed by the specific folding of the HDV antigenomic RNA (Casey, 2006).

Therefore, although HDV has only one ORF, it encodes two proteins: the small delta antigen (S-HDAg) of 195 amino acids and the large delta antigen (L-HDAg) with 214 amino acids, which will be described in Section 1.3.

1.2.4. HDV Origin

The resemblance of HDV RNA to that of plant viroids could indicate a viroid ancestor. The similarities are indeed extensive: the small genome size and circular conformation, the presence of ribozyme domains and the proposed rolling-circle mechanism for HDV replication (recently reviewed by Flores et al., 2012). Viroids are nevertheless significantly smaller than HDV, ranging from 250 to 400 nucleotides, and they are also non-coding RNAs (Tsagris et al., 2008). HDV could have evolved to encode the delta antigen and HDAg coding sequence accounts for HDV's longer genome. In fact, a cellular homolog of the HDAg has been reported, designated as Delta Interacting Protein A (DIPA), and was initially considered has a candidate for HDAg origin (Brazas and Ganem, 1996). However, the homology to HDAg was subsequently evaluated as not significant, thus disproving the hypothesis (Taylor and Pelchat, 2010).

Like HDV RNA, viroid RNAs are circular molecules that display intramolecular folding achieving almost rod-like structures. Some viroids, known as avsunviroids, replicate via a symmetrical rolling-circle mechanism as the one proposed for HDV (Taylor, 2009; Tsagris et al., 2008).

Despite the similarities that support a proposed evolution from the viroid world, there are important differences that advocate a different origin. For example, although ribozymes are characteristic of viroids, the two HDV ribozymes are structurally very different from those of the viroids. In contrast, several HDV-like ribozymes have been found in eukaryotes (Webb et al., 2009). This more supports the hypothesis that HDV may be derived from the human transcriptome rather than a plant or ancestral RNA world. Also, the HDV genome is larger than viroids and its antigenome encodes a viral protein.

In summary, although HDV has similarities to viroids, there are fundamental differences that avert us from stating, for now, that HDV originates from the plant rather than animal world.

1.3. Delta Antigens

HDV has only one open reading frame, which, as mentioned in Section 1.2.3, due to some post-transcriptional site-specific RNA editing, ultimately encodes two proteins, S-HDAg and L-HDAg. The two isoforms share 195 amino acids and differ only in that the large form has 19 extra amino acids on the C-terminus. As such, S-HDAg and L-HDAg share several functional domains within the common amino acid sequence, as illustrated in Figure 4. The delta antigens contain a nuclear localization signal (NLS) comprised by amino acids 66 through to 75 (Alves et al., 2008); a coiled-coil domain (CCD), also referred to as dimerization domain, within amino acids 12 to 60; and an RNA binding domain within amino acids 97 and 143 (Lee et al., 1993). L-HDAg has, within its extra sequence, a nuclear export signal (NES) spanning amino acids 198 to 210 (Lee et al., 2001).

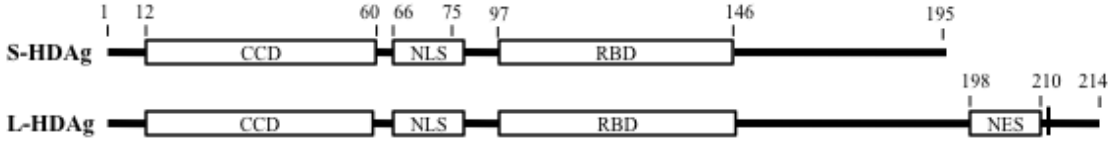


Figure 4: Functional domains of S-HDAg and L-HDAg. The delta antigens share most of their sequence differing only in the 19 amino acids extension at the C-terminal of L-HDAg. They have, within the common sequence, as represented, a coiled-coil domain (CCD), a nuclear localization signal (NLS), and an RNA binding domain (RBD). Also indicated on L-HDAg are the nuclear export signal (NES) and the unique cysteine, residue 211, which is the target for farnesylation. The numbers indicate the position of the amino acid residues.

Both delta antigens undergo post-translational modifications (PTMs) by several host enzymes. Several groups have investigated the impact these PTMs may have on the antigens' functions but the precise significance of most of these modifications remains uncertain.

The exception is one PTM, characteristic only of L-HDAg, which has been shown to be essential. It occurs on cysteine residue 211, and is mediated by a host farnesyltransferase (Glenn et al., 1992; Otto and Casey, 1996). This isoprenylation of L-HDAg is necessary, although not sufficient for viral packaging. It is somehow necessary for the interactions with HBsAgs, leading to the assembly of new viral particles (Hwang and Lai, 1994; Lee et al., 1994).

There are other PTM events, ones shared by both forms of the delta antigen. These involve phosphorylation, methylation, acetylation and sumoylation (Chen et al., 2008; Hong and Chen, 2010; Li et al., 2004; Mu et al., 2004; 1999; Tseng et al., 2008).

Phosphorylation has been observed at multiple sites, mostly at serine and threonine residues. Different phosphorylation patterns were observed for S-HDAg and L-HDAg, and, if relevant, the distinct patterns may in part account for their distinct biological functions (Mu et al., 1999). Several host enzymes have been reported to phosphorylate delta antigens at different sites: casein kinase II on Ser2 and Ser213 (Yeh et al., 1996); double-stranded RNA-activated protein kinase R on residues Ser177, Ser180, and Thr182 (Chen, 2002); extracellular signal-related kinases 1 and 2 (ERK1/2) on Ser177 (Chen et al., 2008); and protein kinase C on residue Ser210 (Yeh et al., 1996). It has been alleged that S-HDAg phosphorylation increases replication of genomic HDV RNA from the antigenomic strand (Chen et al., 2008). By enhancing the expression of ERK1/2 in cells transfected with plasmids expressing S-HDAg and dimeric HDV antigenomic RNA, an increase in the accumulation of HDV genomic RNA was observed, but not for antigenomic RNA (Chen et al., 2008). More recently, it has been suggested that phosphorylation of S-HDAg at Ser177 can work as a switch in HDV antigenomic RNA replication from the initiation to the elongation stage (Hong and Chen, 2010).

Acetylation of Lys72 on S-HDAg, by host p300 acetyltransferase, is thought to regulate nucleocytoplasmic shuttling of viral RNA (Huang et al., 2008b; Mu et al., 2004). Note that this amino acid is within the NLS of the HDAgs (Alves et al., 2008). Thus such a modification could be expected to have an impact on nuclear import. Acetylation of S-HDAg, has also been suggested to function as a switch in the synthesis of the different viral RNA species as this PTM was reported to be essential for HDV genome and mRNA synthesis but dispensable for antigenomic RNA synthesis (Tseng et al., 2008).

Methylation of Arg13 on S-HDAg, by protein arginine methyltransferase I, has been observed *in vitro* and has also been proposed to have a switching effect on HDV RNA replication (Li et al., 2004; Tseng et al., 2008). The studies were performed with S-HDAg with an R13A mutation, which failed to be methylated *in vitro*. In transfected

cells, the mutant S-HDAg reduced genomic RNA synthesis and almost completely suppressed HDV mRNA synthesis (Li et al., 2004; Tseng et al., 2008).

Finally, sumoylation of multiple lysine sites, by small ubiquitin-related modifier isoform 1 (SUMO1), has been reported. Such PTM was detected on S-HDAg but not on L-HDAg (Tseng et al., 2010). And this PTM was proposed to enhance genomic RNA and mRNA synthesis based on experiments where SUMO1 was fused to S-HDAg, so as to mimic sumoylated S-HDAg (Tseng et al., 2010).

Although the two delta antigens share sequence and functional domains they play very distinct roles in the HDV replication cycle. S-HDAg is essential for HDV RNA accumulation whereas L-HDAg acts as a dominant negative inhibitor of HDV replication (Chao et al., 1990) and also is essential for the assembly, via HBsAgs, of HDV RNA into new virus particles. There is, however, a common function attributed to both antigens: it has been observed that both can downregulate HBV replication (Williams et al., 2009).

Regarding the appearance of L-HDAg it is important to recall that because of the accumulated editing of the HDV antigenome, the proportion of this form, in relation to the total amount of accumulated HDAgs, increases during the replication cycle from 0% to around 30%, (Wong and Lazinski, 2002). This is sufficient to suppress replication but allow the accumulation of viral genomes that can then be packaged into new infectious particles with the help of L-HDAg. The NES present in L-HDAg allows the viral RNP to be exported from the nucleus to the cytoplasm for packaging (Lee et al., 2001). HDV RNPs then interact with HBsAgs at the endoplasmatic reticulum to form new infectious virions, which are then secreted to propagate further rounds of HDV infection (Tavanez et al., 2002). Such assembly of new virions only occurs when HBsAgs are present, otherwise, the viral RNPs return to the nucleus (Tavanez et al., 2002).

1.3.1. Small Delta Antigen

The objective of my research has been to increase the understanding of the structure and function of S-HDAg, the one viral protein that is essential for the accumulation of processed HDV RNA transcripts. Therefore, in this Section I will review, in more detail, prior knowledge and uncertainties relating to this protein.

Three groups have purified recombinant S-HDAg, but have been unable to go on and produce distinct crystals. Thus, at this time a detailed molecular structure for the protein is unavailable. A region comprising amino acids 12 to 60 was predicted to have alpha-helical structure and also be responsible for the protein's ability to form dimers (Xia and Lai, 1992; Wang and Lemon, 1993). To test this, one lab obtained large amounts of the corresponding synthetic peptide (Rozzelle et al., 1995). As a follow-up to this synthesis, it was found that the peptide readily formed crystals, leading to a successful molecular structure (Zuccola et al., 1998).

In this structure dimers were observed. More specifically, the largely alpha-helical peptides formed dimers by an anti-parallel coiled-coil interaction; hence this region has been referred to as the coiled-coil domain or CCD. Furthermore, the crystal structure indicated how such dimers might interact to form higher multimers. To test this hypothesis, the authors expressed and purified full-length recombinant S-HDAg, and showed, by prior cross-linking followed by mass spectrometry, that complexes as high as 8-mers could be detected (Zuccola et al., 1998).

Earlier studies had demonstrated that the ability to form dimers was essential for S-HDAg's role in promoting HDV RNA accumulation (Lazinski and Taylor, 1993). The molecular structure of the CCD was used to predict amino acids that might be important in the dimer and/or multimer interactions. Some of these predictions were tested and it was confirmed that there are amino acids, which are critical for dimer formation and, when mutated, affect S-HDAg's ability to support accumulation of HDV RNAs (Moraleda et al., 2000).

S-HDAg can form high molecular weight complexes within infected cells. This was first shown for infected liver tissue by sedimentation with prior cross-linking (Wang and Lemon, 1993). It was subsequently shown for transfected tissue culture cells, both in the presence and absence of HDV RNA (Chang et al., 2008).

Several different approaches have been used to demonstrate how S-HDAg binds to RNA. In some of these studies the binding was considered to be specific for HDV rod-like RNA structure (Chao et al., 1991; Defenbaugh et al., 2009; Lin et al., 2010). One study showed a minimum requirement of 311 nucleotides in a rod-like folding (Defenbaugh et al., 2009). In addition, multimerization of the S-HDAg was also required (Lin et al., 2010). In these two studies, the researchers examined purified recombinant S-HDAg that was C-terminally truncated to become HDV RNA specific and did not report the same assays with full-length S-HDAg. An earlier study, performed with S-HDAg fusion proteins, also reported that the HDV RNA rod-like structure is a prerequisite for the HDV RNA to be recognized by the antigen (Chao et al., 1991). A different group observed that the removal of the dimerization domain did not prevent S-HDAg from binding to the viral RNA (Lin et al., 1990). Thus, while it is agreed that S-HDAg is an RNA binding protein there remains controversy as to whether it is an HDV RNA specific interaction or if multimerization of S-HDAg is a prerequisite for the interaction.

As summarized in Table 1, several roles have been attributed to S-HDAg during the HDV life cycle.

Table 1: Roles attributed to S-HDAg. Putative and observed functions of S-HDAg during the HDV replication cycle.

Function	Observations
Achieve nuclear import of HDV RNAs	S-HDAg-mediated nuclear import of HDV RNA has been observed *in vivo* (Chou et al., 1998; Tavanez et al., 2002)
Facilitate ribozyme cleavage (chaperone)	S-HDAg stimulates HDV RNA ribozyme cleavage *in vitro* (Wang et al., 2003)
Regulate HDV RNA editing	S-HDAg expression in transfected cells suppresses editing of HDV antigenomic RNA (Polson et al., 1998)
Facilitate accumulation of processed HDV RNA transcripts	S-HDAg is essential for the accumulation of full-length HDV RNAs in transfected cells (Chao et al., 1990)
Regulate transcription by pol II	Enhanced elongation by pol II *in vitro* although it might be limited to 3'-OH end additions (Yamaguchi et al., 2001)

1. Introduction

S-HDAg is present in the virions forming viral RNPs with the HDV genome. One of the first tasks it performs is the transport of the viral genome into the nucleus of infected cells, where RNA-directed RNA synthesis takes place. This transport is achieved by the presence of the previously described NLS and RBD. Nuclear import may be facilitated by karyopherin 2α, since this importin interacts with S-HDAg *in vitro* (Chou et al., 1998).

Another role attributed to S-HDAg is the regulation of HDV RNA editing, particularly the deamination by ADAR-1. This editing seems to occur at multiple locations on HDV RNAs, but it is focused on the antigenomic RNA at the stop codon adenosine (Polson et al., 1998). S-HDAg has been found to suppress editing at this stop codon when expressed in transfected cells at levels close to those observed during HDV replication (Polson et al., 1998). This observation suggests that the antigen plays a role in limiting HDV RNA editing, as excessive editing has been shown to inhibit HDV RNA accumulation (Jayan and Casey, 2002).

It has been known for more than two decades that the small form of the delta antigen is essential for the accumulation of processed HDV RNAs (Chao et al., 1990). Several theories have been proposed for the precise role(s) it may play, as will be discussed ahead.

S-HDAg has been shown to interact with host pol II. In a pull-down assay, both S-HDAg and L-HDAg fused with a glutathione S-transferase tag were able to bind pol II from HeLa nuclear extracts (Yamaguchi et al., 2001). In the same study S-HDAg, was observed to enhance pol II elongation, presumably by displacing the subunit A of the negative elongation factor (NELF-A). S-HDAg was thus reported as an elongation enhancer of DNA-templated pol II transcription *in vitro* (Yamaguchi et al., 2001). However, the observed enhancement appears to be limited to 3'-OH end additions, rather than transcription. In a subsequent study Yamaguchi *et al.*, reported that S-HDAg functionally interacts with pol II suggesting that S-HDAg may be involved in facilitating the uncommon RNA-directed synthesis by an RNA polymerase that is normally DNA-directed (Yamaguchi et al., 2007). They proposed that the interaction

between pol II and S-HDAg loosens what, from molecular structure studies, is considered to be a pol II clamp, thereby reducing transcriptional fidelity and allowing the recognition of the atypical RNA template.

Amidst all the reports that S-HDAg actively participates in HDV RNA transcription there is a contradictory result showing that the presence of S-HDAg is not required for the accumulation of processed short HDV transcripts, although full-length transcripts, genomic or antigenomic, do require S-HDAg, or even L-HDAg (Lazinski and Taylor, 1994). As an explanation it was proposed that full-length HDV RNAs are susceptible to nucleolytic degradation in the absence of S-HDAg and, due to their size, such RNAs are more prone to be degraded than smaller RNAs. In other words, S-HDAg interacts with HDV RNAs to protect them and thereby allow their accumulation in infected cells.

Another role attributed to S-HDAg is that of HDV RNA chaperone. *In vitro* studies have reported that S-HDAg can stimulate HDV RNA ribozyme activity (Wang et al., 2003). From such studies it is inferred that *in vivo* S-HDAg may be directly involved in post-transcriptional processing of nascent multimeric transcripts by enhancing cleavage into unit-length molecules. It should be noted however, that the abovementioned studies of Lazinski and Taylor indicate that *in vivo*, HDAg is not directly needed for ribozyme cleavage and subsequent ligation (Lazinski and Taylor, 1994).

S-HDAg may also be involved in deviating/re-directing other host cell components to facilitate HDV RNA replication. As will be discussed subsequently, S-HDAg is a rather promiscuous protein in that many cellular partners have been detected.

1.3.2. Host Interactions

HDV has a very small RNA genome, as mentioned earlier, and encodes only one viral protein, HDAg. Albeit the fact that a second isoform of the HDAg appears later in the replication cycle, S-HDAg and L-HDAg are not sufficient for HDV to complete its replication cycle. None of these two HDAg isoforms has replicase activity or any other known enzymatic activity and, despite the fact that HDAgs are essential, HDV must rely extensively on host cell factors to complete its replication cycle. Many S-HDAg-interacting proteins have been identified through different affinity approaches, as summarized in Table 2. But, as will be discussed below, the role of most of these interactions remains unclear and is still being investigated.

Table 2: Identification of S-HDAg-interacting proteins. Summary of the proteomic studies to identify host proteins that interact with S-HDAg.

Reference	Experimental Approach	Number of Identified Proteins
Cao et al., 2009	Immunopurification followed by mass spectrometry	> 100
Casaca et al., 2011	Yeast two-hybrid screening	30

A comprehensive study using immunopurification followed by mass spectrometry identified over 100 host proteins that associated with a tagged S-HDAg (Cao et al., 2009). This set included 9 of the 12 subunits of the pol II complex, further supporting the idea that pol II is involved in HDV RNA transcription (Cao et al., 2009). In another study, a yeast two-hybrid approach identified 30 proteins encoded by a human liver cDNA library that interacted with S-HDAg (Casaca et al., 2011). Only three proteins from this study had also been identified by the previously mentioned immunopurification approach.

Some of the S-HDAg-interacting proteins that have been more thoroughly studied are described in Table 3. These include the previously mentioned host enzymes responsible for post-translational modifications.

Table 3: S-HDAg-interacting proteins. Summary of S-HDAg-interacting proteins with putative or observed function (adapted from Greco-Stewart and Pelchat, 2010).

Host Protein	Function	Reference
Casein kinase II	Phosphorylation of Ser2 observed *in vivo* as inhibition of casein kinase II suppresses Ser2 phosphorylation	Yeh et al., 1996
Double-stranded RNA-activated protein kinase R	Phosphorylation of Ser177, Ser180 and Thr182 observed *in vitro* and *in vivo*	Chen et al., 2002
ERK1/2	Phosphorylation of Ser177 observed *in vitro*, and *in vivo* activation of endogenous ERK1/2 increases Ser177 phosphorylation	Chen et al., 2008
Arginine methyltransferase 1	Methylation of Arg13 observed *in vitro*	Li et al., 2004; Tseng et al., 2008
p300 acetyltransferase	Acetylation of Lys72 observed *in vivo*	Mu et al., 2004
SUMO 1	Sumoylation of multiple lysine sites observed *in vitro* and i*n vivo* by co-transfection with S-HDAg-expressing plasmid	Tseng et al., 2010
Karyopherin 2α	Nuclear import of viral RNPs (interacts *in vitro* with S-HDAg)	Chou et al., 1998
Pol II	S-HDAg enhances pol II elongation *in vitro* but might be limited to 3'-OH end additions	Yamaguchi et al., 2001

Some of the S-HDAg-interacting proteins may be relevant for its cellular localization and may influence the role played by the viral protein. It has been suggested that the delta antigens interact with abundant nucleolar proteins C23, nucleolin, and B23, nucleophosmin (Huang et al., 2001; 2008a). A nucleolar localization signal has even been attributed to the delta antigens (Huang et al., 2008a). The presence of S-HDAg in the nucleolus of infected cells may be highly relevant, in that it has been proposed that antigenomic HDV RNA is synthesized by host polymerase I, a nucleolar protein (Modahl et al., 2000). S-HDAg may even be responsible for the transport of HDV genome to the nucleolus so that it can be transcribed. However, the biological relevance of S-HDAg nucleolar localization has not been ascertained.

1.4. Intrinsically Disordered Proteins[1]

The notion that a rigid 3-Dimensional (3-D) structure is a prerequisite for a protein to be functional was unquestioned for a long time. The paradigm can be dated back to the lock-and-key hypothesis, proposed in 1894 by Emil Fischer, to explain enzymatic specificity (Lemieux and Spohr, 1994). This concept was later validated as the crystal structures of proteins were beginning to be solved by X-ray diffraction (reviewed by Uversky, 2011a). So, for a long time, the conventional view was that a functional protein folds into a unique and stable 3-D structure, perfectly matching the substrate to which it should bind.

Up until a couple of decades ago, the idea that disordered proteins could also have specific functions was considered outlandish. Nevertheless, occasionally, flexible but functional proteins were discovered or re-discovered. Information on these flexible proteins was scarce and shallow, since they didn't fit the structure-function paradigm; they were considered mere exceptions to a rule. Nonetheless, throughout the years, a variety of names were given to these non-conventional proteins: partially folded, flexible, pliable, chameleon, vulnerable, natively unfolded, etc. (reviewed by Uversky, 2011a).

Only in the late 1990s did researchers start to realize that these unstructured proteins were representative of a broad class of rather important proteins (Dunker et al., 2001; Romero et al., 1998; Tompa, 2002; Uversky et al., 2000; Wright and Dyson, 1999). In recent years, the term intrinsically disordered proteins (IDPs) has become the most widely accepted, and intrinsically disordered regions (IDRs), to define proteins or protein segments that are biologically functional although they exist as collapsed or extended mobile conformational ensembles (Uversky, 2011a).

The increasing number of experimentally characterized IDPs led to the creation, in 2007, of DisProt, a database of disordered proteins (Sickmeier et al., 2007).

[1]This Section has been adapted from a review presented in Appendix.

1.4.1. Intrinsic Disorder Prediction

Like ordered proteins, whose structures can to some extent be inferred from their amino acid sequence, IDPs and IDRs have signature characteristics that allow the prediction of disorder based on sequence data alone. A mark of probable intrinsic disorder is a low content of hydrophobic amino acids, which usually form the core in folded proteins, and a high presence of polar amino acids conferring high net charge to the disordered protein or disordered region (Romero et al., 2001; Uversky et al., 2000). Low hydrophobicity is thought to lead to a low driving force for protein compaction and high net charge may result in strong electrostatic repulsion. Hence, these features contribute to structural disorder. Even by 1978, it had been suggested that IDPs and IDRs have amino acid compositions that differ from ordered proteins and therefore disorder could be predicted by the abnormally high ratio of charged residues by hydrophobic residues (Williams, 1978). However, this early mode of intrinsic disorder prediction was based on a very small set of proteins and never tested for other sets. Later it was shown that IDPs and IDRs are deficient in what has been called 'order-promoting' amino acids such as Ile, Leu, Val, Trp, Tyr, Phe, Cys and Asn; and are enriched in 'disorder-promoting' amino acids like Ala, Arg, Gly, Gln, Ser, Glu, Lys and Pro (Dunker et al., 2001).

IDPs and IDRs sequence characteristics were used to design algorithms to predict intrinsic disorder. In 1997 and 2000 were published the first well-tested predictors of intrinsic disorder (Romero et al., 1997; Uversky et al., 2000). Since then, more than 50 predictors have been developed to evaluate disordered regions based on amino acid sequences on a per-residue analysis (reviewed by He et al., 2009). Many predictors can be accessed via public servers such as the PONDR® predictors (Peng et al., 2006; Romero et al., 1997), FoldIndex (Prilusky et al., 2005), Dis-EMBL (Linding et al., 2003a), GlobPlot (Linding et al., 2003b) and PrDOS (Ishida and Kinoshita, 2007), to name a few. There are also available meta-predictors, such as PONDR-FIT (Xue et al., 2010a), which combines the output of six individual disorder predictors, and metaPrDOS (Ishida and Kinoshita, 2008), which is claimed to have an improved prediction accuracy. It is also available as a web metaserver, MeDor, which permits a fast simultaneous analysis by multiple predictors of a given sequence, retrieving an integrated view of the outputs (Lieutaud et al., 2008).

1.4.2. ID Prevalence

Disorder predictors have been applied to the predicted proteins of entire genomes to assess the extent of intrinsic disorder, with predictions spanning the three kingdoms. Studies have shown that IDPs and IDRs are not rare exceptions but highly abundant in all species. In fact, almost 70% of proteins in the Protein DataBase by 2003 had IDRs (Obradovic et al., 2003). IDPs are more common in eukaryota than in prokaryota and archea, with up to 30% of eukaryotic proteins being mostly disordered (Oldfield et al., 2005). However, in a recent study, comparing 194 eukaryotic and 87 prokaryotic proteomes, researchers found an overlap in the frequency of predicted disorder, which spans a wide range in both kingdoms (Pancsa and Tompa, 2012). Although prokaryotes were found to have a lower average disorder than eukaryotes both groups have a very broad range of predicted disorder, with scores of average ratio of disordered residues in proteins ranging from 0.12 to 0.35 for prokaryotes and 0.1 to 0.41 in eukaryotes (Pancsa and Tompa, 2012).

The same study also found the highest levels of predicted intrinsic disorder in single-celled protists, often higher than in more complex eukaryotic organisms. Hence, a new theory was proposed by the authors correlating intrinsic disorder with lifestyle and not only with the complexity of the organism (Pancsa and Tompa, 2012). They suggest that a low level of disorder reflects and adaptation to the environment, as low values were found for certain intracellular parasites and endosymbionts. Organisms, such as host-changing parasites, which lead a varied lifestyle, changing habitats, have higher levels of disorder (Pancsa and Tompa, 2012).

The amount of predicted intrinsic disorder in viral proteins has been compared with that of eukaryotes. Even though, eukaryotic proteomes contain more proteins with long disordered sequences, viral proteomes have more short disordered segments (Xue et al., 2010b).

Nevertheless, viruses have also been shown to have the widest spread of proteome disorder ranging from 7.3% of disorder promoting residues in human coronavirus up to 77.3% in avian carcinoma virus (Xue et al., 2012). Viral proteins display a high propensity for intrinsic disorder, as they tend to have reduced amounts of hydrophobic residues and a high proportion of polar amino acid residues (Xue et al.,

2010b). In particular, RNA viruses, which display the highest mutation rates, also have a high incidence of disordered regions and significantly lower van der Waals contact densities, reflecting the intensity of the 3-D interaction network of a protein (Tokuriki et al., 2008). As will be examined in section 1.4.4, viral proteins profit at different levels from the flexibility that results from intrinsic disorder.

1.4.3. Biophysical Characterization of Disorder

IDPs are by definition, an ensemble of conformational states, so the determination of one unique structure is not possible and a multiparametric approach is required. A dynamic structural characterization of an IDP relies on different parameters from a host of physiochemical methods to obtain information on different aspects of the protein such as overall compactness, conformational stability, presence of residual secondary structure, transient long-range contacts, regions of more or less mobility (Uversky and Dunker, 2012). Most of the techniques used were initially developed to analyze ordered proteins thus, data analysis needs to be careful as results for intrinsic disorder are usually due to the lack of signals characteristic of ordered sequences. That is also why several methods should be used and results compared to minimize ambiguity (Uversky and Dunker, 2012). Some techniques will be briefly summarized below, of which, the most commonly used to obtain data describing IDP structures are NMR and small-angle X-ray scattering (Tompa, 2011).

(i) X-ray crystallography is the ‘classic’ technique for determining a protein’s crystal structure. With this method the higher flexibility of the atoms in disordered regions results in a non-coherent X-ray scattering, making them invisible. That is, the outcome is a region with missing electron density corresponding to the disordered segment (Dunker et al., 2001). This method depends on the ability of proteins to form crystals, and highly disordered proteins in general, are not even able to be crystallized.

(ii) An NMR spectroscopy approach is commonly used. It provides information on a residue-by-residue basis although supplying little information on the overall shape of the protein. 1-Dimensional NMR is limited to small molecules while heteronuclear multidimensional NMR is preferable for larger proteins. The latter method can be used

to obtain precise data on 3-D structures and also provide direct measurements of IDR mobility (Uversky and Dunker, 2010).

(iii) Circular dichroism (CD) in the far-UV region provides estimates of secondary structure and near-UV CD displays sharp peaks for aromatic groups when the protein is ordered. CD lacks residue-specific information and the data obtained are less clear when a protein contains both ordered and disordered segments (Dunker et al., 2001).

(iv) The level of protein compactation or hydrodynamic dimension can be assessed by different techniques such as gel-filtration, viscometry, small-angle X-ray scattering or sedimentation (Uversky and Dunker, 2012).

(v) Small-angle X-ray scattering can also be used to determine the degree of globularity, providing information on the presence or absence of a tightly packed core (Uversky and Dunker, 2012).

(vi) Proteolytic degradation is also a method to identify disordered segments in a protein since flexibility can be a major determinant for susceptibility to cleavage (Johnson et al., 2012). Typically a structured protein needs a segment of more than 10 residues to be unfolded to reveal a proteolytic cleavage site (Dunker et al., 2001).

Note that many IDPs and IDRs adopt a well-defined structure when bound to their partners (Dyson and Wright, 2002). In these cases, by using methods to characterize ordered proteins, the structure of the bound protein can be readily solved.

1.4.4. Viral IDPs

Viral proteins should be considered as a unique group of proteins. Viruses must adapt faster than their hosts. They need to survive in their host's environment as well as inside the host, all the while evading host defense mechanisms, such as innate and adaptive immune responses. In order to adapt, viral genomes are subject to relatively high mutation rates, ranging from 10^{-5} up to 10^{-3} nucleotide exchanges per generation for RNA viruses and 10^{-8} to 10^{-5} for DNA viruses (Drake et al., 1998). The higher mutation rate found in RNA viruses probably reflects the lack of RNA repair mechanisms. In comparison, the genomes of bacteria and eukaryotes have on average, a

mutation rate of 10^{-9} (Drake et al., 1998). Since viruses have highly compact genomes, often with overlapping reading frames, a single mutation can have an impact on more than one viral protein (Reanney, 1982). Finally, viral proteins usually need to perform numerous interactions with host cell components, during the different steps of the virus life cycle from entry to replication to formation and exit of new infectious particles. Viral proteins must interact with host membranes, host proteins and in some cases, host nucleic acids even though viral proteins are often phylogenetically separated from their host proteins (Forterre, 2006; Reanney, 1982).

All these features make it extremely interesting to test whether viral proteins are also associated with unique biophysical characteristics and advantages. Intrinsic disorder may be a way for viral proteins to cope with these distinctive circumstances as the resulting plasticity can confer a number of exceptional functional advantages.

(i) Intrinsically disordered proteins are more flexible and without a rigid compact structure; thus, viral proteins can be highly promiscuous and take part in several interactions with multiple partners.

(ii) IDRs in particular can act as flexible linkers between functional domains enabling mechanisms that will facilitate binding and promiscuity.

(iii) These flexible linkers can also help viral proteins to elude the host cell's adaptive immune system by making it difficult for epitopes to be recognized (Goh et al., 2008).

(iv) Disorder can also be a way to cope with high mutation rates, characteristic to viruses. High flexibility, resulting from low interactions between amino acids, can be linked to a high adaptability and represent a strategy to buffer deleterious effects of mutations; an already unstructured protein has less to lose from a substitution than a highly structured one, as it is already unfolded.

Intrinsic disorder or conformational flexibility affords a "broad partnership" enabling the disordered proteins to adapt to and interact with several distinct partners. One IDR can bind multiple partners by gaining very different structures (Oldfield et al., 2008). IDPs can perform different interactions: they can be involved in highly stable

complexes or in signaling interactions in which they transit between the bound and unbound state as a dynamic and sensitive on-off switch (Uversky, 2011b). IDPs' ability to have different conformations depending on environmental conditions allows them to exercise different functions in different contexts.

Binding promiscuity is a key characteristic for viral proteins. Even though some viruses have a genome that codes several proteins, it is usually not sufficient and they require host cell machinery to complete their life cycles. As mentioned before, viruses have very compact genomes and by having viral proteins with IDRs or even the entire protein disordered, a single protein can be involved in different tasks by interacting with different partners.

1.5. Final Remarks

S-HDAg is a nuclear protein and such subcellular localization can be related to some of the roles attributed to S-HDAg during HDV replication (Table 1). S-HDAg has been observed both in the nucleolus and in the nucleoplasm with results varying with the experimental conditions (Alves et al., 2008; Bichko and Taylor, 1996; Lee et al., 1998). On the one hand, S-HDAg has been reported to be directly involved in HDV replication, namely by interacting with host pol II, which locates in the nucleoplasm (Yamaguchi et al., 2001). On the other hand, S-HDAg is known to interact with nucleolar proteins such as B23 and C23 (Huang et al., 2001; 2008a) and it has been suggested that host pol I, a nucleolar protein may be responsible for HDV antigenomic RNA synthesis (Modahl et al., 2000). Thus, it can be hypothesized that the different subcellular localization patterns of S-HDAg are related with different phases of HDV replication. By analyzing the co-localization of S-HDAg with pol II and nucleolar proteins we will try to assess the relevance of S-HDAg subcellular localization in the presence and absence of HDV RNA.

HDV has a very limited coding capacity: just one protein, HDAg, which can exist in two isoforms. Thus it can be hypothesized that they are likely to be multi-purpose proteins, that is, they are likely to be involved in more than one step of the viral replication cycle. S-HDAg has already been described as a promiscuous protein, interacting with many host proteins (Cao et al., 2009; Casaca et al., 2011; Greco-Stewart and Pelchat, 2010). This adaptability of S-HDAg to different partners and other characteristics suggest that it has some degree of structure flexibility or intrinsic disorder. Furthermore, no one has been able to obtain crystals of full-length S-HDAg suggesting again the lack of a rigid structure. In fact, S-HDAg has several characteristics attributed to intrinsically disordered proteins besides binding promiscuity and inability to form a crystal structure (Uversky, 2011). IDPs are commonly nucleic acid binding proteins, displaying chaperone activity, just like S-HDAg. High net charge is also common in IDPs and HDV has an estimated net charge of +12 (Kuo et al., 1988). Multimerization ability is yet another feature of IDPs that is present in S-HDAg. Based on these attributes we hypothesize that S-HDAg has some level of intrinsic disorder and propose to test this hypothesis both *in silico* and *in vitro*.

S-HDAg multimerization and its interaction with nucleic acids are two fundamental characteristics for some of the functions already attributed to S-HDAg (Table 1). These two features rely on the protein's coiled-coil and RNA binding domains and have been studied by different groups, using different forms of recombinant S-HDAg (as discussed in Section 1.3.2.). It has been suggested that S-HDAg needs to multimerize in order to bind HDV RNA (Lin et al., 2010). And also, that the binding is specific for HDV RNA, requiring a minimum length and rod-like structure to take place (Defenbaugh et al., 2009). As these studies were performed with a C-terminally truncated S-HDAg we propose to test this features in the context of the full-length S-HDAg. As such, we will test the specificity of the RBD in the full-length S-HDAg. Also we will assess the importance of multimerization in RNA binding using a N-terminally truncated S-HDAg lacking the CCD.

2. Specific Aims

The following studies were undertaken to better understand three important aspects of the essential role(s) of the hepatitis delta virus small antigen, S-HDAg:

1. Evaluate sequence features affecting the subcellular localization of S-HDAg, particularly in the context of HDV replication, along with the significance and relevance of S-HDAg co-localization with host factors.
2. Further examine structural features of purified S-HDAg, with an evaluation of the extent of intrinsic disorder, and the part it might play in the role(s) of the protein.
3. Evaluate the role of S-HDAg multimerization in nucleic acid binding ability, both for the full-length S-HDAg and for a truncated form lacking the putative multimerization domain.

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

JOURNAL OF VIROLOGY, July 2009, p. 6457–6463 Vol. 83, No. 13
0022-538X/09/$08.00+0 doi:10.1128/JVI.00008-09

Intracellular Localization of Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation[▽]

Ziying Han, Carolina Alves, Severin Gudima, and John Taylor*

Fox Chase Cancer Center, Philadelphia, Pennsylvania

Received 3 January 2009/Accepted 11 April 2009

Hepatitis delta virus (HDV) encodes one protein, hepatitis delta antigen (δAg), a 195-amino-acid RNA binding protein essential for the accumulation of HDV RNA-directed RNA transcripts. It has been accepted that δAg localizes predominantly to the nucleolus in the absence of HDV genome replication while in the presence of replication, δAg facilitates HDV RNA transport to the nucleoplasm and helps redirect host RNA polymerase II (Pol II) to achieve transcription and accumulation of processed HDV RNA species. This study used immunostaining and confocal microscopy to evaluate factors controlling the localization of δAg in the presence and absence of replicating and nonreplicating HDV RNAs. When δAg was expressed in the absence of full-length HDV RNAs, it colocalized with nucleolin, a predominant nucleolar protein. With time, or more quickly after induced cell stress, there was a redistribution of both δAg and nucleolin to the nucleoplasm. Following expression of nonreplicating HDV RNAs, δAg moved to the nucleoplasm, but nucleolin was unchanged. When δAg was expressed along with replicating HDV RNA, it was found predominantly in the nucleoplasm along with Pol II. This localization was insensitive to inhibitors of HDV replication, suggesting that the majority of δAg in the nucleoplasm reflects ribonucleoprotein accumulation rather than ongoing transcription. An additional approach was to reevaluate several forms of δAg altered at specific locations considered to be essential for protein function. These studies provide evidence that δAg does not interact directly with either Pol II or nucleolin and that forms of δAg which support replication are also capable of prior nucleolar transit.

Hepatitis delta virus (HDV) is a subviral agent with a single-stranded circular RNA genome. As reviewed recently (6, 14, 37) this 1,679-nucleotide RNA is replicated by RNA-directed RNA synthesis dependent upon redirection of the host RNA polymerase II (Pol II). Replication involves the production of the antigenome, an exact complement of the genome. The small delta antigen (δAg) is translated from a third RNA, one with the same antigenomic polarity but with a linear rather than circular conformation, of less than full-length, and possessing a 5′ cap and a 3′ poly(A). δAg is a 195-amino-acid-long RNA binding protein that is essential in one or more ways for the accumulation of HDV RNA-directed RNA transcripts. There is a second form of δAg that is 19 amino acids longer at the C terminus; this species arises due to specific RNA editing that occurs on the antigenomic RNA during replication. Unlike the small form, the large does not support HDV genome replication and is a dominant-negative inhibitor of the replication driven by the small form. However, large δAg does have an essential function in that in the presence of HBV envelope proteins, it facilitates the assembly of HDV genomic RNA into new virus particles.

Our previous immunostaining studies have shown that in the absence of HDV genome replication, small δAg localizes predominantly to the nucleolus, with some staining in the nucleoplasm (4). It has been reported that δAg moves to the nucleus because of a bipartite nuclear localization signal (NLS) (21); however, a more recent study shows that a single short sequence (positions 66 to 75) is sufficient to direct a marker protein to the nucleus, where it localizes to the nucleoplasm (2). Nucleolar localization of δAg has been attributed to a nucleolar localization signal (24). It has been suggested that δAg binds directly to nucleolin (24), also known as C23, and to nucleophosmin, also known as B23 (17). The present study includes examination of the nature and significance of the δAg localization to the nucleolus. Other investigators have noted that for several viruses, specific and essential viral proteins, typically RNA-binding proteins, also locate to the nucleolus, and in some cases such localization only occurs early in infection, leading to the speculation that a "nucleolar transit" may be an essential step in genome replication (12, 15). Consistent with but not proof of such a transit for δAg is that during HDV genome replication, the majority of δAg is no longer found in the nucleolus but in the nucleoplasm (4). Located within the nucleoplasm is the majority of the host RNA Pol II, and so it has been tempting to suggest that colocalization of δAg might be an indication of HDV RNA-directed transcription complexes (8). However, the present studies indicate that even though the majority of the δAg is in the nucleoplasm, it is not in association with active HDV RNA-directed transcription by Pol II but rather, with the accumulation of completed transcripts. Furthermore, even nonreplicating HDV RNAs can move δAg from the nucleolus to the nucleoplasm.

Along with other investigators we have noted that during HDV replication, there arise mutated forms of δAg. There is the large δAg that arises via editing. However, there are other forms, some of which might support infection or coexist in the presence of forms that do (3, 4, 13). Such mutated forms

* Corresponding author. Mailing address: Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497. Phone: (215) 728 2436. Fax: (215) 728 3105. E-mail: john.taylor@fccc.edu.
▽ Published ahead of print on 15 April 2009.

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

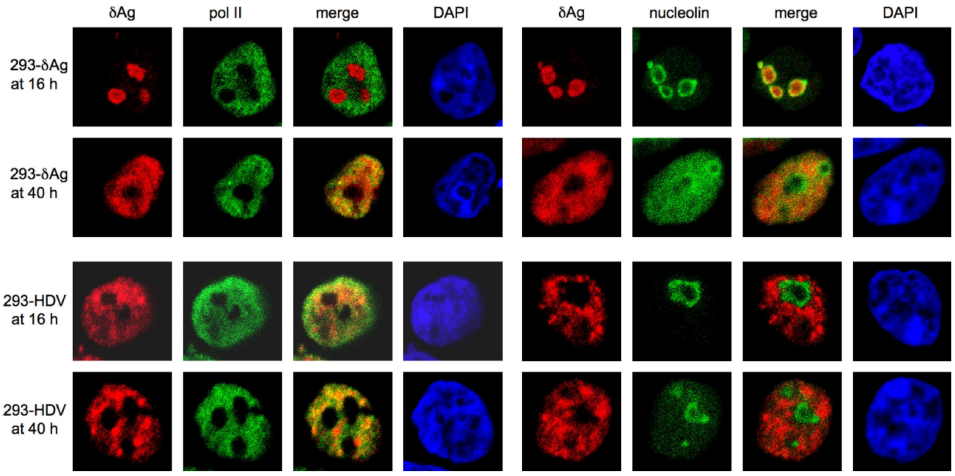


FIG. 1. δAg distribution during induction of 293-δAg and 293-HDV cells. Cells were induced by the addition of TET and studied at 16 and 40 h by immunostaining and confocal microscopy. Detection of δAg is indicated in red while detection of nucleolin or Pol II (as indicated) is in green; DAPI counterstaining is in blue.

present a spectrum of different intracellular localizations; for example, they may be spread throughout the whole cell or located over the endoplasmic reticulum or associated with the splicing factor SC35 at speckle sites in the nucleoplasm. Therefore, for most of the present study, we focused on expressing the essential small δAg, without and with replicating or nonreplicating HDV RNAs, under conditions designed to prevent altered forms of the protein from arising. Then, with immunostaining procedures we examined the intracellular localizations of δAg relative to specific cellular components, such as nucleoplasmic Pol II and SC35 or nucleolar nucleolin. Our findings have implications for the hypothesis of prior nucleolar transit of δAg and for the significance of δAg in the nucleoplasm when HDV RNA replication occurs.

MATERIALS AND METHODS

Cells. Two inducible cell culture systems were as previously described (7). Briefly, we first established from TRex cells (Invitrogen) a clone, 293-δAg, expressing a single copy of δAg cDNA under tetracycline (TET) control. Next, HDV genome replication was initiated in these cells by transfection of HDV RNA that has a frameshift mutation and does not express δAg by itself. A clone derived from a single cell was identified by its high level of HDV replication and was designated as 293-HDV cells. Both 293-δAg and 293-HDV cells were maintained in Dulbecco's modified Eagle's medium with 10% fetal bovine serum and selective antibiotics, blasticidin and hygromycin (Invitrogen). Also used were 293T and Huh7 cells (30).

Plasmids and transfection. Vectors expressing the forms of δAg and HDV were as previously described (22, 23, 28). Transfection of 293T and Huh7 cells was via Lipofectamine 2000 (Invitrogen) according to the manufacturer's instructions.

Indirect immunofluorescence. Indirect immunofluorescence was performed as previously described (8). After paraformaldehyde fixation, we used the following primary antibodies: polyclonal rabbit anti-δAg and mouse monoclonal antibodies specific for nucleolin, Pol II (4H8), and splicing factor SC35. The mouse antibodies were obtained from Abcam. As secondary antibodies we typically used Alexa 488 chicken anti-mouse and Alexa 594 goat anti-rabbit (Invitrogen). Cells were stained with DAPI (4′,6′-diamidino-2-phenylindole) to visualize the nuclei and mounted with antifade mounting solution (Invitrogen). Confocal images were obtained using either a Zeiss 510 laser scanning microscope or a Nikon Eclipse TE2000-E. Optical sections of 0.1 μm were collected. In some cases Z-stacks at 0.25 μm were collected, and the intracellular distribution of stained proteins was determined using MetaMorph software.

RESULTS

Effect of increased time of TET induction on the intracellular localization of δAg in 293-δAg and 293-HDV cells. As described previously (7) and in Materials and Methods, both the 293-δAg and 293-HDV cell lines contain a TET-inducible cDNA for the small δAg. In addition, the 293-HDV cells contain copies of HDV RNA so that TET induction leads to increased transcription and accumulation of HDV RNAs. At 16 and 40 h after induction, the immunostaining patterns were determined as shown in Fig. 1.

For 293-HDV cells the δAg pattern was the same at both 16 and 40 h. It was predominantly nucleoplasmic, colocalizing with host RNA Pol II. We previously showed that HDV RNA accumulation reaches maximal values by 40 h (7). Thus, we consider that for the majority of the δAg in these cells, the nucleoplasmic distribution reflects association with the accumulation of processed HDV RNA species rather than with the process of RNA-directed transcription.

At 16 h the predominant distribution of δAg in 293-δAg cells was with host nucleolin, a major nucleolar protein, rather than with Pol II, a nucleoplasmic protein. A minority of cells gave staining in both nucleoli and nucleoplasm or in nucleoli only. After 40 h the predominant staining pattern was in nucleoplasm only. Quantitation of the three staining patterns is presented in Table 1.

For 293-δAg cells at 40 h the nucleoplasmic distribution of δAg seemed similar to that obtained for 293-HDV cells, and yet in this case it had nothing to do with HDV RNA-directed RNA transcription. We noted that the nucleolin distribution in 293-δAg cells also changed with time, from being nucleolar to more nucleoplasmic. This change also occurred in 293T cells that lack the δAg (data not shown). Thus, we considered that

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

TABLE 1. Time-dependent changes in intracellular distribution of δAg[a]

Cell line	Time after induction (h)	% of cells with specific intracellular δAg distribution in:		
		Nucleolus only	Nucleolus and nucleoplasm	Nucleoplasm only
293-δAg	16	86	10	4
	40	5	15	80
293-HDV	16	<1	<1	>99
	40	<1	<1	>99

[a] As described in the legend of Fig. 1, the two indicated cells types were induced by the addition of TET and at the indicated times were fixed and immunostained. At least 100 cells were examined for the three indicated patterns of intracellular distribution of δAg.

for both the δAg and the nucleolin, the change in localization might actually be associated with the time since the cells were seeded. As confirmation of this interpretation, we found that when induced cells were reseeded at subconfluent density and observed within 24 h, the majority of both δAg and nucleolin localized to the nucleolus (data not shown). In an earlier study of δAg expression in Huh7 cells at 6 days after transfection, we observed δAg to be predominantly nucleolar; however, such cells had been reseeded at subconfluent density within 24 h of observation (4).

Cellular stress changed the δAg localization of induced 293-δAg cells. Upon further study, we found two situations that quickly and reversibly changed the intracellular distribution of δAg and nucleolin in 293-δAg cells. At 16 h after induction cells were treated for 1 h with either actinomycin or 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole (DRB). A low concentration of actinomycin (10 ng/ml) is known for its ability to specifically block transcription by Pol I of rRNA precursors (31). DRB is a nucleoside analog known for several activities in animal cells. It inhibits kinases including casein kinase II and in this manner can interfere with Pol II transcription (44). It is also known to be a cause of nucleolar stress (18). Nucleolin is known to be hyperphosphorylated by kinases including casein kinase II (26). δAg is also reported to be phosphorylated by this kinase (42).

As shown in Fig. 2B and D, treatments for 1 h with actinomycin or DRB, respectively, caused the majority of the δAg to be detected in the nucleoplasm. The same treatments also caused much of the host nucleolin to be localized in the nucleoplasm. Furthermore, if the treated cells were then returned to normal growth conditions for 2 h, within each cell a significant fraction of both δAg and nucleolin were back in the nucleolus, as shown in Fig. 2C and E, respectively. A similar effect for DRB has been reported for a fusion of large δAg with green fluorescent protein (27, 35).

The studies shown in Fig. 2 support the interpretation that for induced 293-δAg cells, δAg, similar to what is known for nucleolin, was in the nucleolus because of binding to rRNA precursors. Also, the studies shown in Fig. 1, with the time-dependent movement of δAg and nucleolin to the nucleoplasm, are consistent with the maturation of the rRNA precursors and the release of δAg from the nucleoli. Nevertheless, while these interpretations might explain why the δAg could leave the nucleolus, they do not explain why the δAg (or nucleolin) was then relocated to the nucleoplasm rather than to another cellular compartment.

Expression of nonreplicating forms of HDV RNA can cause movement of δAg from the nucleoli to nucleoplasm. The above studies suggested that δAg might bind to rRNA precursors in the nucleolus. As a test of this interpretation, we examined the consequences of also expressing nonreplicating HDV RNAs since we might expect that δAg would prefer to bind to the latter. Huh7 cells, a human liver cell line, were transfected with a plasmid to express δAg (pDL444), without or with a plasmid to express an excess of full-length HDV RNA. The latter, plasmid pDL482, as previously described, was a chimera of genomic and antigenomic HDV RNA sequences and was incapable of replication even when provided with δAg (23). After 24 h cells were examined for the distribution of δAg and nucleolin.

For cells expressing δAg we detected only nucleolar localization (Fig. 3A). However, for about half of the cells coexpressing the HDV RNA, we detected two additional patterns: localization in nucleoli and nucleoplasm or in the nucleoplasm only. Table 2 gives the quantitation of this and additional experiments where we tested the consequences of expressing other RNAs. Another HDV RNA, lacking sequences from one end of the genomic rod-like structure, also caused redistribution of δAg.

Thus, nonreplicating HDV RNAs were able to cause movement of δAg from the nucleoli to the nucleoplasm. The results support the interpretation that, given the choice, δAg will ac-

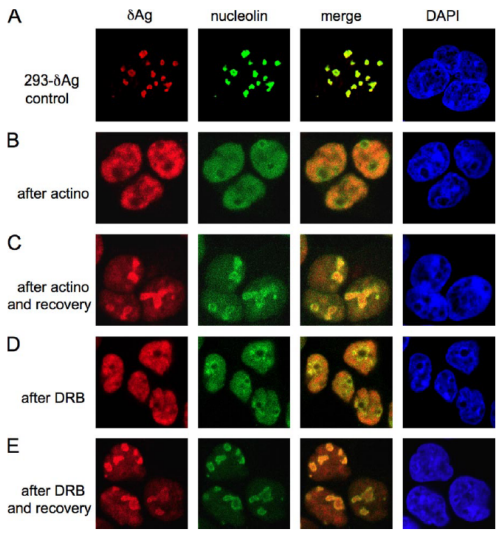


FIG. 2. Conditions of cell stress cause relocalization of δAg and nucleolin in 293-δAg cells. At 16 h after TET induction, cells were either untreated (A) or subjected to brief treatments (B to E) and then studied by immunostaining to detect δAg (red) or nucleolin (green); DAPI counterstaining is also shown (blue). The treatments were as follows: 10 ng/ml actinomycin for 1 h (actino; B), actinomycin followed by 2 h of recovery in the absence of actinomycin (C), 50 μM DRB for 1 h (D), and DRB followed by 2 h of recovery (E).

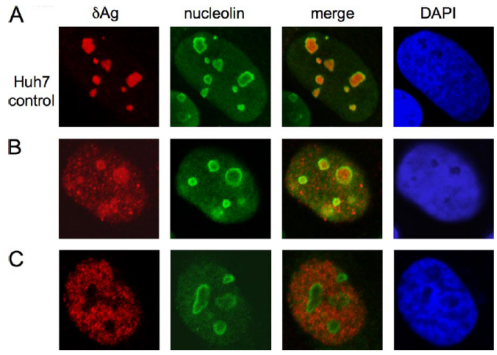


FIG. 3. Movement of δAg from nucleoli to nucleoplasm following expression of nonreplicating HDV RNA sequences. Huh7 cells were transfected with a plasmid (pDL444) to express δAg, without or with a plasmid to express a chimera of HDV RNA (pDL482). After 24 h, immunostaining was used to detect δAg (red) and nucleolin (green). Panel A shows the pattern detected when cells expressed δAg only. Panels B and C show two new patterns detected when cells also expressed the nonreplicating HDV RNA. Quantitation for the patterns is summarized in Table 2.

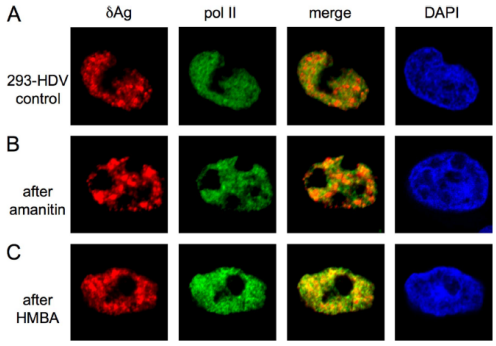


FIG. 4. Effect of amanitin and HMBA treatments on δAg distribution for induced 293-HDV cells. Panels A and B show 293-HDV cells at 16 h after TET induction, without and with an additional incubation with amanitin (10 μg/ml for 45 min). Panel C shows 293-HDV cells that were induced for 22 h in the presence of added HMBA (3 mM). Immunostaining was used to detect δAg (red) and Pol II or nucleolin (green); DAPI counterstaining is shown in blue.

cumulate on HDV RNAs even if they are nonreplicating rather than on rRNA precursors.

Treatment with amanitin or hexamethylene bisacetamide (HMBA) did not cause relocalization of δAg in induced 293-HDV cells. In contrast to 293-δAg cells, when induced 293-HDV cells were treated for 1 h with actinomycin or DRB, there was no detectable change in the δAg distribution even though there was movement of nucleolin from the nucleolus (data not shown). Furthermore, when the cells were briefly treated with amanitin (45 min at 5 μg/ml), an inhibitor of host RNA Pol II and known to be able to inhibit HDV RNA accumulation (9, 29), there was again no change in the distributions of δAg or of Pol II (Fig. 4A and B). Such findings provide more support for the interpretation that in induced 293-HDV cells, the distribution of the majority of δAg has nothing to do with HDV RNA transcription although it may have something to do with the accumulation of processed HDV RNAs. However this does not exclude the possibility that even after disruption of transcription, the intracellular localization of δAg and RNA remains nucleoplasmic.

It has been reported that δAg in vitro interacts with an essential transcription factor, pTEFb, a complex of CDK2 and cyclin T (40). Also, it is reported that in animal cells much of the Pol II can exist bound to specific templates but that in a paused state it is associated with a depletion of active factor pTEFb and, further, that this depletion is caused by the association of pTEFb to 7SK RNA in the nucleolus (43). One way to disrupt the nucleolar complex and release paused Pol II is by treating cells with HMBA (10, 32). We applied such a treatment during TET induction of 293-HDV cells. As shown in Fig. 4C, this caused no significant change in the intranuclear distribution of δAg or Pol II, and it had no effect on the extent of accumulation of HDV RNA (data not shown). This supports the interpretation that δAg in the nucleoplasm was not in a paused configuration with Pol II.

Altered forms of δAg show unexpected intracellular localizations. Figure 5 is a representation of features that have been described for the unmodified δAg, with alignment on the 195-amino-acid species used in our studies (19). At position 12 to 60 is a domain that was noted to have a potential alpha-helical

TABLE 2. Ability of expressed RNAs to change in intracellular distribution of δAg[a]

Type of RNA expressed	% of cells with specific intracellular δAg distribution in:		
	Nucleolus only	Nucleolus and nucleoplasm	Nucleoplasm only
None	>99	<1	<1
HDV chimera of genomic and antigenomic sequence	44	36	20
HDV fragment of genomic sequence	52	29	19

[a] As described in the legend of Fig. 3, Huh7 cells were transfected with plasmid (pDL444) to express δAg along with either a control plasmid (pSVL) or plasmids to express an HDV chimera (pDL482) or a 1,165-nucleotide HDV RNA fragment (pDL539). At 24 h the cells were fixed and immunostained. At least 100 cells were examined for the three indicated patterns of intracellular distribution of δAg.

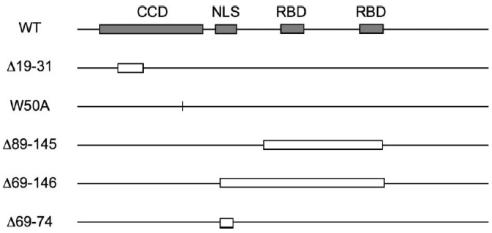


FIG. 5. Features of δAg. Shown is the localization on δAg of what are referred to as the CCD, core NLS, and the bipartite RBD. Also indicated are representations of the changes or deletions of five altered forms of δAg. These forms are as used in Fig. 6 and have been described previously (23, 28). WT, wild type.

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

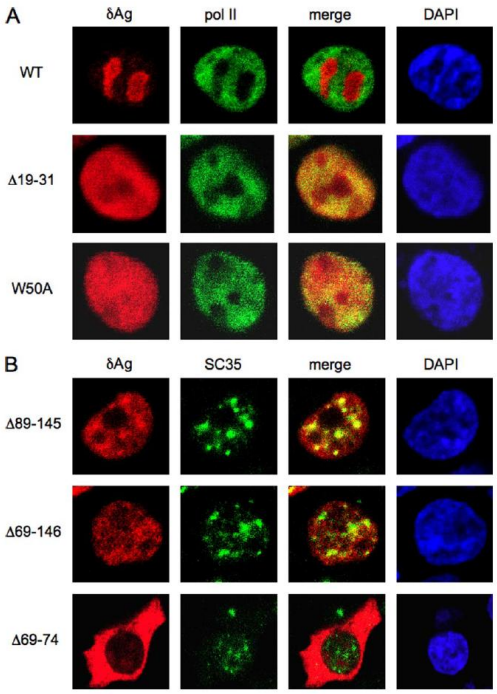


FIG. 6. Expression of altered forms of δAg in 293 cells. Plasmids encoding the wild type (WT; controls) and five altered forms, as indicated at left, of δAg were transfected into 293 cells. At 16 h cells were examined by immunostaining to detect δAg (red) and either Pol II or SC35 (green); DAPI counterstaining is shown in blue.

structure (45). Initial studies to determine why δAg located to the nucleus defined a bipartite nuclear localization signal. Subsequent studies have shown that just one such region at position 66 to 75 (indicated as an NLS) is sufficient (2). Other studies have determined two regions, at positions 97 to 107 and 136 to 146, that are rich in positively charged amino acids, which function as a bipartite RNA-binding domain (RBD) (20). Another study has reported that within the 12 amino acids at the N terminus is yet another RBD (39).

Also indicated in Fig. 5 are five altered forms of δAg that we have previously described (23, 28). These are representative of changes in the features of the unmodified wild-type δAg. These proteins were transiently expressed by transfection of 293 cells, and then the intracellular distributions were examined by immunostaining.

Figure 6A shows the results for two forms of δAg that were altered in the coiled-coil domain (CCD). The mutant with a deletion of residues 19 to 31 (Δ19–31) fails to dimerize (23, 28). The amino acid exchange mutant, W50A, is predicted to be important in the CCD structure, and yet experimentally dimerization can still be detected (28). Both altered forms fail to support HDV genome replication (28) and, as shown in Fig. 6A, localized to the nucleoplasm rather than to the nucleolus. Similar results were obtained with W20A that, like W50A, is predicted to be important for the CCD structure and yet does not prevent dimerization (28).

Figure 6B shows the results for a form, Δ89–145, altered by deletion that includes the two RBDs. A form with a larger deletion, Δ69–146, now including the NLS, also localized to the nucleoplasm. Neither of these species is capable of supporting HDV RNA accumulation (23). Since both species lack the two RBDs, it may be unlikely that they are binding to RNA species in the nucleoplasm. Some of the nucleoplasmic δAg seemed to colocalize to speckles containing the splicing factor SC35. Such speckles are considered to contain Pol II and are adjacent to paraspeckles, putative sites for the processing of Pol II RNA-directed RNA transcripts (5, 33). Shih et al. have previously shown that a green fluorescent protein fusion to the N-terminal 88 amino acids of small δAg primarily localizes to SC35 speckles (36).

Finally in Fig. 6B are shown results obtained with a form of δAg, Δ69–74, that has a deletion of most of the NLS (23). This protein was localized mainly to the cytoplasm. A small amount was detected in the nucleus, primarily colocalized with nucleoli. We were initially puzzled to find that this protein has some ability to support HDV transcription and accumulation of HDV RNA transcripts. Relative to the HDV RNA transcript accumulation achieved with unmodified δAg, such accumulation was about 20% (data not shown). Therefore, we used analysis of a series of Z-sections obtained by confocal microscopy to determine the fraction of the protein that was nuclear. It was determined to be about 30%. Thus, relative to unmodified δAg, the altered protein lacking an NLS led to about equal amounts of δAg nuclear accumulation and of HDV RNA accumulation.

DISCUSSION

This report focuses on the use of immunostaining procedures to study the associations of δAg in the presence and absence of HDV RNAs. The advantage of this approach is that if the cells are fixed prior to staining, one can be confident that observed associations are not artifacts of rearrangement, such as can occur during cell fractionation procedures. This is a concern in our previous studies of δAg that made use of fractionation followed by rate zonal sedimentation, immunoprecipitation, and gel electrophoresis under nondenaturing conditions (8, 11). Even when fixation is followed by fractionation, there are concerns about the interpretability of the complexes detected (8). However, even for studies using fixation followed by immunostaining, there is still the limitation that a colocalization of two components supports but is not sufficient to prove an interaction, whether direct or indirect interactions.

Most studies reported here made use of expression in 293 cells, but many aspects have been confirmed in Huh7 cells. Also, some of the results with the small δAg have been reproduced with large δAg. A difference was that when large δAg was relocalized to the nucleoplasm, it showed a greater tendency than small δAg to localize to SC35 speckles (data not shown).

There are three major questions addressed by this study. First, what are the nature and significance of δAg in the nucleolus? δAg, like nucleolin, is an RNA-binding protein, and we conclude that both proteins colocalize because they bind to

rRNA precursors (15); immunoprecipitation of δAg with nucleolin has been reported, but such studies did not exclude the possibility that the interaction was mediated by rRNA (24). Since δAg is nucleolar only in the absence of HDV RNA transcription and/or accumulation, one might argue that nucleolar localization is irrelevant to replication. An alternative interpretation is that δAg goes to the nucleolus before it is utilized to facilitate the accumulation of processed HDV RNA transcripts. If such a nucleolar transit does occur, this raises the question of whether such a transit is somehow required for the viral life cycle. It can be said that the only forms of δAg that support replication are ones which, expressed in the absence of HDV RNAs, will accumulate in the nucleolus. At the same time, altered forms, such as large δAg that does not support replication, will also go to the nucleolus. As suggested in the introduction, we know that the RNA binding proteins of many viruses, both RNA and DNA viruses, can localize to the nucleolus (15, 27). Furthermore, some studies consider nucleolar transit to be of relevance. For example, in the case of orthomyxoviruses there are studies suggesting that prior nucleolar localization is required for the nucleocapsid protein early in replication, which is followed by a combination of nucleoplasmic and nucleolar localizations (12, 34).

Second, what are the nature and significance of δAg found in the nucleoplasm? Several situations have been found in which δAg localized to the nucleoplasm. Not just replicating HDV RNA (Fig. 1) but also nonreplicating HDV RNA (Fig. 3) can cause nucleoplasmic accumulation of δAg. Also, nucleoplasmic accumulation can occur in the absence of detectable HDV RNA species. For example, at 40 h after induction δAg is primarily localized to the nucleoplasm (Fig. 1). Certain cell stress conditions lead to nucleoplasmic localization (Fig. 2B and D). Also, several altered forms of δAg localize to the nucleoplasm (Fig. 6A and B). In each of these situations we do not know what the δAg is binding to in the nucleoplasm. In some cases we can detect a component that colocalized with SC35 (Fig. 6B). This could reflect indirect rather than direct binding, but the localization apparently does not need δAg to possess either a dimerization or an RBD.

Third, does δAg bind, whether directly or indirectly, to nucleoplasmic Pol II? Some form of binding, direct or indirect, is plausible since it has been shown that Pol II is essential for at least some of the HDV RNA-directed transcription (8). There is a report that δAg can interact in vitro with one or two subunits of host Pol II (41) while a more recent study reconstituted Pol II transcription in the absence of δAg (1). Be that as it may, our previous studies using immunoprecipitation indicate that for the majority of δAg expressed within cells in the presence or absence of HDV RNA, there is only minimal association with Pol II (8). Furthermore, the immunostaining studies show that in the absence of HDV RNAs, the δAg is largely nucleolar while Pol II is nucleoplasmic (Fig. 1); thus, δAg is not binding to Pol II and is somehow binding to nucleolar components, which we believe are probably rRNA precursors. Contrary interpretations are that δAg in the nucleolus is binding to nucleolin (24) and B23 (17), that it has fewer posttranslational modifications than nucleoplasmic δAg (38), that δAg modified to contain a nucleolar localization signal supports the synthesis of antigenomic RNA but not genomic RNA (16), and that Pol I in the nucleolus carries out the transcription of antigenomic RNA (25).

In the presence of HDV RNAs the immunostaining results show that δAg and Pol II are in the same nucleoplasmic location, and it has been tempting to assert that they must be associated and participating in RNA-directed RNA synthesis (8). In contrast, after the present studies we would be more cautious. First, at 40 h after induced replication, when accumulation of processed HDV RNAs has reached a maximum value and transcription may have ceased (7), the same colocalization is observed (Fig. 1). Furthermore, when HDV transcription is inhibited with amanitin, the colocalization remains (Fig. 4). Therefore, we suggest that the majority of δAg that colocalizes with Pol II in the nucleoplasm is not currently involved with active RNA-directed transcription. In cells that have undergone HDV transcription, we have previously used immunoprecipitation and found at least 16% of the HDV RNAs in association with δAg (8). Thus, we expect that the majority of δAg in the nucleoplasm is in association with HDV RNAs in a postreplicative state. However, we do not know why such ribonucleoprotein complexes remain largely in the nucleoplasm. In addition, we have to remember that δAg can have a similar nucleoplasmic localization in the presence of HDV RNAs that cannot replicate (Fig. 3) or in the total absence of full-length HDV RNAs (Fig. 1 and 2), and so also will certain altered δAg forms that lack an RBD or a dimerization domain (Fig. 6A and B).

In summary this study used immunostaining of cells under defined conditions of δAg expression, without and with the accumulation of associated HDV RNA species, to examine the intracellular localizations of δAg and its colocalizations with host components. The findings reveal the complications of extrapolating from such localizations to interpretations of functional significance.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants AI-26522 and CA-06927 and by an appropriation from the Commonwealth of Pennsylvania. C.A. was a recipient of a Fundação para a Ciência e Tecnologia Ph.D. grant.

Some confocal imaging was performed at the Biomedical Imaging Core, Department of Pathology and Laboratory Medicine, University of Pennsylvania, with the assistance of Xinyu Zhao. Other imaging was performed at the Fox Chase Imaging Facility under the direction of Michal Jarnik. Constructive comments on the manuscript were given by Ning Chai, William Mason, and Glenn Rall.

REFERENCES

1. **Abrahem, A., and M. Pelchat.** 2008. Formation of an RNA polymerase II preinitiation complex on an RNA promoter derived from the hepatitis delta virus RNA genome. Nucleic Acids Res. **36:**5201–5211.
2. **Alves, C., N. Freitas, and C. Cunha.** 2008. Characterization of the nuclear localization signal of the hepatitis delta virus antigen. Virology **370:**12–21.
3. **Bichko, V. V., Y. E. Khudyakov, and J. M. Taylor.** 1996. A novel form of hepatitis delta antigen. J. Virol. **70:**3248–3251.
4. **Bichko, V. V., and J. M. Taylor.** 1996. Redistribution of the delta antigens in cells replicating the genome of hepatitis delta virus. J. Virol. **70:**8064–8070.
5. **Cardinale, S., B. Cisterna, P. Bonetti, C. Aringhieri, M. Biggiogera, and S. M. Barabino.** 2007. Subnuclear localization and dynamics of the Pre-mRNA 3′ end processing factor mammalian cleavage factor I 68-kDa subunit. Mol. Biol. Cell **18:**1282–1292.
6. **Casey, J. L. (ed.)** 2006. Hepatitis delta virus. Current topics in microbiology and immunology, vol. 307. Springer, Berlin, Germany.
7. **Chang, J., S. O. Gudima, C. Tarn, X. Nie, and J. M. Taylor.** 2005. Development of a novel system to study hepatitis delta virus genome replication. J. Virol. **79:**8182–8188.

3. Intracellular Localization of the Hepatitis Delta Virus Proteins in the Presence and Absence of Viral RNA Accumulation

8. **Chang, J., X. Nie, H. E. Chang, Z. Han, and J. Taylor.** 2008. Transcription of hepatitis delta virus RNA by RNA polymerase II. J. Virol. **82:**1118–1127.
9. **Chang, J., X. Nie, S. Gudima, and J. Taylor.** 2006. Action of inhibitors on accumulation of processed hepatitis delta virus RNAs. J. Virol. **80:**3205–3214.
10. **Contreras, X., M. Barboric, T. Lenasi, and B. M. Peterlin.** 2007. HMBA releases P-TEFb from HEXIM1 and 7SK snRNA via PI3K/Akt and activates HIV transcription. PLoS Pathog. **3:**1459–1469.
11. **Dingle, K., G. Moraleda, V. Bichko, and J. Taylor.** 1998. Electrophoretic analysis of the ribonucleoproteins of hepatitis delta virus. J. Virol. Methods **75:**199–204.
12. **Goic, B., J. Bustamante, A. Miquel, M. Alvarez, M. I. Vera, P. D. Valenzuela, and L. O. Burzio.** 2008. The nucleoprotein and the viral RNA of infectious salmon anemia virus (ISAV) are localized in the nucleolus of infected cells. Virology **379:**55–63.
13. **Gudima, S., J. Chang, G. Moraleda, A. Azvolinsky, and J. Taylor.** 2002. Parameters of human hepatitis delta virus genome replication: the quantity, quality, and intracellular distribution of viral proteins and RNA. J. Virol. **76:**3709–3719.
14. **Handa, H., and Y. Yamaguchi (ed.).** 2006. Hepatitis delta virus. Springer, New York, NY.
15. **Hiscox, J. A.** 2002. The nucleolus—a gateway to viral infection? Arch. Virol. **147:**1077–1089.
16. **Huang, W. H., Y. S. Chen, and P. J. Chen.** 2008. Nucleolar targeting of hepatitis delta antigen abolishes its ability to initiate viral antigenomic RNA replication. J. Virol. **82:**692–699.
17. **Huang, W. H., B. Y. Yung, W. J. Syu, and Y. H. Lee.** 2001. The nucleolar phosphoprotein B23 interacts with hepatitis delta antigens and modulates the hepatitis delta virus RNA replication. J. Biol. Chem. **276:**25166–25175.
18. **Klibanov, S. A., H. M. O'Hagan, and M. Ljungman.** 2001. Accumulation of soluble and nucleolar-associated p53 proteins following cellular stress. J. Cell Sci. **114:**1867–1873.
19. **Kuo, M. Y.-P., J. Goldberg, L. Coates, W. Mason, J. Gerin, and J. Taylor.** 1988. Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. J. Virol. **62:**1855–1861.
20. **Lai, M. M. C.** 2006. Hepatitis delta antigen: biochemical properties and functional roles in HDV replication, p. 38–51. *In* H. Handa and Y. Yamaguchi (ed.), Hepatitis delta virus. Landes Bioscience, Georgetown, TX.
21. **Lai, M. M. C.** 1995. The molecular biology of hepatitis delta virus. Annu. Rev. Biochem. **64:**259–286.
22. **Lazinski, D. W., and J. M. Taylor.** 1994. Expression of hepatitis delta virus RNA deletions: *cis* and *trans* requirements for self-cleavage, ligation, and RNA packaging. J. Virol. **68:**2879–2888.
23. **Lazinski, D. W., and J. M. Taylor.** 1993. Relating structure to function in the hepatitis delta virus antigen. J. Virol. **67:**2672–2680.
24. **Lee, C.-H., S. C. Chang, C.-J. Chen, and M.-F. Chang.** 1998. The nucleolin binding activity of hepatitis delta antigen is associated with nucleolus targeting. J. Biol. Chem. **273:**7650–7656.
25. **Li, Y. J., T. Macnaughton, L. Gao, and M. M. Lai.** 2006. RNA-templated replication of hepatitis delta virus: genomic and antigenomic RNAs associate with different nuclear bodies. J. Virol. **80:**6478–6486.
26. **Litchfield, D. W.** 2003. Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. Biochem. J. **369:**1–15.
27. **Lo, S. J., C. C. Lee, and H. J. Lai.** 2006. The nucleolus: reviewing oldies to have new understandings. Cell Res. **16:**530–538.
28. **Moraleda, G., K. Dingle, P. Biswas, J. Chang, H. Zuccola, J. Hogle, and J. Taylor.** 2000. Interactions between hepatitis delta virus proteins. J. Virol. **74:**5509–5515.
29. **Moraleda, G., and J. Taylor.** 2001. Host RNA polymerase requirements for transcription of the human hepatitis delta virus genome. J. Virol. **75:**10161–10169.
30. **Nakabayashi, H., K. Taketa, K. Miyano, T. Yamane, and J. Sato.** 1982. Growth of human hepatoma cell lines with differentiated functions in chemically defined medium. Cancer Res. **42:**3858–3863.
31. **Perry, R. P.** 1963. Selective effects of actinomycin D on the intracellular distribution of RNA synthesis in tissue culture cells. Exp. Cell Res. **29:**400–406.
32. **Peterlin, B. M., and D. H. Price.** 2006. Controlling the elongation phase of transcription with P-TEFb. Mol. Cell **23:**297–305.
33. **Prasanth, K. V., S. G. Prasanth, Z. Xuan, S. Hearn, S. M. Freier, C. F. Bennett, M. Q. Zhang, and D. L. Spector.** 2005. Regulating gene expression through RNA nuclear retention. Cell **123:**249–263.
34. **Schneider, J., B. Dauber, K. Melen, I. Julkunen, and T. Wolff.** 2009. Analysis of influenza B virus NS1 protein trafficking reveals a novel interaction with nuclear speckle domains. J. Virol. **83:**701–711.
35. **Shih, K.-N., and S. J. Lo.** 2001. The large-delta antigen fused with GFP remains functional and provides for studying its dynamic distribution. Virology **285:**138–152.
36. **Shih, K.-N., Y. T. Chuang, H. Liu, and S. J. Lo.** 2004. Hepatitis D virus RNA editing is inhibited by a GFP fusion protein containing a C-terminally deleted delta antigen. J. Gen. Virol. **85:**947–957.
37. **Taylor, J. M., P. Farci, and R. H. Purcell.** 2007. Hepatitis D (delta) virus, p. 3031–3046. *In* D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. Martin, B. Roizman, and S. E. Straus (ed.), Fields virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
38. **Tseng, C. H., K. S. Jeng, and M. M. Lai.** 2008. Transcription of subgenomic mRNA of hepatitis delta virus requires a modified hepatitis delta antigen, distinct from antigenomic RNA synthesis. J. Virol. **82:**9409–9416.
39. **Wang, C. C., T. C. Chang, C. W. Lin, H. L. Tsui, P. B. Chu, B. S. Chen, Z. S. Huang, and H. N. Wu.** 2003. Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen. Nucleic Acids Res. **31:**6481–6492.
40. **Yamaguchi, K., and H. Handa (ed.).** 2006. Hepatitis delta antigen and RNA polymerase II. Landes Bioscience, Georgetown, TX.
41. **Yamaguchi, Y., T. Mura, S. Chanarat, S. Okamoto, and H. Handa.** 2007. Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. Genes Cells **12:**863–875.
42. **Yeh, T. S., S. J. Lo, P.-J. Chen, and Y. H. Lee.** 1996. Casein kinase II and protein kinase C modulate hepatitis delta virus RNA replication but not empty viral particle assembly. J. Virol. **70:**6190–6198.
43. **Yik, J. H., R. Chen, R. Nishimura, J. L. Jennings, A. J. Link, and Q. Zhou.** 2003. Inhibition of P-TEFb (CDK9/Cyclin T) kinase and RNA polymerase II transcription by the coordinated actions of HEXIM1 and 7SK snRNA. Mol. Cell **12:**971–982.
44. **Zandomeni, R., M. C. Zandomeni, D. Shugar, and R. Weinmann.** 1986. Casein kinase type II is involved in the inhibition by 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole of specific RNA polymerase II transcription. J. Biol. Chem. **261:**3414–3419.
45. **Zuccola, H. J., J. E. Rozzelle, S. M. Lemon, B. W. Erickson, and J. M. Hogle.** 1998. Structural basis of the oligomerization of hepatitis delta antigen. Structure **6:**821–830.

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

Virology 407 (2010) 333–340

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

Intrinsic disorder and oligomerization of the hepatitis delta virus antigen

Carolina Alves [1], Hong Cheng, Heinrich Roder, John Taylor *

Fox Chase Cancer Center, Philadelphia, PA 19111, USA

ARTICLE INFO

Article history:
Received 19 July 2010
Returned to author for revision
19 August 2010
Accepted 19 August 2010
Available online 19 September 2010

Keywords:
Hepatitis delta virus
Delta antigen
Intrinsic disorder
Protein oligomerization
Nucleic acid binding

ABSTRACT

The 195 amino acid basic protein (δAg) of hepatitis delta virus (HDV) is essential for replication of the HDV RNA genome. Numerous properties have been mapped to full-length δAg and attempts made to link these to secondary, tertiary and quaternary structures. Here, for the full-size δAg, extensive intrinsic disorder was predicted using PONDR-FIT, a meta-predictor of intrinsic disorder, and evidenced by circular dichroism measurements. Most δAg amino acids are in disordered configurations with no more than 30% adopting an α-helical structure. In addition, dynamic light scattering studies indicated that purified δAg assembled into structures of as large as dodecamers. Cross-linking followed by denaturing polyacrylamide gel electrophoresis revealed hexamers to octamers for this purified δAg and at least this size for δAg found in virus-like particles. Oligomers of purified δAg were resistant to elevated NaCl and urea concentrations, and bound without specificity to RNA and single- and double-stranded DNAs.

Introduction

The replication of the RNA genome of hepatitis delta virus (HDV) requires the translation of a single 195 amino acid basic protein known as the delta antigen (δAg) (Chao et al., 1990). As reviewed elsewhere (Casey, 2006; Lai, 2006; Taylor, 2009), different properties have been attributed to the δAg, some of which have been associated with domains on the δAg primary sequence. These include a nuclear localization signal (NLS) and an RNA binding domain (RBD) (Fig. 1). Also, during HDV replication an RNA editing event occurs, which allows the translation of a δAg that is 19 aa longer. This larger δAg acts as a dominant negative inhibitor of the smaller form (Chao et al., 1990), and has an essential role in recruitment of HBV envelope proteins for assembly of progeny virus (Chang et al., 1991).

Early studies with altered forms of δAg revealed that regions near the N-terminus were required to make dimers and showed that this ability was essential for functionality in viral RNA replication (Lazinski and Taylor, 1993). A peptide corresponding to amino acids 12–60, and spanning the region thought necessary for dimerization readily formed crystals; the structure was determined as a dimer of anti-parallel α-helices (Zuccola et al., 1998) that became known as the coiled-coil domain (CCD). From this structural analysis it was also argued that the δAg might form octamers. In addition, mass spectrometry after cross-linking, confirmed that some of the purified δAg did form octamers (Zuccola et al., 1998). Amino acids in the CCD structure that might be critical for the formation of dimers and higher multimers were chosen and tested (Moraleda et al., 2000). Some agreement with the structural models was obtained by mutating these residues (Cornillez-Ty and Lazinski, 2003).

Studies of δAg in virus particles have also been carried out. Associations with HDV genomic RNA were demonstrated by both equilibrium centrifugation (Ryu et al., 1993), rate zonal sedimentation, and agarose gel electrophoresis under non-denaturing conditions (Dingle et al., 1998b). Interestingly, δAg behaves as a high MW complex, even when the associated RNA is released by prior treatment with either RNase or vanadyl ribonucleosides, consistent with the idea that δAg has an intrinsic ability to form and maintain multimers (Dingle et al., 1998b).

Several studies indicate that δAg is also present in high MW complexes within cells. Wang and Lemon (1993) applied prior cross-linking to HDV infected liver tissue and showed that δAg was in complexes with a sedimentation value of 15 S. Chang et al. (2008) obtained similar results using extracts of tissue culture cells transfected with HDV. Interestingly, similar-sized complexes were found when δAg was expressed in the absence of the full-length viral RNA, and no size change was detected when these complexes were first treated with RNase. An extensive mass spectrometry study identified >100 protein partners of δAg in cells undergoing HDV replication (Cao et al., 2009). Thus, while the majority of δAg expressed in cells seems to be in high MW complexes it remains unclear to what extent these represent multimers of δAg and/or association with host proteins. In terms of the intrinsic secondary, tertiary and quaternary structures of the δAg, most studies have been limited to fragments of δAg or complexes that include

* Corresponding author. Fox Chase Cancer Center, 333 Cottman Avenue, Philadelphia, PA 19111-2497, USA. Fax: +1 215 728 3105.
E-mail addresses: carolinamantero@gmail.com (C. Alves), hong.cheng@fccc.edu (H. Cheng), Heinrich.Roder@fccc.edu (H. Roder), john.taylor@fccc.edu (J. Taylor).
[1] Current address: Molecular Biology Unit, Center for Malaria and Tropical Diseases, Institute of Hygiene and Tropical Medicine, New University of Lisbon, Lisbon, Portugal.

0042-6822/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2010.08.019

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

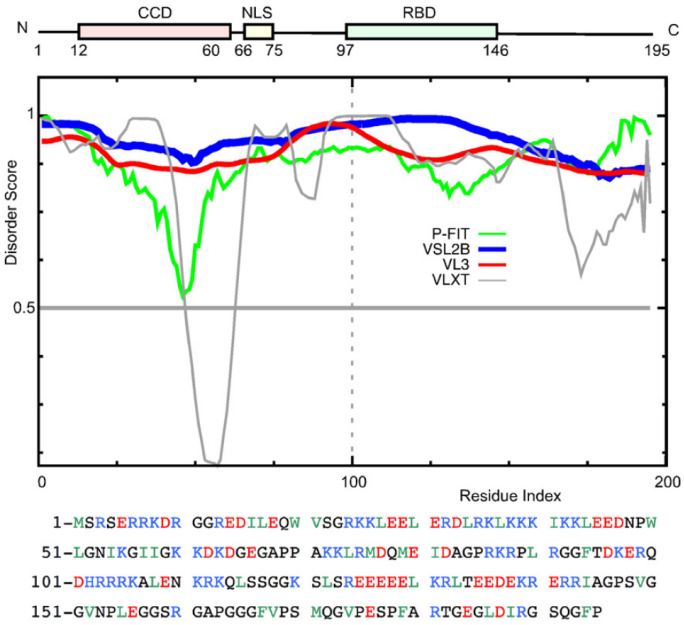


Fig. 1. Primary and secondary structure features of the 195 amino acid δAg. The upper panel indicates the coiled-coil domain (CDD), nuclear localization signal (NLS) and the RNA binding domain (RBD) (Han et al., 2009). The middle panel shows predictions of disorder using the meta-predictor PONDR-Fit (Xue et al., 2010), and three of its component programs (Obradovic et al., 2003; Obradovic et al., 2005), as indicated. The disorder score is a measure of the certainty that a region of the protein is disordered; a score of 1 indicates 100% certainty. The lower panel shows the primary sequence with basic, acidic and hydrophobic amino acids indicated in blue, red and green, respectively.

δAg but, in other respects, were of uncertain origin. Therefore, the present studies focus on the use of δAg that was of full-length, with no modifications, and purified in such a way as to maintain native structure. Furthermore, several experimental approaches have been used to quantitate intra- and inter-molecular interactions.

Theoretical approaches have been used to predict secondary structure (Deny, 2006; Taylor, 2009) but the tertiary structure for full-length δAg remains unknown; at least three labs have purified the full-length δAg and yet been unable to determine a crystal structure. A reason for this latter difficulty might be what is referred to as intrinsic protein disorder. Based on experimental evidence and the development and application of neural network predictors it is now believed that ~50% of natural eukaryotic proteins contain at least one region of disorder (Dunker et al., 2000). It is thought that disorder can offer an evolutionary benefit, making possible protein–protein and protein–nucleic acid interactions in which disordered regions fold into more structured domains (Dyson and Wright, 2004; Fuxreiter et al., 2008).

In the following study we have made use of structure predictors along with experimental strategies to detect secondary, tertiary and quaternary structures of δAg, and its interactions with nucleic acids.

Results

Predictions of intrinsic disorder

In order to predict disordered regions in proteins, neural network strategies have been applied in programs that are then educated using experimental data on regions of proteins known to be either ordered or disordered. Recently Xue et al. (2010) have combined six of these programs to produce PONDR-Fit, a meta-predictor which provides significantly improved accuracy. We applied this program to the full-length δAg, as seen in Fig. 1, where a score of 1.0 indicates a region that is disordered and 0 indicates a highly ordered region. According to this meta-predictor most of the δAg is disordered. For the CCD, which spans positions 12–60, only a small patch is predicted to have less disorder. This is in rough agreement with studies of the corresponding peptide that crystallizes and has been solved as an anti-parallel coiled-coil (Zuccola et al., 1998). Also shown are the results for three of the component predictors.

The same meta-predictor was applied to the δAg for all eight clades of HDV with results as summarized in Fig. 2. Maybe the only conserved feature was the prediction of partial order within the CCD. There were no regions with comparable levels of predicted order, and even regions with reduced order, were not conserved between clades. In summary, other than for part of the CCD region, the shared feature was predicted disordered. As a reference a comparison was made with the 125 amino acid apoptin protein of the chicken anemia virus. As reviewed in the Discussion, this protein has many similarities to δAg, including extensive disorder and yet, according to the meta-predictor, is not as disordered as the δAg. For comparison we also show the highly α-helical ammonia channel protein of *E. coli*.

Circular dichroism of purified δAg

We next evaluated the secondary structure of the full-length δAg, experimentally, via circular dichroism, which assesses the average secondary structure content of a protein in solution. As shown in Fig. 3, the spectrum of recombinant δAg exhibits a strong negative band around 208 nm and a somewhat weaker negative band around 222 nm. Assuming a molar mean-residue ellipticity of −34,100 mdeg cm^2 $dmol^{-1}$ at 222 nm for a fully α-helical peptide (Scholtz and Baldwin, 1992), the data indicate that δAg contains approximately 35% α-helical secondary structure. Alternatively the data were analyzed using the K2D2 program (Perez-Iratxeta and Andrade-Navarro, 2008) giving a best fit with 29% α-helix, 14% β-sheet, and 57% other (mainly disordered).

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

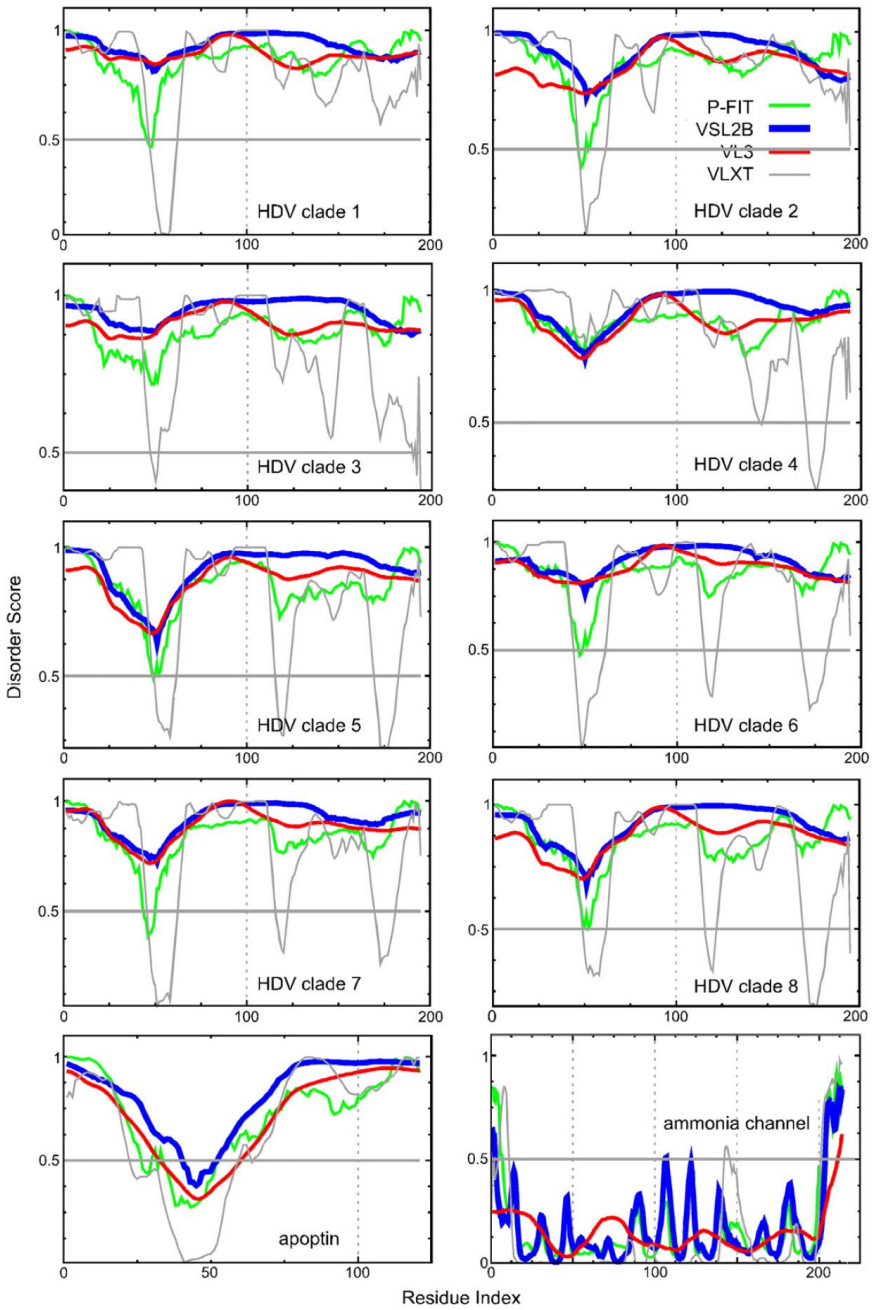


Fig. 2. Predictions of δAg disorder for all eight clades of HDV. Analyses were made as in Fig. 1. Also shown are controls of the highly disordered apoptin of chicken anemia virus, and the highly structured ammonia channel protein from *E. coli*.

Previous studies have used CD to detect α-helical structure in the 24–50 peptide (Cheng et al., 1998; Lin et al., 1999; Lou et al., 2000) and in the larger 12–60 fragment of Rozzelle et al. (1995). X-ray crystallography confirmed that the 12–60 peptide is largely α-helical, adopting a dimeric anti-parallel coiled-coil structure (Zuccola et al., 1998). If we accept that the 12–60 peptide is 90% α-helical structure (Zuccola et al., 1998) it could contribute 20% α-helical structure to the full-length protein. This is less than predicted for δAg by CD, implying

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

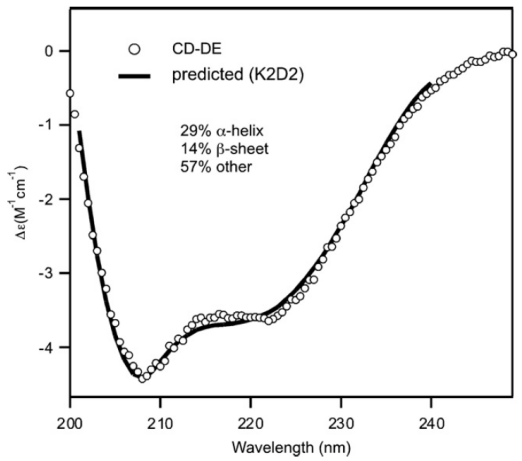


Fig. 3. Circular dichroism of δAg. The purified protein was analyzed as described in Materials and methods. The open circles show the data points and the solid line the fit using the K2D2 program (Perez-Iratxeta and Andrade-Navarro, 2008) and the indicated results for α-helix, β-sheet and other (disordered) structures.

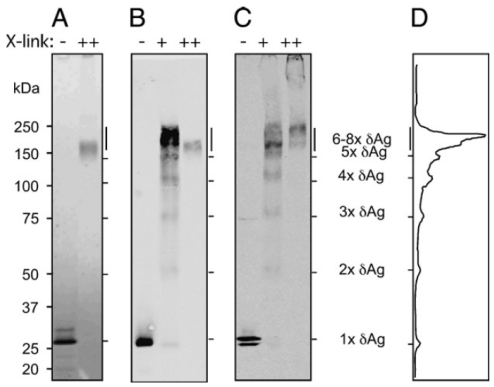


Fig. 4. Electrophoretic mobility of purified δAg and of δAg assembled into virus-like particles. In panels A and B purified δAg, at 2 μM was treated without or with glutaraldehyde cross-linking, 0.01% or 0.1%, as indicated by + and ++ respectively. This was followed by SDS denaturation and electrophoresis on a gel of 4–12% polyacrylamide. In panel A, total protein was detected by SimplyBlue staining and in panel B, δAg was detected by immunoblot using specific antibody. In panel C, virus-like particles containing δAg, without and with prior cross-linking were examined as in panel B. The minor band migrating faster than the monomer might be a proteolytic fragment. At left are indicated MW markers. Panel D is the quantitation of panel B, lane 2. Indicated is an interpretation of detected bands as multimers of δAg.

that a small amount of additional α-helical structure exists within the full-length protein. This conclusion is consistent with previous theoretical predictions of α-helical structure within δAg (Deny, 2006).

Dynamic light scattering to detect δAg multimers

As mentioned in the Introduction a number of approaches have been used to evaluate the quaternary structure of δAg, using the full-length protein as well as truncated and/or tagged forms of δAg. For the recombinant full-length protein used in this study, there is already evidence from mass spectrometry for the occurrence of octamers, detected following protein cross-linking (Zuccola et al., 1998). However, it was unclear if these were stable structures or transient structures that could be stabilized by the cross-linking protocol.

Dynamic light scattering (DLS) provides a measure of translational diffusion constants, which together with assumptions on protein shape can yield insight into the oligomerization state of protein solutions of non-cross-linked proteins that are sufficiently monodisperse (Schurr and Schmitz, 1986). The DLS data for δAg (Table S1 and Fig. S1) showed a major species with a hydrodynamic radius $R_h = 6.44$ nm (accounting for 98% of the amplitude and essentially 100% of the total protein mass). Assuming the complex can be approximated as a sphere, the corresponding molecular mass is ~263 kDa, consistent with a multimer of twelve 22 kDa subunits. This is an upper limit of the subunit number, since a non-spherical structure could exhibit a similar R_h with fewer subunits. Very similar DLS parameters were obtained after addition of 150 mM NaCl (MW = 279 kDa) or lowering the temperature to 15 °C (275 kDa; Table S1), suggesting that the size of the δAg oligomer was quite insensitive to solvent conditions.

Gel electrophoresis to detect δAg multimers

To obtain a more precise estimate of oligomer subunit number, we used SDS polyacrylamide gel electrophoresis of purified δAg, with and without prior glutaraldehyde cross-linking. Un-cross-linked protein was almost homogeneous. There was a single major band with a minor band (<10%) of slightly slower mobility (Fig. 4A). Relative to MW markers, the major band is ~26 kDa, significantly larger than the predicted size of 21.8 kDa and as detected by mass spectrometry (Zuccola et al., 1998). However, SDS gel electrophoresis is known to overestimate the molecular size of the δAg (Bergmann et al., 1987; Gerlich et al., 1987). Following cross-linking of δAg at 2 μM, it migrated, relative to protein MW standards, as a broad band at about 160 kDa, consistent with aggregates of 6- to 8-mers. However, in addition to the anomalous mobility of the monomer, it might be that cross-linked multimers also migrate aberrantly relative to protein standards.

Identical samples were also analyzed by immunoblot to detect δAg. Without prior cross-linking, the δAg again gave a major band at ~26 kDa (Fig. 4B). This confirmed both the purity of the δAg preparation and the lack of hydrolytic protein fragments.

We next asked if reducing the extent of prior cross-linking altered estimates of oligomer size. As shown in Fig. 4B, with quantitation presented in Fig. 4D, we detected small amounts of species consistent with 1, 2, 3, 4 and 5 δAg. (Note that the immunoblot assay could give signals that increase in linear proportion to the size of a multimer.) We also note that with reduced cross-linking the major species near the top of the gel, indicated as 6- to 8-mers, gave a stronger immunoblot signal. This is consistent with less cross-linking producing multimers that even after SDS gel electrophoresis, are less compact and more readily reactive with the polyclonal antibody.

We next asked if we could use cross-linking to detect δAg multimers in virus-like particles. As shown in Fig. 4C, we obtained results similar to those obtained using purified δAg (c.f. Fig. 3B). By reducing cross-linking, we again detected species corresponding to 2, 3, 4, 5, and 6–8 molecules of δAg.

As a test of the natural stability of the complexes present in the purified δAg, we incubated purified δAg in various concentrations of NaCl and urea prior to cross-linking. As shown in Fig. S2, the complexes were largely resistant to these treatments.

Binding of nucleic acids to δAg multimers

The above studies support the conclusion that complexes of six, eight or more δAg molecules are formed in solutions of purified protein and in virus-like particles associations of six to eight or more δAg molecules are formed. Based upon this similarity, and the reports that δAg binds RNA in virus-like particles, we asked whether the multimers made from purified protein were able to change the electrophoretic mobility in non-denaturing agarose gels of different species of ^{32}P-labeled RNA and DNA.

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

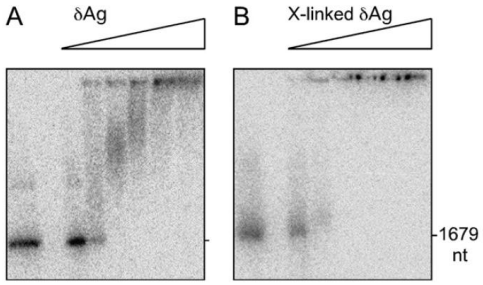


Fig. 5. Electrophoretic mobility shift of HDV genomic unit-length RNA by δAg. Increasing amounts of δAg, without (panel A) or with (panel B) prior cross-linking with 0.1% glutaraldehyde followed by quenching with ammonium acetate, were incubated with trace amounts of ^{32}P-labeled unit-length linear genomic HDV RNA. Samples were analyzed on non-denaturing gels of 1.5% agarose. After electrophoresis the gel was dried and radioactivity assayed using a bio-imager. In each panel the first lane is the control in the absence of δAg. The triangle indicates use of increasing concentrations of δAg: 0.2, 0.4, 0.8, 1.6, 2 and 3.2 μM.

First we tested the ability of increasing concentrations of δAg to bind to relatively small amounts of ^{32}P-labeled unit-length genomic RNA. As shown in Fig. 5A, δAg at 0.2 μM did not affect RNA mobility, but with increasing concentration mobility shifts were observed, and by 1.6 μM the majority of the RNA was retained at the top of the gel.

In order to test whether rearrangement of δAg multimers was required for binding to RNA, we subjected the δAg to prior cross-linking followed by quenching of the excess cross-linker with ammonium acetate. As shown in Fig. 5B, the cross-linked δAg was still able to bind RNA with almost equal efficiency as the non-cross-linked δAg. To confirm the completeness of the quenching procedure, glutaraldehyde was pre-mixed with quencher and then added to δAg; no cross-linking was detected (data not shown). We conclude that preformed δAg multimers can bind directly to the RNA.

Next we tested other nucleic acids for their ability to bind to δAg. Fig. 6A and B shows results with two different HDV RNA species. The first was 1679 nt unit-length linear antigenomic RNA, while the second was a 419 nt antigenomic RNA which is not predicted to contain any of the HDV rod-like folding. As shown, as the concentration of the δAg was increased both of these RNAs underwent a mobility shift, ultimately collecting at the origin of the gel. Thus, rod-like folding is not needed for binding.

Two non-HDV RNAs were also tested in the mobility shift assay. RMRP is a 267 nt non-coding host RNA recently shown to bind to telomerase (Maida et al., 2009), and potato spindle tuber viroid RNA is 359 nt long. As shown in Fig. S2A and B respectively, both RNAs bound to δAg.

DNA species were also tested. For this we 5′-end labeled a DNA ladder with double-stranded species ranging from 100 to 12,000 bp. As shown in Fig. 5C all species were shifted towards the top of the gel. There was no significant difference related to the length of the DNA. Denatured DNAs were also shifted by the same concentrations of δAg (data not shown).

In summary these mobility studies allow two conclusions: (i) Multimers of δAg are able to bind many different forms of RNA and DNA. This is in contrast to previous reports from this and other labs and will be considered in the Discussion. (ii) Multimers can form first and then bind nucleic acids. As previously reported the δAg multimers assembled into virus-like particles and into serum virus retain their integrity when the RNA is removed (Dingle et al., 1998b). While this is consistent with the RNA binding to preformed multimers, it is equally consistent with the RNA assisting in the formation of multimers that no longer need the RNA for their integrity.

Since δAg is a basic protein with an expected net charge of +12 at neutral pH, a multimer will be even more positively charged, and this could facilitate binding to the negatively charged nucleic acid. Therefore we asked if high concentrations of NaCl could disrupt the binding of ^{32}P-labeled genomic HDV RNA to δAg. As shown in Fig. 7A, at 2.4 M NaCl, the RNA was released. In parallel we tested whether NaCl was able to disrupt δAg multimers. As shown in Fig. S2, even at 3.6 M NaCl, this did not occur. We also showed that increasing concentrations of vanadyl ribonucleosides (VRC) released the RNA from the δAg multimers (Fig. 7B).

Discussion

This study advances our understanding of the secondary, tertiary, and quaternary structures of δAg that are essential for HDV replication. Previous studies have identified a dimerization domain

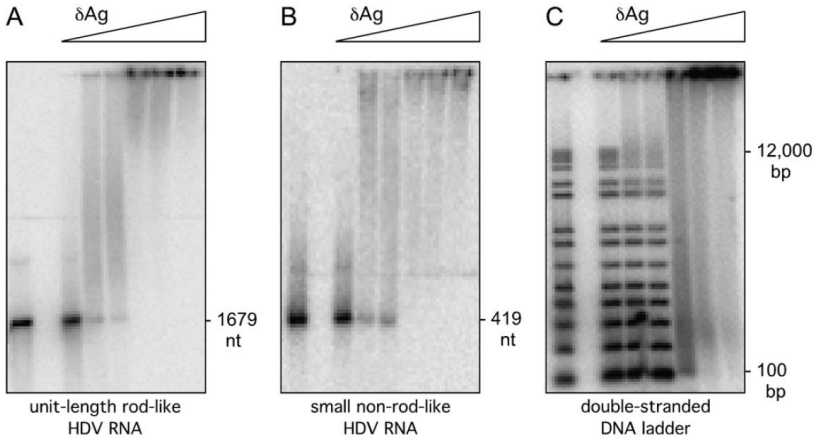


Fig. 6. Electrophoretic mobility shift of RNA and DNA species by δAg. ^{32}P-labeled RNA species were gel purified and then incubated in the absence or presence of increasing amounts of δAg. Samples were then subjected to electrophoresis under non-denaturing conditions, as in Fig. 5. In panel A the RNA was the 1679 nt unit-length antigenomic linear HDV RNA. In panel B it was a 419 nt species of antigenomic linear HDV RNA, representing the region 714-224, using the numbering of Kuo et al. (1988), that is predicted to not include the rod-like folding of the RNA. In panel C, we used end-labeled double-stranded DNA with a ladder of sizes from 100 to 12,000 bp. In each panel the left lane is the control in the absence of δAg. The triangles indicate use of increasing concentrations of δAg: 0.2, 0.4, 0.8, 1.6, 2 and 3.2 μM.

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

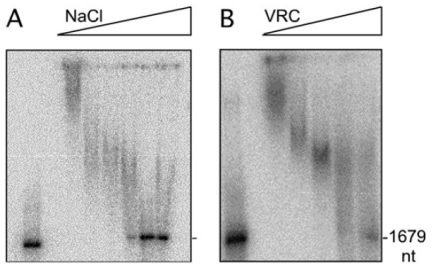


Fig. 7. Binding of HDV RNA to δAg can be reversed by increasing concentrations of NaCl and VRC. ^{32}P-labeled unit-length genomic HDV was incubated with δAg at 2 μM, and then increasing concentrations of NaCl (panel A) or VRC (panel B), and then analyzed as in Fig. 5. In each case the first lane is the control in the absence of δAg. The triangles indicate use of increasing concentrations of NaCl: 0.15, 0.3, 0.6, 1.2, 2.4, and 3.6 M (panel A) or VRC: 0, 0.3, 1, 3, and 10 mM (panel B).

(CCD) spanning amino acids 12–60 (Fig. 1). In particular, the crystal structure of a synthetic 12–60 peptide revealed an anti-parallel coiled-coil dimer in which over 90% of the residues (45 out of 49) are α-helical (Zuccola et al., 1998) consistent with earlier CD data (Rozzelle et al., 1995). From our CD measurements on full-length δAg we deduced that the remainder of the protein contains very little additional secondary structure (Fig. 3). We arrived at a similar conclusion using the recently developed meta-predictor PONDR-Fit (Xue et al., 2010) which combines six neural network programs trained on a large number of regions of known protein disorder. Analysis of the δAg sequence studied here (Fig. 1) as well as for all eight clades of HDV (Fig. 2) predicted high levels of disorder (P-fit scores >0.75) throughout the protein, except for parts of the CCD region. There was no evidence for additional ordered regions outside the CCD that are conserved between clades. PONDR-VLXT, one of the six components of PONDR-Fit, predicts somewhat higher levels of order, both in the CCD and two or three segments in the C-terminal half of δAg (especially for clades 5–8). This algorithm recognizes short regions experimentally known to be disordered that become structured when they are bound to other proteins (Obradovic et al., 2005). This supports the possibility that δAg multimerization promotes helix formation in the CCD while isolated segments within the initially disordered C-terminal region are poised to become ordered upon interaction with other cellular proteins.

Intrinsic disorder is increasingly recognized for its relevance in understanding the intra- and inter-molecular interactions of important proteins (Dunker et al., 2008). In the case of viral proteins that are expressed in cells in relatively large amounts, disordered regions may increase the capacity to structurally adapt for engagement in a wide variety of homo- and hetero-multimeric interactions (Goh et al., 2008a,b); during HDV replication δAg interacts with more than 100 different host proteins (Cao et al., 2009). It is interesting to compare δAg with apoptin, a protein of chicken anemia virus. Like δAg this protein is small (125 amino acids), basic, essential for replication, binds nucleic acids, and has not yielded to crystal-structure determination. Apoptin forms homomultimers of 30–40 molecules and is known to interact with several different host proteins (Leliveld et al., 2003a,b; Los et al., 2009; Teodoro et al., 2004). Comparison of the PONDR-Fit profiles of apoptin with that of δAg (Fig. 2) shows somewhat higher degrees of order centered around residue 45 with flanking regions of predicted disorder, similar to the patterns observed for the N-terminal 120 residues of δAg. (Incidentally, the N-terminal 69 amino acids of apoptin have been demonstrated to account for its multimerization behavior (Leliveld et al., 2003b).) However, a much higher fraction of the residues in δAg are predicted to be highly disordered suggesting that intrinsic disorder might be even more relevant to understanding the properties and functions of δAg in the HDV life cycle.

Qualitative insight into δAg quaternary structure was obtained by denaturing gel electrophoresis. Prior cross-linking revealed the existence of multimers formed both by the purified protein and that present in virus-like particles. With the purified δAg the complexes were detected when the proteins were at 0.2 and 2 μM during cross-linking (Fig. 4 and data not shown). Much higher concentrations are present in cells during HDV replication, suggesting that multimers should also arise *in vivo*; for example, the observed accumulation of 2.4 million copies of δAg in a nucleus of ~9 μm diameter represents an average concentration of 24 μM (Chang et al., 2005).

Relative to protein standards the majority of cross-linked multimers appeared to contain 6–8 molecules of δAg (Fig. 4A–C). This confirms and extends previous mass spectrometry studies in which octamers were detected although not quantified (Zuccola et al., 1998). In addition, the hydrodynamic dimensions observed in our dynamic light scattering measurements, were consistent with an oligomer of ≤12 subunits. The δAg multimers we observed were remarkably stable even in the presence of high NaCl or urea concentrations (Fig. S2).

The multimers of full-length δAg did bind nucleic acids but without specificity; HDV RNAs with and without rod-like folding and of different lengths were bound, as were non-HDV RNAs, and double- and single-stranded DNAs (Figs. 5A, 6 and S3). This binding is consistent with simple electrostatic interactions. The δAg has a predicted net charge of +12 at neutral pH (Kuo et al., 1988) and multimers will have a greater total charge and are thus expected to have enhanced affinity for nucleic acids that have one negative charge per nucleotide. Similar electrostatic interactions with nucleic acids have been reported for the apoptin mentioned earlier, which is also basic and forms large oligomers (Leliveld et al., 2003a). We also found that nucleic acid binding was achieved when the protein was first cross-linked (Fig. 5B), indicating that δAg rearrangement was not necessary during binding. Observations that the binding was reversed with increasing concentrations of NaCl (Fig. 7A) or vanadyl ribonucleosides (VRC) (Fig. 7B) agree with the interpretation of an electrostatic interaction. The findings are also consistent with our earlier studies on natural HDV ribonucleoprotein complexes where VRC released the HDV RNA but left the δAg complexes intact (Dingle et al., 1998b). However, unlike the natural situation, the purified δAg did not demonstrate specificity for rod-like HDV RNA.

The ability of δAg to bind non-specifically to nucleic acids might explain its behavior as a chaperone. In a series of papers Wang et al. (2003) showed that δAg would facilitate strand exchange reactions between short DNA oligonucleotides. We confirmed their results including evidence that the chaperone effect can be achieved with N-terminal and C-terminal fragments of the δAg lacking the oligomerization domain (G. Moraleda, H.J. Netter, and J.M. Taylor, unpublished observations). Thus, the chaperone activity of δAg appears to be independent of its ability to oligomerize. It seems to be a property of a positively charged protein independent of multimerization.

The present findings of non-specific binding of δAg to nucleic acids also appear to disagree with earlier studies from this lab by Chao et al. (1991). which indicated that a delta antigen would bind to HDV RNAs but only if they had the ability to form rod-like folding structures. However, the assay used in the previous study was quite different: the protein tested was a fusion protein and the detection method was a northwestern. The present findings also disagree with recent studies that employed a δAg with an N-terminal fusion and a C-terminal deletion in a mobility shift assay and concluded that binding to HDV RNA required a minimum of ~311 nt of rod-like folding (Defenbaugh et al., 2009; Lin et al., 2010). However, it should be noted that when a non-truncated form of fusion protein was used, no specificity was detected in the RNA binding assay.

Further *in vitro* experiments are thus needed to determine whether or not we can identify features of the full-length protein and/or the RNA

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

substrates that will produce specific interactions. Recent *in vitro* studies with the full-length core protein of HBV have also encountered RNA interactions that are efficient but not specific for viral RNA (Porterfield et al., 2010). However, it remains to be seen how and to what extent the *in vivo* interactions can be specific for viral RNA. Further work is needed to gain a more detailed understanding of the mechanism of *in vivo* self-associations of δAg and its interactions with HDV RNAs as well as host proteins and nucleic acids. We have reported that *in vivo* δAg is present in large complexes, even in the absence of HDV RNA genome replication or when HDV RNAs are present but released by treatment with VRC (Dingle et al., 1998b) or RNase (Chang et al., 2008; Dingle et al., 1998a). In addition, a recent study reported the identification by mass spectrometry of more than 100 host proteins that interact with δAg and began to test how these might contribute to supporting HDV replication (Cao et al., 2009). Sorting out which are the relevant associations is non-trivial because millions of copies of δAg are produced per cell during HDV replication, most of which are in the nucleus at concentrations exceeding 24 μM (Chang et al., 2005), much higher than the 0.2–2 μM used for the *in vitro* studies reported here. A further *in vivo* complication is that in the presence or absence of HDV RNA, δAg is found in either the nucleoplasm or nucleolus, respectively (Han et al., 2009). That is, the same high concentrations of δAg can assume different intranuclear associations depending upon whether or not HDV RNA species have been allowed to replicate.

Conclusion

A combination of biochemical and biophysical approaches has been used to show that extensive intrinsic disorder and multimerization are essential aspects of full-length δAg. However, the *in vivo* interaction between this essential protein and viral RNAs is more specific and no doubt more complex than we have been able to reproduce *in vitro* with purified protein.

Materials and methods

Purified delta protein

As previously described, full-length δAg was expressed in *E. coli* and purified (Dingle et al., 1998a; Moraleda et al., 2000; Zuccola et al., 1998). The chromatography procedures used in the purification were chosen with the aim of obtaining a protein that was still in its native conformation. The protein was a generous gift from Harmon Zuccola and James Hogle and had been evaluated as >85% pure. The sequence corresponds to an American strain of HDV (M28267). From optical density measurements we deduce that the purified protein contains less than one nucleotide of RNA or DNA per molecule of delta antigen. The protein has been tested for its ability to support HDV genome replication: surprisingly transfection into cells of this protein and in vitro transcribed HDV RNA at as low as an equimolar ratio was sufficient to initiate replication (Dingle et al., 1998a). Nevertheless, this is not proof that the majority of the purified protein is native in conformation.

PONDR-FIT analyses

PONDR-Fit (Xue et al., 2010) is a meta-predictor which combines the results of six different programs each separately designed to predict disorder in proteins. It was applied here to the δAg sequence of the purified protein, as well as to the δAg for all eight clades of HDV: clade 1 (US-1), D01075; clade 2 (Taiwan-3), U19598; clade 3 (Vnzd8375), AB037947; clade 4 (Tokyo), AB118847; clade 5 (dFR2600, Togo), AM183326; clade 6 (dFR2627, Nigeria), AM183329; clade 7 (dFR2158, Cameroon), AM183333; clade 8 (dFR2072, Senegal), AM183330, the 125 amino acid apoptin (Noteborn et al., 1991) of chicken anemia virus (M55918), and the 428 amino acid ammonia channel protein of *E. coli* (P69681).

Circular dichroism analysis

The far-UV circular dichroism (CD) spectrum of a 5.3 μM protein solution (in 20 mM potassium phosphate buffer, pH 6.3) was acquired at 25 °C on an Aviv 62A spectropolarimeter (Aviv, Lakewood, NJ), using a 1 mm quartz cuvette. The CD spectrum is an average of five scans recorded in the far-UV region (195–250 nm) with a band pass of 2 nm.

Dynamic light scattering analysis

Dynamic light scattering experiments were performed at 15 and 25 °C on a DynaPro Molecular Sizing Instrument with Dynamics V6 data analysis software (Protein Solutions, Inc.) on 80 μl of a 10 μM solution of δAg (in 20 mM potassium phosphate buffer, pH 6.3, without or with 150 mM NaCl). Autocorrelation curves were acquired for a total acquisition time of 600 s. Representative data are shown in Fig. S1A with deductions summarized in Table S1.

Virus-like particles

As previously described, Huh7 cells were transfected with plasmids to initiate HDV genome replication and to express the envelope proteins of hepatitis B virus (Gudima et al., 2007). Media were harvested at days 7–10, and clarified of cellular debris by low speed centrifugation. Virus-like particles were then collected by two cycles of ultracentrifugation through a cushion of 20% sucrose.

DNA-directed RNA transcription

Plasmids containing HDV and non-HDV sequences were transcribed *in vitro* using T7 RNA polymerase using a RiboMax large-scale transcription system (Promega) with the addition of [α-^{32}P] CTP (MP Biomedicals). Unit-length genomic HDV was transcribed from pSG253 pre-cut with *Not*I (Gudima et al., 2004). Unit-length, 1060 and 419 nt antigenomic RNAs were transcribed from pSG254 pre-cut with *Not*I, *Stu*I, and *Bgl*II, respectively (Gudima et al., 2004). Unit-length genomic PSTVd RNA was transcribed from pJC144 pre-cut with *Eco*RI (Chang et al., 2003). RMRP was transcribed using expression PCR (Chang and Taylor, 2002). Transcripts were gel purified prior to use in gel retardation assays.

Gel electrophoresis analyses

Gel retardation assays with ^{32}P-labeled nucleic acids, were performed using gels of 1.5% agarose (in 1× TBE) followed by drying onto charged paper (DE81, Whatman) and bio-imager detection (Fuji). ^{32}P-labeled RNAs were prepared as described above. ^{32}P-labeled DNA was obtained by phosphorylating a 1 kb Plus DNA Ladder (Invitrogen) using [γ-^{32}P] ATP (MP Biomedicals) following the manufacturer's instructions.

Association of δAg with nucleic acids was performed in 150 mM NaCl–5 mM HEPES (pH 7.5) for 10 min at room temperature. RNA concentrations were at 2 nM and δAg concentrations were as indicated in the figure legends.

Protein samples, without and with prior glutaraldehyde cross-linking, were examined on 4–12% acrylamide NuPAGE gels (Invitrogen). Glutaraldehyde was typically added to 0.1% final concentration, and after 10 min at room temperature it was inactivated by the addition of 100 mM ammonium acetate. Total protein in gels was stained with SimplyBlue SafeStain (Invitrogen) and detected using an Odyssey laser scanner (LI-COR). For immunodetection δAg proteins were electrotransferred to a cellulose nitrate membrane and reacted with a rabbit polyclonal anti-δAg followed by infrared dye-labeled goat anti-rabbit antibody (LI-COR), with detection using the Odyssey scanner.

Supplementary materials related to this article can be found online at doi:10.1016/j.virol.2010.08.019.

4. Intrinsic Disorder and Oligomerization of the Hepatitis Delta Virus Antigen

Acknowledgments

Critical comments on the manuscript were provided by William Mason, Celso Cunha, and Eileen Jaffe. We thank Harmon Zuccola and James Hogle for the purified recombinant δAg, and Ziying Han for a source of HDV virus-like particles. Vladimir Uversky and Roland Dunbrack introduced us to PONDR programs and analysis. Gloria Moraleda and Hans Netter are cited for unpublished nucleic acid chaperone reactions. We acknowledge assistance from the FCCC Spectroscopy Support Facility. JT was supported by grants AI-256522 and CA-06927 and by an appropriation from the Commonwealth of Pennsylvania. HR was also supported by grant R01 GM056250. CA was supported by a Ph.D. grant from Fundação para a Ciência e Tecnologia.

References

Bergmann, K.F., Pohl, C., Gerin, J.L., 1987. Characterization of proteins of hepatitis delta virus. Prog. Clin. Biol. Res. 234, 105–110.
Cao, D., Haussecker, D., Huang, Y., Kay, M.A., 2009. Combined proteomic-RNAi screen for host factors involved in human hepatitis delta virus replication. RNA 15, 1971–1979.
Casey, J.L., 2006. Hepatitis delta virus. In: Compans, R.M., Cooper, M.D., Honjo, T., Koprowski, H., Melchers, F., Oldstone, M.B.A., Olsnes, S., Potter, M., Vogt, P.K., Wagner, H. (Eds.), Current Topics in Microbiology and Immunology. Springer, Berlin.
Chang, J., Taylor, J., 2002. In vivo RNA-directed transcription, with template switching, by a mammalian RNA polymerase. EMBO J. 21, 157–164.
Chang, F.L., Chen, P.J., Tu, S.J., Chiu, M.N., Wang, C.J., Chen, D.S., 1991. The large form of hepatitis δ antigen is crucial for the assembly of hepatitis δ virus. Proc. Natl. Acad. Sci. U. S. A. 88, 8490–8494.
Chang, J., Provost, P., Taylor, J.M., 2003. Resistance of human hepatitis delta virus RNAs to dicer activity. J. Virol. 77, 11910–11917.
Chang, J., Gudima, S.O., Tarn, C., Nie, X., Taylor, J.M., 2005. Development of a novel system to study hepatitis delta virus genome replication. J. Virol. 79, 8182–8188.
Chang, J., Nie, X., Chang, H.E., Han, Z., Taylor, J., 2008. Transcription of hepatitis delta virus RNA by RNA polymerase II. J. Virol. 82, 1118–1127.
Chao, M., Hsieh, S.-Y., Taylor, J., 1990. Role of two forms of the hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. J. Virol. 64, 5066–5069.
Chao, M., Hsieh, S.-Y., Taylor, J., 1991. The antigen of hepatitis delta virus: examination of in vitro RNA-binding specificity. J. Virol. 65, 4057–4062.
Cheng, J.W., Lin, I.J., Lou, Y.C., Pai, M.T., Wu, H.N., 1998. Local helix content and RNA-binding activity of the N-terminal leucine-repeat region of hepatitis delta antigen. J. Biomol. NMR 12, 183–188.
Cornillez-Ty, C.T., Lazinski, D.W., 2003. Determination of the multimerization state of the hepatitis delta virus antigens in vivo. J. Virol. 77, 10314–10326.
Defenbaugh, D.A., Johnson, M., Chen, R., Zheng, Y.Y., Casey, J.L., 2009. Hepatitis delta antigen requires a minimum length of the hepatitis delta virus unbranched rod RNA structure for binding. J. Virol. 83, 4548–4556.
Deny, P., 2006. Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades. In: Casey, J.L. (Ed.), Hepatitis Delta Virus. Springer, Heidelberg, Germany, pp. 151–171.
Dingle, K., Bichko, V., Zuccola, H., Hogle, J., Taylor, J., 1998a. Initiation of hepatitis delta virus genome replication. J. Virol. 72, 4783–4788.
Dingle, K., Moraleda, G., Bichko, V., Taylor, J., 1998b. Electrophoretic analysis of the ribonucleoproteins of hepatitis delta virus. J. Virol. Methods 75, 199–204.
Dunker, A.K., Obradovic, Z., Romero, P., Garner, E.C., Brown, C.J., 2000. Intrinsic protein disorder in complete genomes. Genome Inform. Ser. Workshop Genome Inform., vol. 11, pp. 161–171.
Dunker, A.K., Oldfield, C.J., Meng, J., Romero, P., Yang, J.Y., Chen, J.W., Vacic, V., Obradovic, Z., Uversky, V.N., 2008. The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9 (Suppl 2), S1.
Dyson, H.J., Wright, P.E., 2004. Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104, 3607–3622.
Fuxreiter, M., Tompa, P., Simon, I., Uversky, V.N., Hansen, J.C., Asturias, F.J., 2008. Malleable machines take shape in eukaryotic transcriptional regulation. Nat. Chem. Biol. 4, 728–737.
Gerlich, W.H., Heermann, K.H., Ponzetto, A., Crivelli, O., Bonino, F., 1987. Proteins of hepatitis delta virus. Prog. Clin. Biol. Res. 234, 97–103.
Goh, G.K., Dunker, A.K., Uversky, V.N., 2008a. A comparative analysis of viral matrix proteins using disorder predictors. Virol. J. 5, 126.
Goh, G.K., Dunker, A.K., Uversky, V.N., 2008b. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. BMC Genomics 9 (Suppl 2), S4.
Gudima, S.O., Chang, J., Taylor, J.M., 2004. Features affecting the ability of hepatitis delta virus RNAs to initiate RNA-directed RNA synthesis. J. Virol. 78, 5737–5744.
Gudima, S., He, Y., Meier, A., Chang, J., Chen, R., Jarnik, M., Nicolas, E., Bruss, V., Taylor, J., 2007. Assembly of hepatitis delta virus: particle characterization, including the ability to infect primary human hepatocytes. J. Virol. 81, 3608–3617.
Han, Z., Alves, C., Gudima, S., Taylor, J., 2009. Intracellular localization of hepatitis delta virus proteins in the presence and absence of viral RNA accumulation. J. Virol. 83, 6457–6463.
Kuo, M.Y.-P., Goldberg, J., Coates, L., Mason, W., Gerin, J., Taylor, J., 1988. Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. J. Virol. 62, 1855–1861.
Lai, M.M.C., 2006. Hepatitis delta antigen: biochemical properties and functional roles in HDV replication. In: Handa, H., Yamaguchi, Y. (Eds.), Hepatitis Delta Virus. Landes Bioscience, Georgetown, TX, pp. 38–51.
Lazinski, D.W., Taylor, J.M., 1993. Relating structure to function in the hepatitis delta virus antigen. J. Virol. 67, 2672–2680.
Leliveld, S.R., Dame, R.T., Mommaas, M.A., Koerten, H.K., Wyman, C., Danen-van Oorschot, A.A., Rohn, J.L., Noteborn, M.H., Abrahams, J.P., 2003a. Apoptin protein multimers form distinct higher-order nucleoprotein complexes with DNA. Nucl. Acids Res. 31, 4805–4813.
Leliveld, S.R., Zhang, Y.H., Rohn, J.L., Noteborn, M.H., Abrahams, J.P., 2003b. Apoptin induces tumor-specific apoptosis as a globular multimer. J. Biol. Chem. 278, 9042–9051.
Lin, I.J., Lou, Y.C., Pai, M.T., Wu, H.N., Cheng, J.W., 1999. Solution structure and RNA-binding activity of the N-terminal leucine-repeat region of hepatitis delta antigen. Proteins 37, 121–129.
Lin, B.C., Defenbaugh, D.A., Casey, J.L., 2010. Multimerization of hepatitis delta antigen is a critical determinant of RNA binding specificity. J. Virol. 84, 1406–1413.
Los, M., Panigrahi, S., Rahshedi, I., Mandal, S., Stetefeld, J., Essmann, F., Schultze-Osthoff, K., 2009. Apoptin, a tumor-selective killer. Biochim. Biophys. Acta 1793, 1335–1342.
Lou, Y.C., Lin, I.J., Pai, M.T., Cheng, J.W., 2000. Solution structure of an N-capping peptide from the N-terminal leucine-repeat region of hepatitis delta antigen. Arch. Biochem. Biophys. 377, 219–227.
Maida, Y., Yasukawa, M., Furuuchi, M., Lassmann, T., Possemato, R., Okamoto, N., Kasim, V., Hayashizaki, Y., Hahn, W.C., Masutomi, K., 2009. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. Nature 461, 230–235.
Moraleda, G., Dingle, K., Biswas, P., Chang, J., Zuccola, H., Hogle, J., Taylor, J., 2000. Interactions between hepatitis delta virus proteins. J. Virol. 74, 5509–5515.
Noteborn, M.H., de Boer, G.F., van Roozelaar, D.J., Karreman, C., Kranenburg, O., Vos, J.G., Jeurissen, S.H., Hoeben, R.C., Zantema, A., Koch, G., et al., 1991. Characterization of cloned chicken anemia virus DNA that contains all elements for the infectious replication cycle. J. Virol. 65, 3131–3139.
Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., 2003. Predicting intrinsic disorder from amino acid sequence. Proteins 53 (Suppl 6), 566–572.
Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Dunker, A.K., 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins 61 (Suppl. 7), 176–182.
Perez-Iratxeta, C., Andrade-Navarro, M.A., 2008. K2D2: estimation of protein secondary structure from circular dichroism spectra. BMC Struct. Biol. 8, 25.
Porterfield, J.Z., Dhason, M.S., Loeb, D.D., Nassal, M., Stray, S.J., Zlotnick, A., 2010. Full-length HBV core protein packages viral and heterologous RNA with similar high cooperativity. J. Virol. 84, 7174–7184.
Rozzelle, J., Wang, J.-G., Wagner, D., Erickson, B., Lemon, S., 1995. Self-association of a synthetic peptide from the N terminus of the hepatitis delta virus protein into an immunoreactive alpha-helical multimer. Proc. Natl. Acad. Sci. U. S. A. 92, 382–386.
Ryu, W.S., Netter, H.J., Bayer, M., Taylor, J., 1993. Ribonucleoprotein complexes of hepatitis delta virus. J. Virol. 67, 3281–3287.
Scholtz, J.M., Baldwin, R.L., 1992. The mechanism of alpha-helix formation by peptides. Ann. Rev. Biophys. Biomol. Struct. 21, 95–118.
Schurr, J.M., Schmitz, K.S., 1986. Dynamic light scattering of biopolymers. Annu. Rev. Phys. Chem. 37, 271–305.
Taylor, J.M., 2009. Replication of the hepatitis delta virus RNA genome. Adv. Virus Res. 74, 102–121.
Teodoro, J.G., Heilman, D.W., Parker, A.E., Green, M.R., 2004. The viral protein Apoptin associates with the anaphase-promoting complex to induce G2/M arrest and apoptosis in the absence of p53. Genes Dev. 18, 1952–1957.
Wang, J.-G., Lemon, S., 1993. Hepatitis delta virus antigen forms dimers and multimeric complexes in vivo. J. Virol. 67, 446–454.
Wang, C.C., Chang, T.C., Lin, C.W., Tsui, H.L., Chu, P.B., Chen, B.S., Huang, Z.S., Wu, H.N., 2003. Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen. Nucleic Acids Res. 31, 6481–6492.
Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., Uversky, V.N., 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. Biochim. Biophys. Acta 1804, 996–1010.
Zuccola, H.J., Rozzelle, J.E., Lemon, S.M., Erickson, B.W., Hogle, J.M., 1998. Structural basis of the oligomerization of hepatitis delta antigen. Structure 6, 821–830.

SUPPLEMENTAL MATERIAL

Intrinsic disorder and oligomerization of the hepatitis delta virus antigen

Carolina Alves, Hong Cheng, Heinrich Roder, John Taylor

Fox Chase Cancer Center, Philadelphia, PA 19111, United States

One Table and three Figures

Table S1
Summary of DLS parameters for δAg at pH 6.3, based on regularization fit of autocorrelation function

	R (nm)[1]	%Pd[2]	D (cm²/s)	%Int	MW (kDa)[3]
20 mM $NaPO_4$, 25^{0}C	6.44	14	3.384e-7	98	263
20 mM $NaPO_4$, 150 mM NaCl 25^{0}C	6.61	12	3.296e-7	97	279
20 mM $NaPO_4$, 150 mM NaCl 15^{0}C	6.57	10	2.498e-7	94	275

[1] hydrodynamic radius
[2] percent polydispersity
[3] molecular mass of equivalent sphere

Fig. S1. Evaluation of multimers of δAg detected by DLS. Shown are representative DLS data obtained at 25 ^{o}C on a solution of 10 μM δAg in 20 mM potassium phosphate, pH 6.3 (after 30 min centrifugation at 14,000 rpm on an Eppendorf centrifuge). Panel A shows the autocorrelation data (solid blue line) along with a regularization fit obtained using the Dynamics program (Protein Solutions, Inc.). The corresponding residuals are plotted in panel B. Panel C shows the predicted distribution of hydrodynamic radii on a percent-mass scale. Additional parameters are given in Table S1.

Fig. S2. Multimers of δAg are resistant to NaCl and urea. Aliquots of δAg were incubated in the presence of increasing concentrations of NaCl or with urea. They were then cross-linked with glutaraldehyde and analyzed by SDS polyacrylamide gel and immunoblot. At left is a sample that was not cross-linked. The increasing concentrations of NaCl indicated by the triangle were 0.15, 0.3, 0.6, 1.2, 2.4 and 3.6 M. The increasing concentrations of urea were: 0, 1, 2, 4, and 6 M. There was minor dissociation to smaller oligomers after treatment with 6 M urea. And with increasing concentrations of urea the 6-8 mer band migrated less; this could be due to displacement of SDS from the protein and/or a reduction in the extent of cross-linking.

Fig. S3. Ability of purified δAg to bind to RNAs of PSTVd (panel A) and RMRP (panel B). Assays were performed as in Fig. 6. In panel B, the second and slower band represents dimers of the 359 nt PSTVd RNA.

Fig. S1

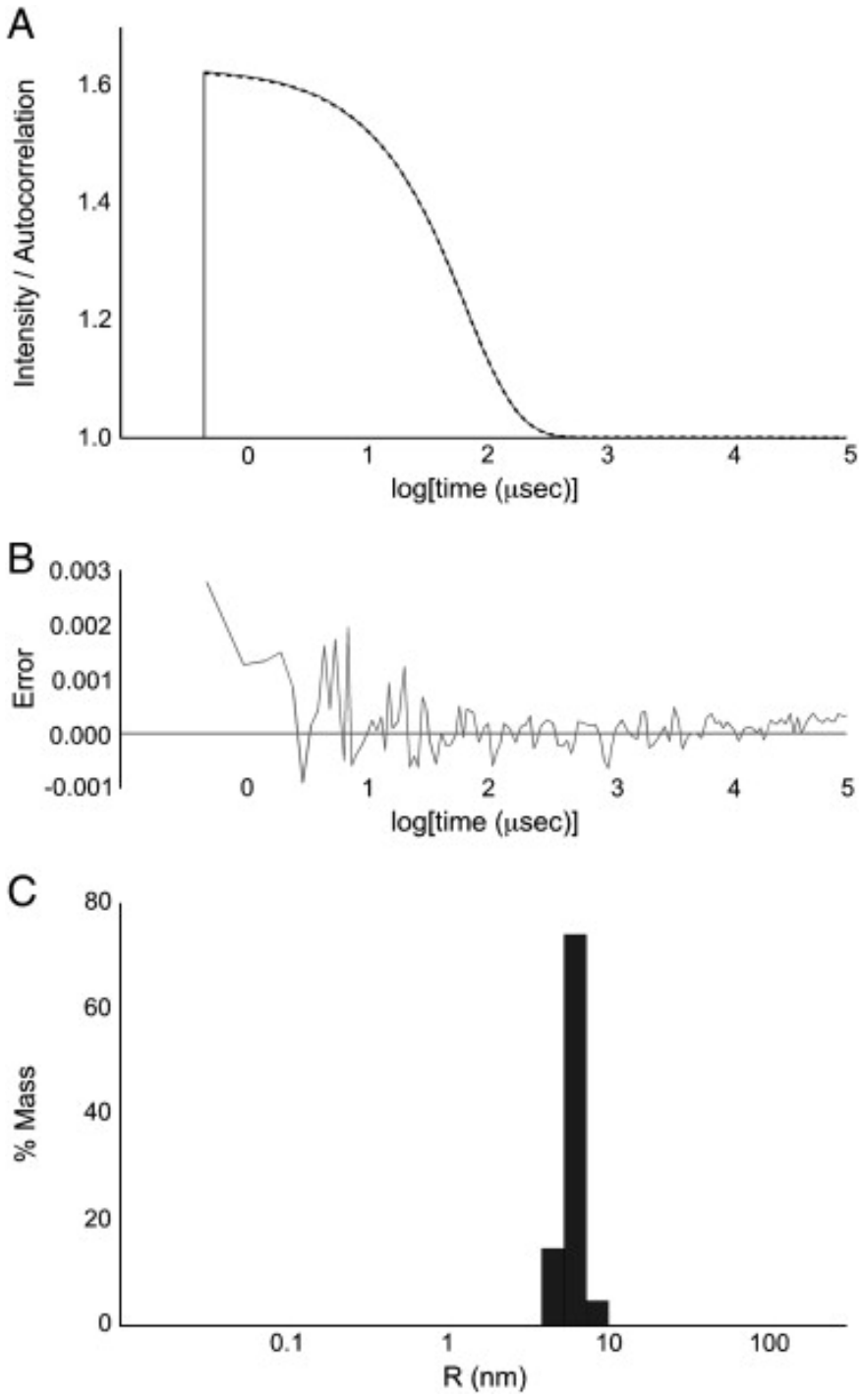

A
Intensity / Autocorrelation
1.6
1.4
1.2
1.0
0
1
2
3
4
5
log[time (μsec)]
B
Error
0.003
0.002
0.001
0.000
-0.001
0
1
2
3
4
5
log[time (μsec)]
C
% Mass
80
60
40
20
0
0.1
1
10
100
R (nm)

Fig. S2

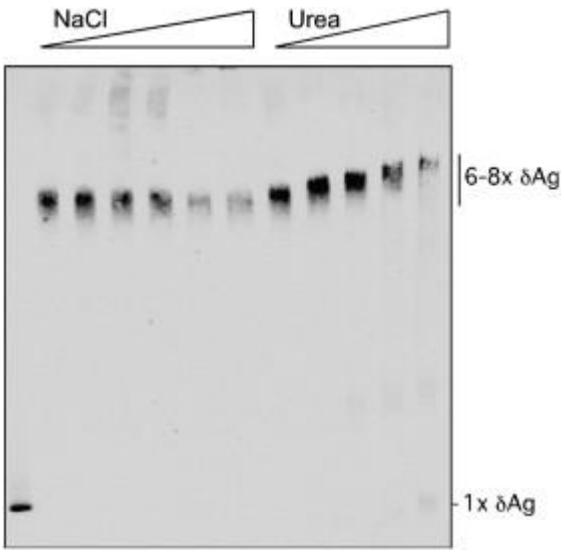
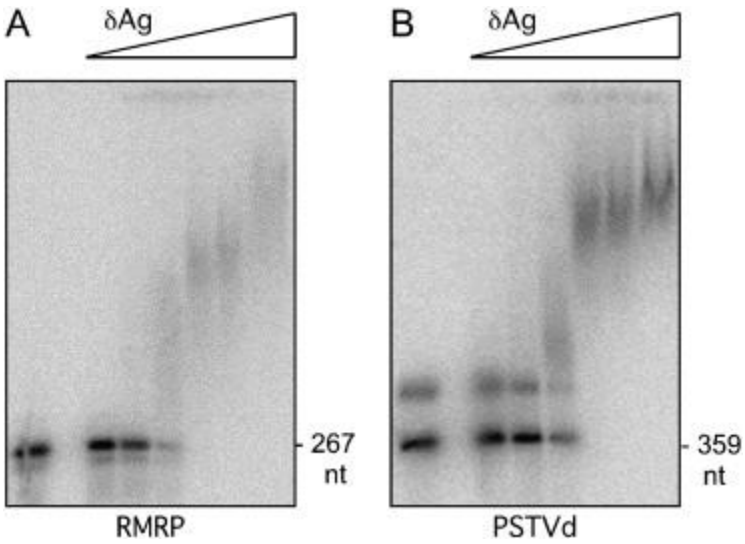


Fig. S3



5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding[2]

[2] Manuscript as submitted to Virology Journal

Multimerization of the hepatitis delta virus antigen is not essential for in vitro nucleic acid binding

Carolina Alves, Ana Casaca[*], Celso Cunha[§]

[1]Medical Microbiology Unit, Center for Malaria and Tropical Diseases, Institute of Hygiene and Tropical Medicine, Nova University, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

[*]Current address: Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 2780-156 Oeiras, Portugal

[§]Corresponding author.

Email addresses:

CA: carolinamantero@gmail.com

AC: acasaca@igc.gulbenkian.pt

CC: ccunha@ihmt.unl.pt

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Abstract

Background

The small antigen (S-HDAg) of hepatitis delta virus (HDV), a basic protein of 195 amino acids, is essential for the accumulation of HDV RNAs in infected cells. We have previously shown that purified recombinant S-HDAg is able to multimerize and binds, without specificity, to nucleic acids. Others have solved the structure of a peptide spanning amino acids 12-60, and obtained evidence for an antiparallel coiled-coil dimerization. This and other data are consistent with this region being responsible for dimerization and even higher order multimerization. In the present study we further characterized a truncated form of S-HDAg, spanning amino acids 61 - 195, a region we have predicted to be intrinsically disordered and has yet to be structurally characterized.

Methods

Expression and purification of the full-length and the truncated protein were performed and used in multimerization assays. The proteins were analyzed by denaturing polyacrylamide gel electrophoresis with and without prior glutaraldehyde cross-linking. Additionally, we analyzed the proteins ability to bind nucleic acids through electrophoretic mobility shift assays.

Results

The truncated protein failed to multimerize and yet was still able to bind nucleic acids, again without specificity. This binding ability was equally sensitive to increasing pH, for both full-length and truncated S-HDAgs.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Conclusions

Some studies found in literature, indicate that S-HDAg dimerization and even multimerization are needed for the biological activities of this essential protein, including its ability to bind nucleic acids, particularly the viral RNAs. However, the data here obtained indicate that nucleic acid binding can be achieved without dimerization and multimerization and with no apparent specificity for viral RNA.

Key words

Hepatitis delta virus, small delta antigen, intrinsically disordered protein, protein multimerization, nucleic acid binding

Background

Hepatitis delta virus (HDV) is the human pathogen with the smallest RNA genome known to date. It is a defective virus that requires the presence of hepatitis B virus (HBV) to propagate infection since the HDV envelope is formed by HBV surface antigens. HDV co-infection with HBV or HDV super-infection of hepatitis B patients, increases the severity of acute and chronic liver disease [1]. The HDV genome is a 1.7kb single-stranded circular RNA with only one open reading frame (ORF) coding the delta antigen (HDAg). During the HDV replication cycle, site-specific editing of the antigenomic RNA by a host adenosine deaminase results in the expression of a second form of the HDAg. The mRNA for the original small delta antigen (S-HDAg), corresponding to a 195 amino acids long protein, has its stop codon changed to a tryptophan residue, so the ORF is extended by 19 amino acids. This results in the

translation of the large delta antigen (L-HDAg) with 214 amino acids. None of the HDAg forms has any known enzymatic activity and although both forms share most of their sequence and therefore functional domains, they display different functions in the HDV replication cycle. It has been shown that S-HDAg is essential for HDV RNA accumulation while L-HDAg acts as a dominant inhibitor of replication and is essential for virus assembly [1].

It has been previously shown that S-HDAg is an intrinsically disordered protein (IDP) by using a meta-predictor of intrinsic disorder as well as measurements of circular dichroism [2]. In fact S-HDAg has several characteristics commonly attributed to IDPs. IDPs rarely contain enzymatic activity and are commonly involved in nucleic acid binding and/or interactions with other proteins [3]. S-HDAg is a nucleic acid-binding protein and is also rather promiscuous, with multiple host protein partners identified through different approaches; although the role of these interactions is mostly unclear [4-6]. S-HDAg is also a basic protein with an expected net charge of +12 at neutral pH [7]. This is common in IDPs and may play an important role in the protein's ability to bind negatively charged nucleic acids. Posttranslational modification (PTM) sites, such as phosphorylation, acetylation, fatty acid acylations, glycosylation, methylation or ubiquitination, are also a common feature to IDPs. This is likely due to the fact that these modifications can induce structural changes in the protein, conferring distinct conformations. It has been described that S-HDAg is modified by phosphorylation, methylation, acetylation and more recently sumoylation, and different functions have been attributed to the different modified forms of the antigen [8-13].

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

The evidence, showing that S-HDAg is an IDP, can explain why little success has been achieved in solving its 3-D structure. To date, despite several attempts from at least three different groups, no one has been able to obtain crystals of the full-length antigen. However, crystals were readily achieved for a truncated form spanning amino acids 12 to 60, the least disordered region of the protein [2, 14]. This segment is involved in S-HDAg dimerization and was designated a coiled-coil domain (CCD) as dimers form an anti-parallel coiled-coil [14].

The present study was undertaken to assess the importance of the other regions of the S-HDAg in both multimerization and nucleic acid binding. For this we generated and purified a truncated form of S-HDAg lacking the first 60 amino acids and analyzed its multimerization and nucleic acid binding properties.

Results

Multimerization ability of Δ60HDAg

We have observed in a preceding work, that recombinant full-length S-HDAg is able to form homomultimers *in vitro* just as well as S-HDAg present in viral particles [2]. We were able to observe aggregates as high as 6- to 8-mers of recombinant S-HDAg.

In the present study, we further characterized the highly disordered region of S-HDAg. A truncated protein, Δ60HDAg, was constructed, which lacked the previously characterized dimerization domain, as represented in Figure 1. After expression and purification the purity of the recombinant protein was confirmed by denaturing SDS gel electrophoresis, in Figure 2A.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

This mutant protein was first analyzed to confirm the negative impact of the deletion in its multimerization ability. Hence, Δ60HDAg was analyzed by SDS-PAGE with and without prior glutaraldehyde cross-linking. As can be seen in Figure 2A**,** the un-cross-linked protein displayed a well-defined single band. The truncated protein has an estimated weight of 15 kDa although in the SDS-PAGE analysis the monomer appears to be a ~19 kDa protein, when compared with the molecular weight marker. However, it has been observed before, that SDS-PAGE over estimates the size of full-length HDAg [15,16]. It seems that this characteristic also applies to the truncated protein.

When the truncated protein was cross-linked before the electrophoretic analysis we observed no major mobility changes. That is, only the single band corresponding to the monomers was detected, even if slightly fuzzier as glutaraldehyde concentration increased. This is compatible with there being no aggregation of Δ60HDAg *in vitro,* even with increasing glutaraldehyde concentrations.

Following this result we asked whether the truncated protein could form aggregates in the presence of the full-length antigen or if it would affect the latter's ability to multimerize. To do so the proteins were mixed in a 1:1 molar ratio and analyzed by SDS-PAGE, with or without prior cross-linking. In Figure 2B**,** it can be seen as expected that, without any cross-linking, two bands were detected corresponding to the full-length S-HDAg and the truncated form Δ60HDAg. Relative to the MW markers, one band has an apparent MW of ~19 kDa and the other, ~28 kDa, corresponding to the truncated and the full-length antigens, respectively. Although the proteins have a MW of around 15 kDa for Δ60HDAg and 21 kDa for S-HDAg, again we observed the over

estimation of the HDAg molecular weight characteristic when using this analytical method.

When the protein mixture was cross-linked by adding increasing concentrations of glutaraldehyde of 0.01% and 0.1%, the full-length protein formed high MW aggregates but the truncated form appeared to have the same electrophoretic mobility (Figure 2B). This shows that in the presence of full-length S-HDAg the Δ60HDAg cannot form aggregates and remains as a monomer even in the presence of high glutaraldehyde concentration. It can also be observed that the presence of Δ60HDAg does not seem to affect the S-HDAg multimerization ability as aggregates form apparently with similar efficiency as when S-HDAg is alone [2].

RNA binding ability of Δ60HDAg

In a previous work, we have shown that the full-length S-HDAg can bind nucleic acids as a multimer [2]. S-HDAg was cross-linked and, following glutaraldehyde neutralization, the multimers were incubated with different nucleic acids. We found no binding specificity in our *in vitro* conditions as the full-length protein bound HDV RNA as well as a dsDNA molecular weight marker [2]. In the present study we analyzed the nucleic acid binging ability of Δ60HDAg, which, as we have just shown, is unable to multimerize.

Δ60HDAg was incubated, at increasing concentrations ranging from 0 to 3μM with a fixed amount of a linear HDV RNA, prior to a gel retardation assay. As can be observed in Figure 3, there was a mobility shift as the Δ60HDAg concentration increased, relative to the nucleic acid alone.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

As mentioned before, S-HDAg showed no apparent nucleic acid binding specificity as it binds to HDV RNA of different shapes and sizes, non-HDV RNA and even DNA [2]. We decided to assess if the lack of multimerization would have an impact on nucleic acid binding specificity. Δ60HDAg was incubated with dsDNA and the interaction was analyzed through a gel retardation assay in the same conditions as before. A shift in the DNA mobility was detected, as observed in Figure 4A, showing that at Δ60HDAg concentrations higher than 4μM, a high molecular weight complex begins to form between the DNA and truncated protein. This complex was unable to enter the gel, and the DNA is detected at the well. This result shows that *in vitro*, Δ60HDAg is able to interact with dsDNA. Therefore, even though it cannot form homomultimers, this truncated delta antigen is still able to interact with different nucleic acids. This interaction is not specific in our *in vitro* conditions similarly to what had been observed before with the full-length antigen [2].

Our *in vitro* conditions for the nucleic acid binding assays were at a standard 7.5 pH. We chose to analyze the Δ60HDAg-nucleic acid interactions at a higher pH, closer to, but still lower than, the protein's estimated pI (Figure 1). A protein's net charge is determined by its amino acid compositions but it is also influenced by the solvent's pH. As Δ60HDAg has an estimated pI of 9.84 it should behave has a positively charged protein at a neutral pH, just as the full-length S-HDAg has an estimated +12 net charge, at neutral pH [7].

At a pH higher than neutral the Δ60HDAg net positive charge would be lower, and this may also effect the protein's conformation. Therefore, increasing concentrations of Δ60HDAg were incubated with a fixed amount of dsDNA at pH 9.5. At this higher pH

we observed a change in Δ60HDAg's ability to bind nucleic acids. In Figure 4B, we detected a change in the DNA mobility at the highest Δ60HDAg concentration when compared with free DNA. However, not all the DNA had shifted since a band with the same mobility as free DNA can be clearly seen, as well as the DNA-Δ60HDAg aggregate close to the well. Comparing with the assay performed at a neutral pH, figure 4A, we observed that an increase in pH would impair the truncated protein's ability to interact with DNA. At pH 7.5, when 4 µM of Δ60HDAg was present high MW aggregates started to appear. In contrast, at pH 9.5 this only begun to occur at 8 µM Δ60HDAg. Also, at pH 7.5 and Δ60HDAg at 12 µM, we did not detect free DNA, consistent with it all forming complexes with the protein. This was not achieved at pH 9.5 since in these conditions we detected free DNA even at the highest Δ60HDAg concentration. A similar impact of increased pH on nucleic acid binding ability was observed for the full-length S-HDAg (Figure 4C and D).

Discussion

S-HDAg is the only HDV encoded protein, along with its longer form L-HDAg; and even though it lacks any known enzymatic activity it is thought to play a crucial role in HDV replication cycle, as it is essential for genomic RNA accumulation in infected liver cells. This small 195 amino acids long protein is highly promiscuous as it not only binds to HDV RNA to form viral ribonucleoproteins but also interacts with several host factors, although the biological meaning of most interactions remains to be clarified [4-6]. S-HDAg was predicted to be an IDP, which is consistent with many of its known characteristics such as high net charge, promiscuity in binding several partners and lack of enzymatic activity [2]. The high level of intrinsic disorder found in S-HDAg can

justify the difficulties found in achieving its crystal structure. So far, only a peptide spanning the first 60 amino acids was crystallized, and consequently, had its structure characterized [14].

In the present study we aimed to gain further insight on the region of S-HDAg that was not crystallized and lacks the CCD domain, involved in S-HDAg multimerization. A truncated version of S-HDAg, consisting of amino acids 61 through to 195, was expressed and purified, and characterized by different assays. The truncated protein, designated Δ60HDAg, is schematized in Figure 1.

Accordingly, our first approach meant to confirm that, without the CCD, the truncated protein was unable to multimerize. Hence, Δ60HDAg was analyzed by SDS-PAGE with prior glutaraldehyde cross-linking and the result shows, in Figure 2A, that Δ60HDAg is unable to form multimers *in vitro*. We also observed, in Figure 2B, that Δ60HDAg does not interfere with the multimerization of the full-length S-HDAg. Furthermore, it appears that Δ60HDAg cannot multimerize even in the presence of full-length S-HDAg, which can clearly form aggregates *in vitro* (Figure 2B). We cannot exclude that a small fraction of the truncated protein is present in the multimers formed by the full-length S-HDAg in the presence of glutaraldehyde. However, the majority of the Δ60HDAg is present as a monomer since we observe no major changes in the bands' intensities except for the same fuzziness displayed when the truncated protein was cross-linked in the absence of S-HDAg, in Figure 2A.

Following the confirmation that Δ60HDAg is unable to multimerize we assessed its ability to bind nucleic acids. S-HDAg is an RNA-binding protein in its native form so

we decided to evaluate the binding ability of the truncated form. An HDV RNA was transcribed *in vitro* and incubated with Δ60HDAg prior to electrophoretic mobility shift assays. As displayed in Figure 3, there is a clear shift in mobility as higher concentrations of Δ60HDAg are incubated with fixed amounts of HDV RNA, relative to the nucleic acid alone. Noticeably, the mobility shift resulting from the interaction between Δ60HDAg and RNA is a gradual one. This is unlike what we observed in previous work for the nucleic acid interactions with the full-length S-HDAg, where there was a clear shift from the unbound to the bound state without any intermediate positions [2]. Here, since ethidium bromide staining was the method of detection, we used significantly higher amounts of nucleic acid (100 ng). The different pattern observed may be due to the fact that, in the present study, we did not use relatively minor amounts of radiolabeled nucleic acid, as in our previous work.

In a second mobility shift assay, we tested Δ60HDAg's ability to bind dsDNA. For that, the truncated protein was incubated, at increasing concentrations, with a dsDNA fragment. The result, shown in Figure 4A, reveals that, in the tested *in vitro* conditions, Δ60HDAg interacts with dsDNA since there is a clear shift in mobility when Δ60HDAg was added to the nucleic acid. However, there is a difference in the mobility shift achieved when the protein is binding RNA or dsDNA. As seen in Figure 3, there is a gradual shift in mobility has the Δ60HDAg concentration increases, consistent with a gradual increase in the molecular weight of the complexes formed by the protein and HDV RNA. However, when Δ60HDAg interacts with dsDNA, we observe, in Figure 4A, no gradual change but an abrupt shift between the unbound and bound DNA. This shift was as observed when using trace amounts of nucleic acids in our previous work

[2]. That is, as the protein concentration increases we observe two bands corresponding to unbound DNA and complexed DNA until, at 12µM Δ60HDAg, all DNA appears to be bound to protein and no unbound DNA is detected. This difference in the shifts may reflect a difference in binding specificity, sugeesting that, if Δ60HDAg is more specific for RNA it may bind immediately, without the need for a minimum concentration for interaction to occur, leading to a gradual shift.

As we have observed in Figures 3 and 4A, Δ60HDAg is able to interact with HDV RNA and dsDNA, leading to the conclusion that this truncated protein displays no nucleic acid binding specificity, at least *in vitro*, just has the full-length S-HDAg [2]. The lack of specificity concurs with the idea that the high net charge of this basic protein facilitates its binding to negatively charged nucleic acids, independently of their nature. Nonetheless, it is still possible that HDV RNA binding specificity occurs *in vivo* for the native S-HDAg. The antigen is extensively phosphorylated *in vivo*, which likely reduces its high net charge, limiting the ability to be involved in non-specific electrostatic interactions [17].

Another group has reported that a truncated form of S-HDAg has RNA binding specificity and also that there is a requirement that the RNA must have a minimum of ~311 nt of rod-like folding for binding to occur [18]. However, their results were obtained while studying a form of S-HDAg, spanning amino acids 1-160, that is truncated at the C-terminus, and possessing a His-tag fused to the N-terminus. In agreement with our results, they observed no specificity when the full-length S-HDAg interacted with nucleic acids [18]. Yet it remains puzzling how removing part of the disordered structured led to the “new found” specificity.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

The results of the present study also disagree with early findings that, again, indicated that S-HDAg specifically bound only HDV RNAs with rod-like folding [19]. Once more, the binding assays used a protein A tagged fusion protein and the results were analyzed by northwestern [19].

Regarding the impact of multimerization on the nucleic acid binding ability of S-HDAg, we have shown that the oligomerization of the protein is not a prerequisite for its ability to interact with nucleic acids, at least *in vitro*, although the full-length S-HDAg forms multimers, even *in vivo* in viral particles [2]. However, others have reported that multimerization is a prerequisite for nucleic acid binding specificity [20]. Again, in this study a different form of S-HDAg was tested, they used the same C-terminal truncated and N-terminal fusion protein that had shown binding specificity. Furthermore, when a point mutation, that depleted the multimerization ability, was analyzed, a lack of binding specificity was found [18, 20].

Conclusion

In conclusion, the data presented here show that multimerization of S-HDAg is not required for the protein to bind nucleic acids *in vitro*. Also, just as previously reported for the full-length S-HDAg, we observe that a multimerization deficient mutant binds nucleic acids with no apparent specificity. The lack of specificity may be due to an electrostatic interaction, between the positively charged protein and the negatively charged nucleic acids, or a consequence of the antigen's disorder that grants flexibility to adapt to different partners.

Methods

Expression plasmids

As previously described, plasmid pR5δV5 has been used to express high levels of full-length S-HDAg in *Escherichia coli* [21]. This plasmid DNA was used a template with specific primers to amplify the region encoding amino acids 61-195 (with forward primer: 5'-TTTCAATTGCCAAAGATAAAGATGGCG-3', reverse primer:5'-TTTCTCGAGTTACGGAAAGCC-3'). The amplified sequence was cloned into the *Eco*RI - *Xho*I cloning site of pGEX-6P-2 (GEHealthcare). The resulting plasmid, designated pGEX-6P-2-Δ60HDAg, allowed expression of the fusion protein GST-Δ60HDAg.

Bacterial expression and purification of recombinant proteins

Expression and purification of full-length S-HDAg was performed as described elsewhere [6]. Expression of GST-Δ60HDAg was obtained in BL21 (DE3) codon plus (Novagen) competent cells, following transformation with plasmid pGEX-6P-2-Δ60HDAg. A single bacterial colony was first inoculated into a 100 ml overnight culture. Then 10 ml was inoculated into a 1 L culture of 2xYTA media. Protein expression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) when optical density was between 0.6 and 0.8. After an additional 3 h, cells were pelleted at 3,000 rpm for 7 min.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

For the subsequent purification method cells expressing the GST-Δ60HDAg were resuspended in PBS. Protease inhibitors were added (cOmplete, Roche). Cells were lyzed by four freeze-thaw cycles after the addition of 0.1mg/mL lisozyme. The lysates were treated with DNase 1 (Roche) and centrifuged at 14,000xg for 10 min. The supernatants were analyzed by SDS-PAGE followed by western blot to detect the presence of the recombinant protein. Total protein extracts were stored at -80C until further use. The GST tag was removed by PreScission protease (GE Healthcare) cleavage as instructed. Δ60HDAg was concentrated using Protein concentrating solution and dialysis cassettes (Pierce) as suggested by the manufacturer.

RNA synthesis

Plasmid pDL542 has a T7 promoter to express full-length antigenomic HDV RNA [22]. The plasmid was transcribed *in vitro* using a T7 RiboMax transcription system (Promega) to obtain HDV RNA, following protocols provided by the manufacturer.

Gel electrophoresis

Protein samples were analyzed by denaturing electrophoresis in 12% polyacrylamide gels with SDS denaturation. Proteins were detected by Coomassie blue staining. In the cross-linking experiments, the protein samples were treated with 0.01% or 0.1% of glutaraldehyde for 10 min at room temperature. Glutaraldehyde was then inactivated by the addition of 100 mM ammonium acetate, followed by SDS denaturation and gel electrophoresis.

For the analysis of protein-nucleic acid interactions the protein samples were diluted in a standard binding buffer containing 150 mM NaCl and 10 mM Tris-HCl (pH7.5) or as

otherwise stated. After 10 min incubation at room temperature, the nucleic acids were added and the mix was incubated for an additional 10 min prior to the electrophoretic mobility shift assays. For the study of DNA-protein interaction we used the PCR product obtained in the cloning of the truncated protein. The mobility shift assays were performed in non-denaturing 1.5% agarose gels (1x TBE). The nucleic acids were subsequently detected by staining with ethidium bromide.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

CA participated in the study design, carried out the experiments, and drafted the manuscript. AC carried out the expression and purification of the full-length S-HDAg and helped draft the manuscript. CC conceived the study, participated in its design and coordination, and drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

CA and AC were supported by Ph.D. grants from Fundação para a Ciência e Tecnologia, Portugal.

References

1. Rizzetto M: **Hepatitis D: thirty years after**. *J Hepatol* 2009, **50**:1043–1050.

2. Alves C, Cheng H, Roder H, Taylor J: **Intrinsic disorder and oligomerization of the hepatitis delta virus antigen.** *Virology* 2010, **407**:333–340.

3. Uversky VN: **Intrinsically disordered proteins from A to Z**. *Int J Biochem Cell Biol* 2011, **43**:1090–1103.

4. Chang MF, Baker SC, Soe LH, Kamahora T, Keck JG, Makino S, Govindarajan S, Lai MM: **Human hepatitis delta antigen is a nuclear phosphoprotein with RNA-binding activity.** *J Virol* 1988, **62**:2403–2410.

5. Greco-Stewart V, Pelchat M: **Interaction of Host Cellular Proteins with Components of the Hepatitis Delta Virus**. *Viruses* 2010, **2**:189–212.

6. Casaca A, Fardilha M, da Cruz e Silva E, Cunha C: **The heterogeneous ribonuclear protein C interacts with the hepatitis delta virus small antigen**. *Virol J* 2011, **8**:358.

7. Kuo M, Goldberg J, Coates L, Mason W, Gerin J, Taylor J: **Molecular cloning of Hepatitis Delta Virus RNA from an Infected Woodchuck Liver: Sequence, Structure, and Applications**. *J Virol* 1988, **62**:1855–1861.

8. Mu J-J, Wu H-L, Chiang B-L, Chang R-P, Chen D-S, Chen P-J: **Characterization of the Phosphorylated Forms and the Phosphorylated Residues of Hepatitis Delta Virus Delta Antigens**. *J Virol* 1999, **73**:10540–10545.

9. Chen Y-S, Huang W-H, Hong S-Y, Tsay Y-G, Chen P-J: **ERK1/2-Mediated Phosphorylation of Small Hepatitis Delta Antigen at Serine 177 Enhances Hepatitis Delta Virus Antigenomic RNA Replication**. *J Virol* 2008, **82**:9345–9358.

10. Hong S-Y, Chen P-J: **Phosphorylation of serine 177 of the small hepatitis delta**

antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *J Virol* 2010, **84**:1430–1438.

11. Tseng C-H, Jeng K-S, Lai MMC: **Transcription of subgenomic mRNA of hepatitis delta virus requires a modified hepatitis delta antigen that is distinct from antigenomic RNA synthesis**. *J Virol* 2008, **82**:9409–9416.

12. Mu J, Tsay Y, Juan L, Fu T, Huang W, Chen D, Chen P: **The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis**. *Virology* 2004, **319**:60–70.

13. Li Y-J, Stallcup MR, Lai MMC: **Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication.** *J Virol* 2004, **78**:13325–13334.

14. Zuccola H, Rozzelle J, Lemon S, Erickson B, Hogle J: **Structural basis of the oligomerization of hepatitis delta antigen**. *Structure* 1998, **6**:821–830.

15. Bergmann KF, Pohl C, Gerin JL: **Characterization of proteins of hepatitis delta virus.** *Prog Clin Biol Res* 1987, **234**:105–110.

16. Gerlich WH, Heermann KH, Ponzetto A, Crivelli O, Bonino F: **Proteins of hepatitis delta virus.** *Prog Clin Biol Res* 1987, **234**:97–103.

17. Mu JJ, Wu HL, Chiang BL, Chang RP, Chen DS, Chen P-J: **Characterization of the phosphorylated forms and the phosphorylated residues of hepatitis delta virus delta antigens.** *J Virol* 1999, **73**:10540–10545.

5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

18. Defenbaugh DA, Johnson M, Chen R, Zheng YY, Casey JL: **Hepatitis Delta Antigen Requires a Minimum Length of the Hepatitis Delta Virus Unbranched Rod RNA Structure for Binding**. *J Virol* 2009, **83**:4548–4556.

19. Chao M, Hsieh SY, Taylor J: **The antigen of hepatitis delta virus: examination of in vitro RNA-binding specificity**. *J Virol* 1991, **65**:4057–4062.

20. Lin B, Defenbaugh D, Casey J: **Multimerization of hepatitis delta antigen is a critical determinant of RNA binding specificity**. *J Virol* 2010, **84**:1406–1413.

21. Dingle K, Bichko V, Zuccola H, Hogle J, Taylor J: **Initiation of hepatitis delta virus genome replication**. *J Virol* 1998, **72**:4783–4788.

22. Lazinski D, Taylor J: **Expression of hepatitis delta virus RNA deletions: cis and trans requirements for self-cleavage, ligation, and RNA packaging.** *J Virol* 1994, **68**:2879–2888.

Figures

Figure 1 - Schematic representation of S-HDAg and Δ60HDAg

CCD represents the coiled-coil domain between amino acids 12 to 60; NLS is the nuclear localization signal from amino acids 67 to 75; RBD corresponds to the RNA binding domain comprised within amino acids 97 and 146. The amino acid sequence corresponds to Δ60HDAg and the Δ60HDAg's isoelectric point (pI) and molecular weight (MW) were estimated using Expasy (www.expasy.org).

Figure 2 - S-HDAg and Δ60HDAg multimerization ability

In panels A and B, purified recombinant protein was cross-linked with increasing concentrations of glutaraldehyde (0, 0.01 and 0.1%) prior to SDS-PAGE. The proteins were detected by Coomassie blue staining. In panel A, only purified Δ60HDAg was present at 2 µM and in panel B both S-HDAg and Δ60HDAg were present at 2 µM each.

Figure 3 - Gel retardation assay: Δ60HDAg binding HDV RNA.

Purified recombinant Δ60HDAg was incubated, in standard pH 7.5 binding buffer, with 100 ng of HDV RNA at increasing concentrations (0; 0.5; 1.5 and 3 µM; lanes 1-4, respectively). At left is a RNA marker (RiboRuler High Range RNA Ladder, Fermentas).

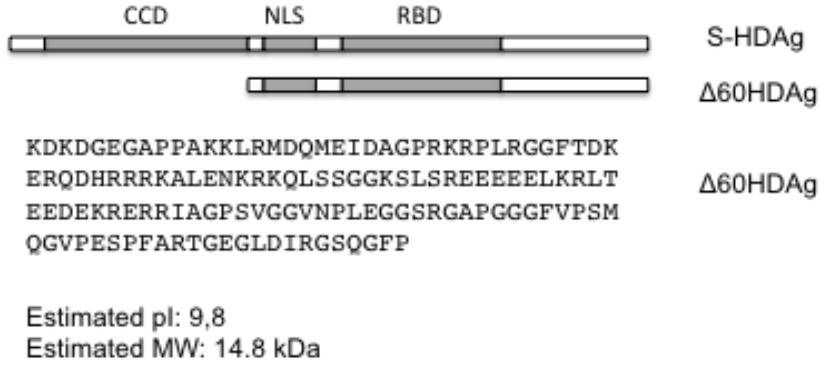
5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Figure 4 - Gel retardation assays: Δ60HDAg binding DNA and S-HDAg binding HDV RNA.

In panel A, 100 ng of dsDNA were incubated in standard pH 7.5 binding buffer, with increasing concentrations of purified recombinant Δ60HDAg (0, 2, 4, 6, 8, 10, and 12 μM). Panel B shows the assay in binding buffer at pH 9.5. Recombinant Δ60HDAg was incubated with 100 ng of dsDNA, at different concentrations (0, 2, 4, 6, 8, 10, and 12 μM). In panel C, 100 ng of HDV RNA was incubated, in standard pH 7.5 binding buffer, with increasing concentrations of purified recombinant S-HDAg (0, 0.5, 1, 1.5, and 2 μM). In panel D, 100 ng of HDV RNA were incubated in a pH 9.5 binding buffer, at different concentrations of purified recombinant S-HDAg (0, 0.5, 1, 1.5, and 2 μM).

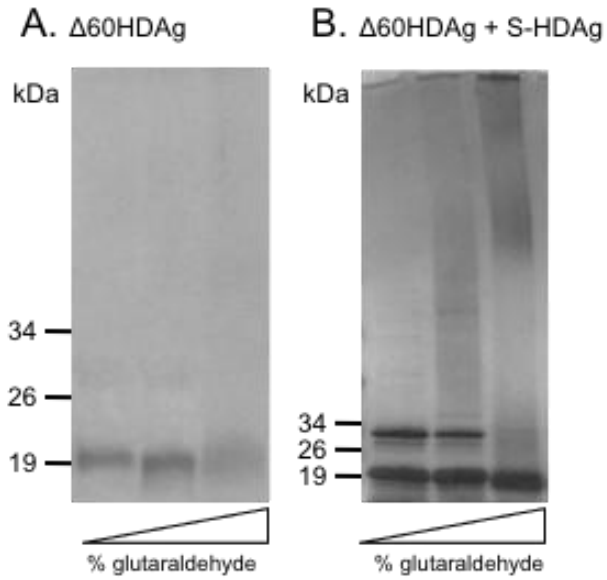
5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Figure 1



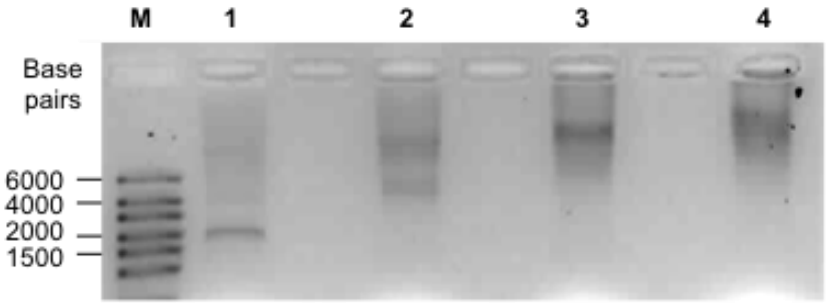
5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Figure 2



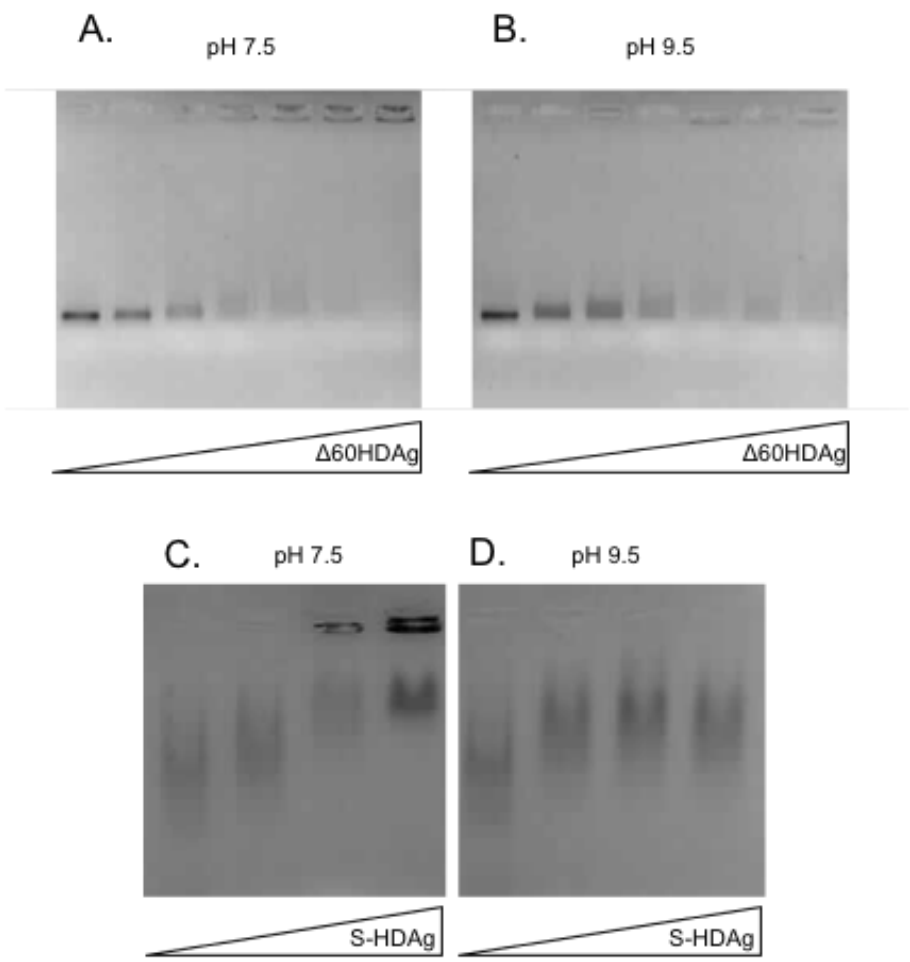
5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Figure 3



5. Multimerization of the Delta Antigen is not Essential for *In Vitro* Nucleic Acid Binding

Figure 4



6. Discussion and Future Perspectives

As considered in the Introduction, the replication of the genome of the hepatitis delta virus is a complex phenomenon. It is deceptive that the RNA genome at ~1,700 nucleotides is smaller than that of any other animal virus. It is equally deceptive that the small 195 amino acids HDAg is the only protein encoded and is essential for the replication of the RNA genome. All this means that the RNA and the HDAg also have to make many and essential interactions with components of the host cell.

The smallest form of HDAg is a crucial player in the HDV replication cycle. While the L-HDAg is essential in the formation of new virions, allowing for HDV propagation, it is S-HDAg that is essential for the accumulation of HDV RNA, possibly playing a direct role in HDV RNA replication. Thus, the subject of the present study was S-HDAg, and, by characterizing this antigen, we aimed to, albeit indirectly, get a deeper insight into the putative roles of this viral protein.

We began our Specific Aim 1 by addressing the significance of S-HDAg subcellular localization in the presence and absence of HDV genome replication. We found that S-HDAg locates predominantly in the nucleolus in the absence of HDV RNA but is observed in the nucleoplasm in the presence of HDV RNA.

The nucleolar pattern may be justified by S-HDAg's known interaction with nucleolar proteins nucleolin and nucleophosmin. But, since S-HDAg only appears to occur in the nucleolus in the absence of HDV RNA this may be deemed an irrelevant pattern for HDV RNA transcription. Then again, this may instead be a necessary transient step as nucleolar transit has been reported crucial for some viruses. Evidence that virus proteins can target the nucleolus has been reported for DNA and RNA viruses (Hiscox, 2002). This has led to the suggestion that viruses use such targeting to enhance viral replication and alter host cell transcription and cell cycle regulation. Also, the observed lack of nucleolar localization of S-HDAg expressed in the presence of HDV RNA, is not necessarily contradictory with reports that host pol I, a nucleolar protein, may be involved in HDV antigenomic RNA synthesis. For example, S-HDAg may not be directly involved in antigenome synthesis and instead have an indirect role such as

HDV genome transport to the nucleolus or chaperone-like functions for the accumulation of processed HDV antigenomes.

To conclude our Specific Aim 1 we assessed how S-HDAg interacted with the host cell by observing co-localization with relevant host factors in different nuclear compartments.

When S-HDAg presented a nucleolar pattern, co-localization with host nucleolin was observed as expected. Some time after S-HDAg induction both S-HDAg and nucleolin migrated to the nucleoplasm. The same re-location was also observed when cell stress was induced. However, S-HDAg is unlikely to interact directly with nucleolin as, in the presence of non-replicating HDV RNA, S-HDAg locates in the nucleoplasm and nucleolin maintains its nucleolar pattern. Considering that S-HDAg is a nucleic acid-binding protein, we hypothesize that unspecific binding to ribosomal RNA precursors is involved in S-HDAg nucleolar pattern. It is also possible that specific post-translational modifications of S-HDAg may alter its intracellular localization.

S-HDAg presence in the nucleoplasm seems particularly relevant since pol II, the host polymerase thought to be responsible for HDV genome replication, is present in the nucleoplasm. It is likely that some direct or indirect interaction occurs between host pol II and S-HDAg as both proteins seem to be essential for HDV RNA transcription and were shown to interact *in vitro*. In the absence of HDV RNA, S-HDAg localizes to the nucleolus, and thus is unable to co-locate with the majority of host pol II.

Nevertheless, in the presence of HDV RNA, we observed the co-localization of the two proteins. However, there is a caveat, in that co-localization of S-HDAg and pol II is consistent with but not proof of an interaction of functional significance. Co-localization is also observed after HDV RNA accumulation has reached a maximum level and therefore transcription has ceased (Chang et al., 2005). Even S-HDAg mutants that are unable to support HDV RNA accumulation, co-locate with host pol II. Also, when pol II activity is inhibited by α-amanitin the co-localization with S-HDAg is still observed. As such, although we detect co-localization in the cell nucleoplasm of S-HDAg and Pol II, our data suggest that the proteins are not necessarily involved in active HDV RNA transcription.

In this Aim we have thus found that, although S-HDAg is essential for HDV RNA accumulation, our observations more support the interpretation that S-HDAg functions as a chaperone that protects/stabilizes the processed transcripts rather than be directly involved in RNA transcription.

In Specific Aim 1 it remains unclear the biological significance of the observed patterns. Particularly, to further understand the nucleoplasmic distribution of S-HDAg, it would be interesting to perform *in situ* hybridization to assess the localization of the viral RNA.

In Specific Aim 2 we tested the hypothesis that S-HDAg has some level of intrinsic disorder. We observed that it has several characteristics common to intrinsically disordered proteins. As mentioned before, these include a high net positive charge, RNA binding ability, chaperone activity, no known enzymatic activity, ability to oligomerize, and binding promiscuity. We applied an *in silico* analysis of S-HDAg using a meta-predictor of disorder. We found that S-HDAg is largely disordered. Only a small portion of the CCD was predicted to be less disordered, corresponding to part of the S-HDAg peptide that was crystallized by Zuccola *et al.* in 1998. Our result was also independently confirmed by CD data, which showed that there is little secondary structure other than the CCD region.

From our findings in Specific Aim 2 we predict that the high level of intrinsic disorder constitutes important advantages for HDV. That is, such flexibility of S-HDAg, would enable it to structurally adapt and engage in a variety of homo- or hetero-multimer interactions with host proteins.

Regarding Specific Aim 2, it should be possible to determine S-HDAg's conformation ensembles or protein-cloud by combining data from different experimental approaches. An understanding of the protein's most likely conformations could also provide clues on how S-HDAg binds its different partners. It could even be assessed if the antigen assumes any rigid structure when bound to its partners. As part of such a follow-up on Specific Aim 2 we are currently expressing and purifying Δ60HDAg labeled with ^{15}N and/or ^{13}C isotopes for NMR analysis. CD data will also be

obtained for this recombinant purified protein. We hope to assemble information on S-HDAg structural features.

With our Specific Aim 3 we confirmed S-HDAg's ability to multimerize. We found that high molecular weight homomultimers of purified recombinant S-HDAg occur *in vitro*. Furthermore, multimerization is also likely to occur *in vivo* as high S-HDAg multimers were detected in virus-like particles. Also, in our *in vitro* experiments the concentrations of S-HDAg used were estimated to be ~10 times less than the average S-HDAg concentration within an infected cell. Thus we predict that multimers are even more likely to form *in vivo*. Both our cross-linking and light scattering results imply that recombinant S-HDAg alone forms mostly complexes of 6 up to 12 molecules.

Following this observation we tested the ability of the full-length recombinant S-HDAg to bind nucleic acids. We found that S-HDAg interacted with nucleic acids with no apparent specificity (RNA vs. DNA, single- vs. double-stranded vs. rod-like) even when S-HDAg multimers were cross-linked prior to the binding assays.

We interpret that the binding in our *in vitro* conditions is based on electrostatic interactions. S-HDAg has a very high net charge (+12) and a multimer may have an even greater total charge. Consequently, we observed a lack of specificity in binding negatively-charged nucleic acids *in vitro*.

To conclude our Specific Aim 3 we assessed the importance of multimerization in S-HDAg's nucleic acid binding ability. We designed and expressed an N-terminal truncated S-HDAg, Δ60HDAg, unable to multimerize and assessed its ability to bind nucleic acids. The results obtained were similar to those for the full-length protein. In our *in vitro* conditions we detected no binding specificity as Δ60HDAg interacted with HDV RNA as well as non-HDV RNA and DNA. Therefore we revised our earlier interpretation. That is, lack of binding specificity is largely due to electrostatic interactions between negatively charged nucleic acids and monomers of positively charged S-HDAg, and that multimerization, *per se*, is not essential for such interactions.

The observed lack of binding specificity is consistent with our hypothesis in Specific Aim 1 that S-HDAg binds rRNA precursors, in the absence of HDV RNA, thus locating in the nucleolus.

Our findings in Specific Aim 3 clearly differ from published reports that demonstrate S-HDAg's binding specifically to HDV RNA (Defenbaugh et al., 2009). However, we note that the authors used a different recombinant S-HDAg, with a C-terminal truncation, which was shown to be HDV RNA specific in their *in vitro* conditions. Similar to our findings, these authors did report a lack of specificity when using a full-length recombinant S-HDAg. When the same research group used their C-terminal truncated S-HDAg they observed that multimerization of S-HDAg is essential for HDV RNA binding (Lin et al., 2010). An obvious interpretation is that if one weakens the binding ability of the monomer then multimerization may be needed.

We speculate that *in vivo,* where extensive post-translational phosphorylation of the S-HDAg occurs (Mu et a., 1999) this will reduce the net positive charge of the monomer, so that multimerization may be needed for specific HDV RNA binding.

Specific Aim 3 should be continued in an attempt to clearly sate if S-HDAg's interaction with nucleic acid is specific for HDV RNA or not. Competition assays with different nucleic acids could be considered to determine if there is a preference for HDV RNA. Nonetheless, we would speculate that a disordered protein such as S-HDAg, which can bind to many distinct host proteins, is also likely able to adapt to different nucleic acids.

In conclusion, it is clear that, despite our observations presented here, and all the progress in understanding HDV biology, many issues remain unresolved. We hope that the work still in progress in our and other labs may contribute for a deeper knowledge of this fascinating pathogen.

7. References

Abgueguen, P., Birguel, J., Rameau, A., Ndong, J.-G., Sobnangou, J.-J., Aurenche, C., Ducancelle, A., Huraux, J.-M. and Lunel-Fabiani, F. (2010) High prevalence of infection by the hepatitis B (HBV) and hepatitis delta (HDV) viruses among pregnant women in Cameroon. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Alves, C., Freitas, N. and Cunha, C. (2008) Characterization of the nuclear localization signal of the hepatitis delta virus antigen. *Virology*, 370, 12–21.

Bahcecioglu, I. H., Aygun, C., Gozel, N., Poyrazoglu, O. K., Bulut, Y. and Yalniz, M. (2011) Prevalence of hepatitis delta virus (HDV) infection in chronic hepatitis B patients in eastern Turkey: still a serious problem to consider. *J. Viral Hepat.*, 18, 518–524.

Been, M. D. (2006) HDV ribozymes. *Curr. Top. Microbiol. Immunol.*, 307, 47–65.

Bichko, V. and Taylor, J. (1996) Redistribution of the delta antigens in cells replicating the genome of hepatitis delta virus. *J. Virol.*, 70, 8064-8070.

Bordier, B., Ohkanda, J., Liu, P., Lee, S.-Y., Salazar, F., Marion, P., Ohashi, K., Meuse, L., Kay, M., Casey, J., Sebti, S., Hamilton, A. and Glenn, J. (2003). In vivo antiviral efficacy of prenylation inhibitors against hepatitis delta virus. *J. Clin. Invest.*, 112, 407–414.

Branch, A. D., Benenfeld, B. J., Baroudy, B. M., Wells, F. V., Gerin, J. L. and Robertson, H. D. (1989) An ultraviolet-sensitive RNA structural element in a viroid-like domain of the hepatitis delta virus. *Science*, 243, 649–652.

Brazas, R. and Ganem, D. (1996) A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science*, 274, 90–94.

Cao, D., Haussecker, D., Huang, Y. and Kay, M. A. (2009) Combined proteomic-RNAi screen for host factors involved in human hepatitis delta virus replication. *RNA*, 15, 1971–1979.

Casaca, A., Fardilha, M., da Cruz e Silva, E. and Cunha, C. (2011) The heterogeneous ribonuclear protein C interacts with the hepatitis delta virus small antigen. *Virol.*

J., 8, 358.

Casey, J. L. (2006) Hepatitis delta virus RNA editing. *In*: Handa, H. and Yamaguchi, Y. eds. *Hepatitis Delta Virus*. Georgetown, TX: Landes Bioscience, 52-65.

Castelnau, C., Le Gal, F., Ripault, M.-P., Gordien, E., Martinot-Peignoux, M., Boyer, N., Pham, B.-N., Maylin, S., Bedossa, P., Dény, P., Marcellin, P. and Gault, E. (2006) Efficacy of peginterferon alpha-2b in chronic hepatitis delta: relevance of quantitative RT-PCR for follow-up. *Hepatology*, 44, 728–735.

Chang, J. and Taylor, J. (2002) In vivo RNA-directed transcription, with template switching, by a mammalian RNA polymerase. *EMBO J.*, 21, 157–164.

Chang, J., Gudima, S., Tarn, C., Nie, X. and Taylor, J. (2005) Development of a novel system to study hepatitis delta virus genome replication. *J. Virol.*, 79, 8182-8188.

Chang, J., Nie, X., Chang, H. E., Han, Z. and Taylor, J. (2008) Transcription of hepatitis delta virus RNA by RNA polymerase II. *J. Virol.*, 82, 1118–1127.

Chao, M., Hsieh, S. Y. and Taylor, J. (1990) Role of two forms of hepatitis delta virus antigen: evidence for a mechanism of self-limiting genome replication. *J. Virol.*, 64, 5066–5069.

Chao, M., Hsieh, S. Y. and Taylor, J. (1991) The antigen of hepatitis delta virus: examination of in vitro RNA-binding specificity. *J. Virol.*, 65, 4057–4062.

Chen, C. W. (2002) The double-stranded RNA-activated kinase, PKR, can phosphorylate hepatitis D virus small delta antigen at functional serine and threonine residues. *J. Bio. Chem.*, 277, 33058–33067.

Chen, C.-J., Yang, H.-I., Su, J., Jen, C.-L., You, S.-L., Lu, S.-N., Huang, G.-T., Iloeje, U., H.REVEAL-HBV Study Group. (2006) Risk of hepatocellular carcinoma across a biological gradient of serum hepatitis B virus DNA level. *JAMA*, 295, 65–73.

Chen, P., Kalpana, G., Goldberg, J., Mason, W., Werner, B., Gerin, J. and Taylor, J. (1986) Structure and replication of the genome of the hepatitis delta virus. *Proc. Natl. Acad. Sci. U.S.A.*, 83, 8774-8778.

Chen, Y.-S., Huang, W.-H., Hong, S.-Y., Tsay, Y.-G. and Chen, P.-J. (2008) ERK1/2-

mediated phosphorylation of small hepatitis delta antigen at serine 177 enhances hepatitis delta virus antigenomic RNA replication. *J. Virol.,* 82, 9345–9358.

Chou, H., Hsieh, T., Sheu, G. and Lai, M. (1998) Hepatitis delta antigen mediates the nuclear import of hepatitis delta virus RNA. *J. Virol.,* 72, 3684-3690.

Chudy, M. (2010) Towards standardization of HDV RNA measurement. Oral Presentation at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Cross, T., Rizzi, P., Horner, M., Jolly, A., Hussain, M. J., Smith, H. M., Vergani, D. and Harrison, P. M. (2008) The increasing prevalence of hepatitis delta virus (HDV) infection in South London. *J. Med. Virol.,* 80, 277–282.

Defenbaugh, D. A., Johnson, M., Chen, R., Zheng, Y. Y. and Casey, J. L. (2009) Hepatitis delta antigen requires a minimum length of the hepatitis delta virus unbranched rod RNA structure for binding. *J. Virol.,* 83, 4548–4556.

Dény, P. (2006) Hepatitis delta virus genetic variability: from genotypes I, II, III to eight major clades. *Curr. Top. Microbiol. Immunol.,* 307, 151–171.

Değertekin, H., Yalçin, K., Yakut, M. and Yurdaydin, C. (2008) Seropositivity for delta hepatitis in patients with chronic hepatitis B and liver cirrhosis in Turkey: a meta-analysis. *Liver Int.,* 28, 494–498.

Dimitrakakis, M. and Gust, I. (1991) HDV infection in the Western Pacific region. *Prog. Clin. Biol. Res.,* 364, 89–96.

Drake, J. W., Charlesworth, B., Charlesworth, D. and Crow, J. (1998) Rates of spontaneous mutation. *Genetics,* 148, 1667–1686.

Dunker, A., Lawson, J., Brown, C., Williams, R., Romero, P., Oh, J., Oldfield, C., Campen, A., Ratliff, C., Hipps, K., Ausio, J., Nissen, M., Reeves, R., Kang, C., Kissinger, C., Bailey, R., Griswold, M., Chiu, W., Garner, E. and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graph. Model.,* 19, 26–59.

Dyson, H. J. and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.,* 12, 54–60.

Elhaasi, A. F., Elzuki, Y., Arabi, B. and Tobgi, R. (2010) Prevalence of delta virus infection in patients with hepatitis B virus infection. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Engelke, M., Mills, K., Seitz, S., Simon, P., Gripon, P., Schnölzer, M. and Urban, S. (2006) Characterization of a hepatitis B and hepatitis delta virus receptor binding site. *Hepatology,* 43, 750–760.

Erhardt, A., Gerlich, W., Starke, C., Wend, U., Donner, A., Sagir, A., Heintges, T. and Häussinger, D. (2006) Treatment of chronic hepatitis delta with pegylated interferon-alpha2b. *Liver Int.,* 26, 805–810.

Farci, P. (2003) Delta hepatitis: an update. *J. Hepatol.,* 39, Suppl. 1, S212–S219.

Farci, P., Chessa, L., Balestrieri, C., Serra, G. and Lai, M. E. (2007) Treatment of chronic hepatitis D. *J. Viral Hepat.,* 14, Suppl 1, 58–63.

Farci, P., Karayiannis, P., Lai, M. E., Marongiu, F., Orgiana, G., Balestrieri, A. and Thomas, H. C. (1988). Acute and chronic hepatitis delta virus infection: direct or indirect effect on hepatitis B virus replication. *J. Med. Virol.,* 26, 279–288.

Farci, P. and Niro, G. (2012) Clinical features of hepatitis D. *Semin. Liver Dis.*, 32: 228-236.

Fattovich, G., Boscaro, S., Noventa, F., Pornaro, E., Stenico, D., Alberti, A., Ruol, A. and Realdi, G. (1987) Influence of hepatitis delta virus infection on progression to cirrhosis in chronic hepatitis type B. *J. Infect. Dis.,* 155, 931–935.

Fattovich, G., Giustina, G., Christensen, E., Pantalena, M., Zagni, I., Realdi, G. and Schalm, S. W. (2000) Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type B. The European Concerted Action on Viral Hepatitis (Eurohep). *Gut.,* 46, 420–426.

Ferns, R. B., Nastouli, E. and Garson, J. A. (2012) Quantitation of hepatitis delta virus using a single-step internally controlled real-time RT-qPCR and a full-length genomic RNA calibration standard. *J. Virol. Methods,* 179, 189–194.

Ferré-D'Amaré, A., Zhou, K. and Doudna, J. (1998) Crystal structure of a hepatitis delta virus ribozyme. *Nature,* 395, 567–574.

7. References

Filipovska, J. and Konarska, M. (2000) Specific HDV RNA-templated transcription by pol II in vitro. *RNA,* 6, 41–54.

Flores, R., Ruiz-Ruiz, S. and Serra, P (2012) Viroids and hepatitis delta virus. *Semin. Liver Dis.*, 32: 201-210.

Forterre, P. (2006) The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.*, 117, 5–16.

Fu, T. B. and Taylor, J. (1993) The RNAs of hepatitis delta virus are copied by RNA polymerase II in nuclear homogenates. *J. Virol.,* 67, 6965–6972.

Gaeta, G. B., Stroffolini, T., Smedile, A., Niro, G. and Mele, A. (2007) Hepatitis delta in Europe: vanishing or refreshing. *Hepatology,* 46, 1312–1313.

Garripoli, A., Di Marco, V., Cozzolongo, R., Costa, C., Smedile, A., Fabiano, A., Bonino, F., Rizzetto, M., Verme, G. and Craxi, A. (1994) Ribavirin treatment for chronic hepatitis D: a pilot study. *Liver,* 14, 154–157.

Glebe, D. and Urban, S. (2007) Viral and cellular determinants involved in hepadnaviral entry. *World J. Gastroenterol.*, 13, 22-38.

Glenn, J. S., Watson, J. A., Havel, C. M. and White, J. M. (1992) Identification of a prenylation site in delta virus large antigen. *Science,* 256, 1331–1333.

Goh, G., Dunker, A. and Uversky, V. (2009) Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virol. J.*, 6, 69.

Govindarajan, S., Chin, K. P., Redeker, A. G. and Peters, R. L. (1984) Fulminant B viral hepatitis: role of delta agent. *Gastroenterology,* 86, 1417–1420.

Greco-Stewart, V., Miron, P., Abrahem, A. and Pelchat, M. (2007) The human RNA polymerase II interacts with the terminal stem-loop regions of the hepatitis delta virus RNA genome. *Virology,* 357, 68-78.

Greco-Stewart, V., Schissel, E. and Pelchat, M. (2009) The hepatitis delta virus RNA genome interacts with the human RNA polymerases I and III. *Virology,* 386, 12–15.

Greco-Stewart, V. and Pelchat, M. (2010) Interaction of Host Cellular Proteins with Components of the Hepatitis Delta Virus. *Viruses,* 2, 1–24.

Gudima, S., Wu, S. Y., Chiang, C. M., Moraleda, G. and Taylor, J. (2000) Origin of hepatitis delta virus mRNA. *J. Virol.,* 74, 7204–7210.

Gudima, S., Chang, J., Moraleda, G., Azvolinsky, A. and Taylor, J. (2002) Parameters of human hepatitis delta virus genome replication: the quantity, quality, and intracellular distribution of viral proteins and RNA. *J. Virol.,* 76, 3709–3719.

He, B., Wang, K., Liu, Y., Xue, B., Uversky, V. N. and Dunker, A. K. (2009) Predicting intrinsic disorder in proteins: an overview. *Cell Res.,* 19, 929–949.

He, L. F., Ford, E., Purcell, R. H., London, W. T., Phillips, J. and Gerin, J. L. (1989) The size of the hepatitis delta agent. *J. Med. Virol.* 27, 31–33.

Hiscox, J. A. (2002) The nucleolus - a gateway to viral infection. *Arch. Virol.,* 147, 1077–1089.

Hong, S.-Y. and Chen, P.-J. (2010) Phosphorylation of serine 177 of the small hepatitis delta antigen regulates viral antigenomic RNA replication by interacting with the processive RNA polymerase II. *J. Virol.,* 84, 1430–1438.

Hsieh, S. Y. and Taylor, J. (1991) Regulation of polyadenylation of hepatitis delta virus antigenomic RNA. *J. Virol.,* 65, 6438–6446.

Huang, W., Yung, B., Syu, W. and Lee, Y. (2001) The nucleolar phosphoprotein B23 interacts with hepatitis delta antigens and modulates the hepatitis delta virus RNA replication. *J. Biol. Chem.,* 276, 25166-25175.

Huang, W., Chen, Y. and Chen, P. (2008a) Nucleolar targeting of hepatitis delta antigen abolishes its ability to initiate viral antigenomic RNA replication. *J. Virol.,* 82, 692-699.

Huang, W., Mai, R.-T. and Lee, Y.-H. W. (2008b) Transcription factor YY1 and its associated acetyltransferases CBP and p300 interact with hepatitis delta antigens and modulate hepatitis delta virus RNA replication. *J. Virol.,* 82, 7313–7324.

Hughes, S. A., Wedemeyer, H. and Harrison, P. M. (2011) Hepatitis delta virus. *Lancet,* 378, 73-85.

Hwang, S. B. and Lai, M. M. (1994) Isoprenylation masks a conformational epitope and enhances trans-dominant inhibitory function of the large hepatitis delta antigen. *J.*

Virol., 68, 2958–2964.

Ishida, T. and Kinoshita, K. (2007) PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.,* 35, W460–W464.

Ishida, T. and Kinoshita, K. (2008) Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics,* 24, 1344–1348.

Jayan, G. and Casey, J. (2002) Increased RNA editing and inhibition of hepatitis delta virus replication by high-level expression of ADAR1 and ADAR2. *J. Virol.,* 76, 3819-3827.

Johnson, D. E., Xue, B., Sickmeier, M. D., Meng, J., Cortese, M. S., Oldfield, C. J., Le Gall, T., Dunker, A. K. and Uversky, V. N. (2012) High-throughput characterization of intrinsic disorder in proteins from the Protein Structure Initiative. *J. Struct. Biol.,* 180, 201–215.

Jung, Y. J., Kim, W. and Kim, Y. J. (2010) Viral superinfection rate in Korea, HBV endemic area. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Kozhanova, T. V., Klushkina, V. V., Isaeva, O. V., Zamyatina, N. A., Gordeychuk, I. V., Oorzhak, N. D., Sargylar, A. A., Sonam-Baiyr, Y. D., Kyuregyan, K. K., Ilchenko, L. Y. and Mikhailov, M. I. (2010) Prevalence of hepatitis delta virus infection in Tyva Republic, Russian Federation. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Kuo, M., Goldberg, J., Coates, L., Mason, W., Gerin, J. and Taylor, J. (1988) Molecular cloning of hepatitis delta virus RNA from an infected woodchuck liver: sequence, structure, and applications. *J. Virol.,* 62, 1855–1861.

Lazinski, D. and Taylor, J. (1993) Relating structure to function in the hepatitis delta virus antigen. *J. Virol.,* 67, 2672–2680.

Lazinski, D. and Taylor, J. (1994) Expression of hepatitis delta virus RNA deletions: cis and trans requirements for self-cleavage, ligation, and RNA packaging. *J. Virol.,* 68, 2879-2888.

Le Gal, F., Gault, E., Ripault, M., Serpaggi, J., Trinchet, J., Gordien, E. and Dény, P.

(2006) Eighth major clade for hepatitis delta virus. *Emerg. Infect. Dis.*, 12, 1447-1450.

Lee, C., Lin, J., Chao, M., McKnight, K. and Lai, M. (1993) RNA-binding activity of hepatitis delta antigen involves two arginine-rich motifs and is required for hepatitis delta virus RNA replication. *J. Virol.*, 67, 2221-2227.

Lee, C. Z., Chen, P.-J., Lai, M. M. and Chen, D. S. (1994) Isoprenylation of large hepatitis delta antigen is necessary but not sufficient for hepatitis delta virus assembly. *Virology*, 199, 169–175.

Lee, C., Chang, S., Chen, C. and Chang, M. (1998) The nucleolin binding activity of hepatitis delta antigen is associated with nucleolus targeting. *J. Bio. Chem.*, 273, 7650-7656.

Lee, C. H., Chang, S. C., Wu, C. H. and Chang, M. F. (2001) A novel chromosome region maintenance 1-independent nuclear export signal of the large form of hepatitis delta antigen that is required for the viral assembly. *J. Biol. Chem.*, 276, 8142–8148.

Leistner, C. M., Gruen-Bernhard, S. and Glebe, D. (2008) Role of glycosaminoglycans for binding and infection of hepatitis B virus. *Cell. Microbiol.*, 10, 122–133.

Lemieux, R. U. and Spohr, U. (1994) How Emil Fischer was led to the lock and key concept for enzyme specificity. *Adv. Carbohydr. Chem. Biochem*., 50, 1-20.

Li, Y.-J., Stallcup, M. R. and Lai, M. M. C. (2004) Hepatitis delta virus antigen is methylated at arginine residues, and methylation regulates subcellular localization and RNA replication. *J. Virol.*, 78, 13325–13334.

Lieutaud, P., Canard, B. and Longhi, S. (2008) MeDor: a metaserver for predicting protein disorder. *BMC Genomics*, 9, Suppl 2, S25.

Lin, B., Defenbaugh, D. and Casey, J. (2010) Multimerization of hepatitis delta antigen is a critical determinant of RNA binding specificity. *J. Virol.*, 84, 1406-1413.

Lin, J.-H., Chang, M.-F., Baker, S., Govindarajan, S. and Lai, M. M. (1990) Characterization of hepatitis delta antigen: specific binding to hepatitis delta virus RNA. *J. Virol.*, 64, 4051-4058.

7. References

Linding, R., Jensen, L. J., Diella, F., Bork, P., Gibson, T. J. and Russell, R. B. (2003a) Protein disorder prediction: implications for structural proteomics. *Structure*, 11, 1453–1459.

Linding, R., Russell, R. B., Neduva, V. and Gibson, T. J. (2003b) GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.*, 31, 3701–3708.

Macnaughton, T. B., Shi, S. T., Modahl, L. E. and Lai, M. M. C. (2002) Rolling circle replication of hepatitis delta virus RNA is carried out by two different cellular RNA polymerases. *J. Virol.*, 76, 3920–3927.

Mansour, M., Malick, Z. F., Sidiya, A., Elkhali, I., Chekaraou, M. A., Lo, B., Ducancelle, A., Gordien, E. and Fabiani, F. L. (2010) High prevalence of hepatitis delta virus in adults and pregnant women in Mauritania. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Modahl, L., Macnaughton, T., Zhu, N., Johnson, D. and Lai, M. (2000) RNA-dependent replication and transcription of hepatitis delta virus RNA involve distinct cellular RNA polymerases. *Mol. Cell Biol.*, 20, 6030-6039.

Moraleda, G., Dingle, K., Biswas, P., Chang, J., Zuccola, H., Hogle, J. and Taylor, J. (2000) Interactions between hepatitis delta virus proteins. *J. Virol.*, 74, 5509–5515.

Mu, J.-J., Wu, H.-L., Chiang, B.-L., Chang, R.-P., Chen, D.-S. and Chen, P.-J. (1999) Characterization of the phosphorylated forms and the phosphorylated residues of hepatitis delta virus delta antigens. *J. Virol.*, 73, 10540–10545.

Mu, J.-J., Tsay, Y., Juan, L., Fu, T., Huang, W., Chen, D. and Chen, P.-J. (2004) The small delta antigen of hepatitis delta virus is an acetylated protein and acetylation of lysine 72 may influence its cellular localization and viral RNA synthesis. *Virology*, 319, 60–70.

Murphy, F. A. (1996) Virus taxonomy. *In*: Fields, B. N., Knipe, D. M., Howley, P. M., *et al*., eds. *Fields Virology, 3rd ed*. Chapter 2: 15-57.

Niro, G. A., Rosina, F. and Rizzetto, M. (2005) Treatment of hepatitis D. *J. Viral. Hepat.*, 12, 2–9.

Niro, G. A., Ciancio, A., Gaeta, G. B., Smedile, A., Marrone, A., Olivero, A., Stanzione, M., David, E., Brancaccio, G., Fontana, R., Perri, F., Adriulli, A. and Rizzetto, M. (2006) Pegylated interferon alpha-2b as monotherapy or in combination with ribavirin in chronic hepatitis delta. *Hepatology,* 44, 713–720.

Obradovic, Z., Peng, K., Vucetic, S., Radivojac, P., Brown, C. J. and Dunker, A. K. (2003) Predicting intrinsic disorder from amino acid sequence. *Proteins,* 53, Suppl 6, 566–572.

Oldfield, C., Cheng, Y., Cortese, M., Brown, C., Uversky, V. and Dunker, A. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry*, 44, 1989-2000.

Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N. and Dunker, A. K. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics,* 9, Suppl 1, S1.

Olivero, A., and Smedile, A. (2012) Hepatitis delta virus diagnosis. *Semin. Liver Dis.*, 32: 220-227.

Otto, J. and Casey, P. (1996) The hepatitis delta virus large antigen is farnesylated both in vitro and in animal cells. *J. Biol. Chem.*, 271, 4569-4572.

Ottobrelli, A., Marzano, A., Smedile, A., Recchia, S., Salizzoni, M., Cornu, C., Lamy, M. E., Otte, J. B., De Hemptinne, B. and Geubel, A. (1991) Patterns of hepatitis delta virus reinfection and disease in liver transplantation. *Gastroenterology,* 101, 1649–1655.

Pancsa, R. and Tompa, P. (2012). Structural disorder in eukaryotes. *PLoS ONE* 7, e34687.

Paraná, R., Kay, A., Molinet, F., Viana, S., Silva, L., Salcedo, J., Tavares-Neto, J., Lobato, C., Rios-Leite, M., Matteoni, L., Dóliveira, A., Tauil, P. and Trépo, C. (2006) HDV genotypes in the Western Brazilian Amazon region: a preliminary report. *Am. J. Trop. Med. Hyg.,* 75, 475-479.

Pawlotsky, J.-M. (2010) Diagnosis of HDV infection: need for standardization. Oral Presentation at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

7. References

Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. and Obradovic, Z. (2006) Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics*, 7, 208.

Polson, A. G., Bass, B. L. and Casey, J. L. (1996) RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature*, 380, 454–456.

Polson, A. G., Ley, H. L., Bass, B. L. and Casey, J. L. (1998) Hepatitis delta virus RNA editing is highly specific for the amber/W site and is suppressed by hepatitis delta antigen. *Mol. Cell Biol.*, 18, 1919–1926.

Ponzetto, A., Hoyer, B. H., Popper, H., Engle, R., Purcell, R. H. and Gerin, J. L. (1987) Titration of the infectivity of hepatitis D virus in chimpanzees. *J. Infect. Dis.*, 155, 72–78.

Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. and Sussman, J. L. (2005) FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, 21, 3435–3438.

Radjef, N., Gordien, E., Ivaniushina, V., Gault, E., Anais, P., Drugan, T., Trinchet, J., Roulot, D., Tamby, M., Milinkovitch, M. and Dény, P. (2004) Molecular phylogenetic analyses indicate a wide and ancient radiation of African hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades. *J. Virol.*, 78, 2537-2544.

Ramalho, F., Carvalho, G., Bonino, F., Baptista, A. and de Moura, M. C. (1987) Clinical and epidemiological significance of hepatitis delta virus (HDV) infection in chronic HBV carriers in Portugal. *Prog. Clin. Biol. Res.*, 234, 409–417.

Reanney, D. C. (1982) The evolution of RNA viruses. *Ann. Rev. Microbiol.*, 36, 47–73.

Reid, C. E. and Lazinski, D. W. (2000) A host-specific function is required for ligation of a wide variety of ribozyme-processed RNAs. *Proc. Natl. Acad. Sci. U.S.A.*, 97, 424–429.

Rizzetto, M., Canese, M. G., Aricò, S., Crivelli, O., Trepo, C., Bonino, F. and Verme, G. (1977) Immunofluorescence detection of new antigen-antibody system (delta/anti-delta) associated to hepatitis B virus in liver and in serum of HBsAg carriers. *Gut.*, 18, 997–1003.

Rizzetto, M., Shih, J. W., Gocke, D. J., Purcell, R. H., Verme, G. and Gerin, J. L. (1979) Incidence and significance of antibodies to delta antigen in hepatitis B virus infection. *Lancet*, 2, 986–990.

Rizzetto, M., Hoyer, B., Canese, M., Shih, J., Purcell, R. and Gerin, J. (1980) delta Agent: association of delta antigen with hepatitis B surface antigen and RNA in serum of delta-infected chimpanzees. *Proc. Natl. Acad. Sci. U.S.A.*, 77, 6124-6128.

Rizzetto, M., Rosina, F., Saracco, G., Bellando, P. C., Actis, G. C., Bonino, F., Smedile, A., Trinchero, P., Sansalvadore, F. and Pintus, C. (1986) Treatment of chronic delta hepatitis with alpha-2 recombinant interferon. *J. Hepatol.*, 3, Suppl 2, S229–S233.

Rizzetto, M., and Ciancio, A. (2012) Epidemiology of hepatitis D. *Semin. Liver Dis*., 32: 211-219.

Romeo, R., Del Ninno, E., Rumi, M., Russo, A., Sangiovanni, A., de Franchis, R., Ronchi, G. and Colombo, M. (2009) A 28-year study of the course of hepatitis delta infection: a risk factor for cirrhosis and hepatocellular carcinoma. *Gastroenterology*, 136, 1629–1638.

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J. E. and Dunker, A. K. (1997) Identifying disordered regions in proteins from amino acid sequence. *Proc. Int. Conference Neural Networks (ICNN'97)*, 1, 90-95.

Romero, P., Obradovic, Z., Kissinger, C., Villafranca, J., Garner, E., Guilliot, S. and Dunker, A. (1998) Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput*., 437-448.

Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J. and Dunker, A. K. (2001) Sequence complexity of disordered protein. *Proteins*, 42, 38–48.

Rozzelle, J., Wang, J., Wagner, D., Erickson, B. and Lemon, S. (1995) Self-association of a synthetic peptide from the N terminus of the hepatitis delta virus protein into an immunoreactive alpha-helical multimer. *Proc. Natl. Acad. Sci. U.S.A.*, 92, 382-386.

7. References

Sadiku, E. and Basho, J. (2010) Hepatitis delta virus in Albania. Poster presented at: *EASL Monothematic Conference: Delta Hepatitis*, 24-26 September, Istanbul, Turkey.

Salehi-Ashtiani, K. (2006) A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science*, 313, 1788–1792.

Saracco, G., Rosina, F., Brunetto, M. R., Amoroso, P., Caredda, F., Farci, P., Piantino, P., Bonino, F. and Rizzetto, M. (1987) Rapidly progressive HBsAg-positive hepatitis in Italy. The role of hepatitis delta virus infection. *J. Hepatol.*, 5, 274–281.

Scholtes, C., Icard, V., Amiri, M., Chevallier-Queyron, P., Trabaud, M. A., Ramiere, C., Zoulim, F., Andre, P. and Dény, P. (2012) Standardized one-step real-time reverse transcription-PCR assay for universal detection and quantification of hepatitis delta virus from clinical samples in the presence of a heterologous internal-control RNA. *J. Clin. Microbiol.*, 50, 2126–2128.

Schulze, A., Gripon, P. and Urban, S. (2007) Hepatitis B virus infection initiates with a large surface protein-dependent binding to heparan sulfate proteoglycans. *Hepatology*, 46, 1759–1768.

Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G. and Taylor, J. (1988) Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.*, 62, 2674–2679.

Sharmeen, L., Kuo, M. Y. and Taylor, J. (1989) Self-ligating RNA sequences on the antigenome of human hepatitis delta virus. *J. Virol.*, 63, 1428–1430.

Shattock, A. G. and Morris, M. C. (1991) Evaluation of commercial enzyme immunoassays for detection of hepatitis delta antigen and anti-hepatitis delta virus (HDV) and immunoglobulin M anti-HDV antibodies. *J. Clin. Microbiol.*, 29, 1873–1876.

Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V., Obradovic, Z. and Dunker, A. (2007) DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.*, 35, D786–D793.

Smedile, A., Dentico, P., Zanetti, A., Sagnelli, E., Nordenfelt, E., Actis, G. C. and Rizzetto, M. (1981) Infection with the delta agent in chronic HBsAg carriers. *Gastroenterology,* 81, 992–997.

Smedile, A., Farci, P., Verme, G., Caredda, F., Cargnel, A., Caporaso, N., Dentico, P., Trepo, C., Opolon, P., Gimson, A., Vergani, D., Williams, R. and Rizzetto, M. (1982) Influence of delta infection on severity of hepatitis B. *Lancet,* 2, 945–947.

Su, C.-W., Huang, Y.-H., Huo, T.-I., Shih, H. H., Sheen, I.-J., Chen, S.-W., Lee, P.-C., Lee, S.-D. and Wu, J.-C. (2006) Genotypes and viremia of hepatitis B and D viruses are associated with outcomes of chronic hepatitis D patients. *Gastroenterology,* 130, 1625–1635.

Tavanez, J., Cunha, C., Silva, M., David, E., Monjardino, J. and Carmo-Fonseca, M. (2002) Hepatitis delta virus ribonucleoproteins shuttle between the nucleus and the cytoplasm. *RNA,* 8, 637-646.

Taylor, J. (1990) Hepatitis delta virus: cis and trans functions required for replication. *Cell,* 61, 371–373.

Taylor, J. (2009) Chapter 3. Replication of the hepatitis delta virus RNA genome. *Adv. Virus Res.*, 74, 103–121.

Taylor, J. and Han, Z. (2010) Purinergic receptor functionality is necessary for infection of human hepatocytes by hepatitis delta virus and hepatitis B virus. *PLoS ONE,* 5, e15784.

Taylor, J. and Pelchat, M. (2010) Origin of hepatitis virus. *Future Microbiology,* 5, 393–402.

Taylor, J. (2012) Virology of hepatitis D virus. *Semin. Liver Dis*., 32: 195-200.

Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N. and Tawfik, D. S. (2008) Do viral proteins possess unique biophysical features. *Trends Biochem. Sci.*, 34, 53–59.

Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci*., 27, 527-533.

Tompa, P. (2011) Unstructural biology coming of age. *Curr. Opin. Struct. Biol.*, 21,

419–425.

Tsagris, E. M., Martínez De Alba, Á. E., Gozmanova, M. and Kalantidis, K. (2008) Viroids. *Cell. Microbiol.*, 10, 2168–2179.

Tsatsralt-Od, B., Takahashi, M., Nishizawa, T., Endo, K., Inoue, J. and Okamoto, H. (2005) High prevalence of dual or triple infection of hepatitis B, C, and delta viruses among patients with chronic liver disease in Mongolia. *J. Med. Virol.*, 77, 491–499.

Tseng, C.-H., Jeng, K.-S. and Lai, M. M. C. (2008) Transcription of subgenomic mRNA of hepatitis delta virus requires a modified hepatitis delta antigen that is distinct from antigenomic RNA synthesis. *J. Virol.*, 82, 9409–9416.

Tseng, C.-H., Cheng, T.-S., Shu, C.-Y., Jeng, K.-S. and Lai, M. M. C. (2010) Modification of small hepatitis delta virus antigen by SUMO protein. *J. Virol.*, 84, 918–927.

Uversky, V. N., Gillespie, J. and Fink, A. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins*, 41, 415-427.

Uversky, V. N. and Dunker, A. K. (2010) Understanding protein non-folding. *BBA - Proteins and Proteomics*, 1804, 1231–1264.

Uversky, V. N. (2011a) Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.*, 43, 1090–1103.

Uversky, V. N. (2011b) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem. Soc. Rev.*, 40, 1623–1634.

Uversky, V. N. and Dunker, A. K. (2012) Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal. Chem.*, 84, 2096–2104.

Wang, C.-C., Chang, T.-C., Lin, C.-W., Tsui, H.-L., Chu, P. B. C., Chen, B.-S., Huang, Z.-S. and Wu, H.-N. (2003) Nucleic acid binding properties of the nucleic acid chaperone domain of hepatitis delta antigen. *Nucleic Acids Res.*, 31, 6481–6492.

Wang, J. and Lemon, S. (1993) Hepatitis delta virus antigen forms dimers and

multimeric complexes in vivo. *J. Virol.*, 67, 446-454.

Wang, K. S., Choo, Q. L., Weiner, A. J., Ou, J. H., Najarian, R. C., Thayer, R. M., Mullenbach, G. T., Denniston, K. J., Gerin, J. L. and Houghton, M. (1986) Structure, sequence and expression of the hepatitis delta viral genome. *Nature,* 323, 508–514.

Webb, C.-H., Riccitelli, N. J., Ruminski, D. J. and Lupták, A. (2009) Widespread occurrence of self-cleaving ribozymes. *Science,* 326, 953.

Wedemeyer, H., Heidrich, B. and Manns, M. P. (2007a) Hepatitis D virus infection--not a vanishing disease in Europe. *Hepatology,* 45, 1331–1332.

Wedemeyer, H., Yurdaydin, C., Dalekos, G., Erhardt, A., Cakaloglu, Y., Degertekin, H., Gurel, S., Zeuzem, S., Bock, T., Zachou, K., Bozkaya, H., Drebber, U., Meyer, S., Dienes, H. and Manns, M. (2007b) [4] 72 week data of the HIDIT-1 trial: a multicenter randomized study comparing peginterferon α-2a plus adefovir VS. peginterferon α-2a plus placebo VS. adefovir in chronic delta hepatitis. *J. Hepatol.*, 46, Supp 1, S4.

Wedemeyer, H. and Manns, M. P. (2010) Epidemiology, pathogenesis and management of hepatitis D: update and challenges ahead. *Nat. Rev. Gastroenterology Hepatol.*, 7, 31–40.

Wedemeyer, H., Yurdaydin, C., Dalekos, G. N., Erhardt, A., Çakaloğlu, Y., Değertekin, H., Gürel, S., Zeuzem, S., Zachou, K., Bozkaya, H., Koch, A., Bock, T., Dienes, H. and Manns, M. (2011) Peginterferon plus adefovir versus either drug alone for hepatitis delta. *N. Engl. J. Med.,* 364, 322–331.

Williams, R. J. (1978) The conformational mobility of proteins and its functional significance. *Biochem. Soc. Trans.,* 6, 1123–1126.

Williams, V., Brichler, S., Radjef, N., Lebon, P., Goffard, A., Hober, D., Fagard, R., Kremsdorf, D., Dény, P. and Gordien, E. (2009) Hepatitis delta virus proteins repress hepatitis B virus enhancers and activate the alpha/beta interferon-inducible MxA gene. *J. Gen. Virol.* 90, 2759–2767.

Wong, S. and Lazinski, D. (2002) Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc. Natl. Acad. Sci. U.S.A.,* 99, 15118-

15123.

Wright, P. and Dyson, H. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.*, 293, 321-331.

Wu, J. C., Chiang, T. Y. and Sheen, I. J. (1998) Characterization and phylogenetic analysis of a novel hepatitis D virus strain discovered by restriction fragment length polymorphism analysis. *J. Gen. Virol.*, 79, 1105–1113.

Xia, Y. and Lai, M. (1992) Oligomerization of hepatitis delta antigen is required for both the trans-activating and trans-dominant inhibitory activities of the delta antigen. *J. Virol.*, 66, 6641-6648.

Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. and Uversky, V. N. (2010a) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *BBA - Proteins and Proteomics*, 1804, 996–1010.

Xue, B., Williams, R. W., Oldfield, C. J., Goh, G. K.-M., Dunker, A. K. and Uversky, V. N. (2010b) Viral disorder or disordered viruses: do viral proteins possess unique features. *Protein Pept. Lett.*, 17, 932–951.

Xue, B., Dunker, A. K. and Uversky, V. N. (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.*, 30, 137–149.

Yamaguchi, Y., Filipovska, J., Yano, K., Furuya, A., Inukai, N., Narita, T., Wada, T., Sugimoto, S., Konarska, M. and Handa, H. (2001) Stimulation of RNA polymerase II elongation by hepatitis delta antigen. *Science*, 293, 124-127.

Yamaguchi, Y., Mura, T., Chanarat, S., Okamoto, S. and Handa, H. (2007) Hepatitis delta antigen binds to the clamp of RNA polymerase II and affects transcriptional fidelity. *Genes Cells*, 12, 863–875.

Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., Huang, Y., Qi, Y., Peng, B., Wang, H., Fu, L., Song, M., Chen, P., Gao, W., Ren, B., Sun, Y., Cai, T., Feng, X., Sui, J. and Li, W. (2012) Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *eLife*, 1, e00049.

Yeh, T. S., Lo, S. J., Chen, P.-J. and Lee, Y. H. (1996) Casein kinase II and protein

kinase C modulate hepatitis delta virus RNA replication but not empty viral particle assembly. *J. Virol.*, 70, 6190–6198.

Yurdaydin, C., Bozkaya, H., Gürel, S., Tillmann, H. L., Aslan, N., Okçu-Heper, A., Erden, E., Yalçin, K., Iliman, N., Uzunalimoglu, O., Manns, M. and Bozdaui, A. (2002) Famciclovir treatment of chronic delta hepatitis. *J. Hepatol.*, 37, 266–271.

Zachou, K., Yurdayin, C., Dienes, H., Dalekos, G., Erhardt, A., Cakaloglu, Y., Degertekin, H., Gurel, S., Zeuzem, S., Bozkaya, H., Bock, T., Schlaphoff, V., Manns, M. and Wedemeyer, H. (2006) Significance of HDV-RNA and HBsAg levels in delta hepatitis: first data of the Hep-Net/international HDV intervention trial. *J. Hepatol.*, 44, Supp2, S178.

Zuccola, H., Rozzelle, J., Lemon, S., Erickson, B. and Hogle, J. (1998) Structural basis of the oligomerization of hepatitis delta antigen. *Structure,* 6, 821–830.

Appendix

The following reviews were published during the course of my research presented here:

Carolina Alves and Celso Cunha (2012) Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions. *In:* Dr. Sameh Magdeldin ed., *Gel Electrophoresis - Advanced Techniques,* 205-228.

Carolina Alves and Celso Cunha (2012). Order and disorder in viral proteins: new insights into an old paradigm. *Future Virology*, 7, 1183-1191.

12

Electrophoretic Mobility Shift Assay: Analyzing Protein – Nucleic Acid Interactions

Carolina Alves and Celso Cunha
Center for Malaria and Tropical Diseases
Institute of Hygiene and Tropical Medicine, New University of Lisbon
Portugal

1. Introduction

Interactions between proteins and nucleic acids mediate a wide range of processes within a cell from its cycle to the maintenance of cellular metabolic and physiological balance. These specific interactions are crucial for control of DNA replication and DNA damage repair, regulation of transcription, RNA processing and maturation, nuclear transport, and translation.

The characterization of protein-nucleic acid interactions is essential not only for understanding the wide range of cellular processes they are involved in, but also the mechanisms underlying numerous diseases associated with the breakdown of regulatory systems. These include, but are far from being limited to, cell cycle disorders such as cancer and those caused by pathogenic agents that rely on or interfere with host cell machinery. More recently, it has been hypothesized that many neurological disorders such as Alzheimer's, Huntington's, Parkinson's, and polyglutamine tract expansion diseases are a consequence, at least in part, of aberrant protein-DNA interactions that may alter normal patterns of gene expression (Jiménez, 2010).

The electrophoretic mobility shift assay (EMSA), also known as gel retardation assay, is a regularly used system to detect protein-nucleic acid interactions. It was originally developed with the aim of quantifying interactions between DNA and proteins (Fried & Crothers, 1981; Garner & Revzin, 1981) and since then evolved to be suitable for different purposes including the detection and quantification of RNA-protein interactions. EMSA is most commonly used for qualitative assays including identification of nucleic acid-binding proteins and of the respective consensus DNA or RNA sequences. Under proper conditions, however, EMSA can also be used for quantitative purposes including the determination of binding affinities, kinetics, and stoichiometry.

EMSA is a commonly used method in the characterization of transcription factors, the most intensely studied DNA-binding proteins, and the largest group of proteins in humans, second only to metabolic enzymes. Their purification and identification is crucial in understanding gene regulatory mechanisms. Transcription factors are sequence specific DNA binding proteins that are usually assembled in complexes formed prior to transcription initiation. They bind discreet and specific DNA sequences in the promoter

region functioning either as an activator or repressor of expression of the targeted gene through protein-protein interactions (reviewed by Simicevic & Deplancke, 2010). Transcription factors play essential roles during development and differentiation. It is well established that disruption of normal function of tissue-specific transcription factors, as a result of mutations, is often associated with a number of diseases including most forms of cancer, neurological, hematological, and inflammatory diseases. Additionally, transcription factors are often found differentially expressed in different pathologies suggesting an at least indirect involvement on the onset or progression of diseases. One of the most prominent examples of the involvement of transcription factors in development and progression of diseases is perhaps the p53 protein. p53 is a transcription factor involved in the modulation of expression of several genes that regulate essential cellular processes such as cell proliferation, apoptosis, and DNA damage repair (reviewed by Puzio-Kuter, 2011). Mutations in p53 that cause loss of function were reported in about 50% of all cancers. It is believed that this loss of function makes cancer cells more prone to the accumulation of mutations in other genes thus facilitating and accelerating the formation of neoplasias (reviewed by Goh et al., 2011).

In our laboratory, research is mainly directed to the study of host-pathogen interactions during hepatitis delta virus (HDV) replication and infection. HDV is the smallest human pathogen so far identified and infects human hepatocytes already infected with the hepatitis B virus (HBV). Both viruses have the same envelope proteins that are coded by the HBV DNA genome. HDV is, thus, considered a satellite virus of HBV. The HDV genome consists of a single-stranded, circular, RNA molecule of about 1700 nucleotides. This genome contains only one open reading frame from which two forms of the same protein, the so-called delta antigen, are derived by an editing mechanism catalyzed by cellular adenosine deaminase I. Both forms, small and large delta antigen, were shown to play crucial roles during virus replication: the small delta antigen is necessary for virus RNA accumulation and the large delta antigen plays an important role during envelope assembly (reviewed by Rizzetto, 2009). However, neither protein seems to display any known enzymatic activity. Accordingly, HDV is highly dependent on the host cell machinery for virus replication. It has been shown through EMSA that the small delta antigen binds *in vitro* to RNA and DNA without any specificity, which is in agreement with one of the roles attributed to the protein as a chaperone (Alves et al., 2010). Making use of different experimental approaches it was possible to identify a number of cellular proteins that interact with HDV antigens or RNA (reviewed by Greco-Stewart & Pelchat, 2010). However, the precise role played by most host factors during the virus life cycle remains elusive. Furthermore, it is highly consensual among HDV researchers that many other cellular factors that interact with delta antigens or HDV RNA remain to be identified and it is crucial to find those that interact with HDV RNA for a better insight on its replication and as possible targets for new therapies.

In this chapter we will review the principles of EMSA and its advantages and limitations for the quantitative and qualitative analysis of protein-nucleic acid interactions. The key parameters influencing the quality of protein samples, binding to nucleic acids, complex migration in gels, and sensitivity of detection will be discussed. Finally, an overview of the principles, advantages and disadvantages of methods that are an alternative to gel retardation assays will be provided.

2. Advantages and limitations

Since its first publication, in 1981, several improvements and variant techniques of EMSA were reported. Originally described as a method to qualitatively detect protein-DNA interactions, gel retardation assays rapidly became one of the most popular methods to map interaction sequences and domains not only in DNA but in RNA-protein interactions as well. EMSA was also adapted in order to allow the determination of quantitative parameters including complex stoichiometry, binding kinetics and affinity.

Several features made EMSA one of the most popular methods among researchers that study protein-nucleic acid interactions. Probably, the main advantages of EMSA when compared to other methods, as we will further discuss in the next sections, may be considered as follows: (1) EMSA is a basic, easy to perform, and robust method able to accommodate a wide range of conditions; (2) EMSA is a sensitive method, using radioisotopes to label nucleic acids and autoradiography, it is possible to use very low concentrations (0.1nM or less) and small sample volumes (20 µL or less; Hellman & Fried, 2007). Even though, less sensitive, non-radioactive labels are often used as well. These labels can further be detected using fluorescence, chemiluminescence or immunohistochemical approaches. Although less sensitive then radioisotopes, the wide variety of labels that can be used makes EMSA a very versatile method; (3) EMSA can also be used with a wide range of nucleic acid sizes and structures as well as a wide range of proteins, from small oligonucleotides to heavy transcription complexes; (4) Under the right conditions a gel retardation assay can separate the distribution of proteins between several nucleic acids within a single sample (Fried & Daugherty, 1998) or distinguish between complexes with different protein stoichiometry and/or binding site distribution (Fried & Crothers, 1981); (5) Finally, but not less important, it is possible to use both crude protein extracts and purified recombinant proteins enabling the identification of new nucleic acid-interacting proteins or characterization of specific proteins and its targets.

Despite its sensitivity, versatility and usually easy to perform protocols, EMSA is often considered to bear a number of limitations. Dissociation can occur during electrophoresis since samples are not at equilibrium during the run, thus preventing detection. Additionally, complexes that are not stable in solution may be stable in the gel requiring very short runs so that the observed pattern relates to what happens in solution. EMSA does not provide a straightforward measure of the weights or entities of the proteins as mobility in gels is influenced by several other factors. Also, EMSA does not directly provide information on the nucleic acid sequence the proteins are bound to. However, this problem may usually be overcome using footprinting approaches as described further ahead. Kinetic studies using EMSA are limited since the time resolution for a regular EMSA protocol consists of the time required to mix the binding reaction and for the electrophoretic migration to occur before the mix enters the gel. Only processes that have relaxation times larger than the interval required for solution handling are suitable for kinetic studies.

3. How complexes migrate in gels

In this section, we will start with a simple account of the characteristics of the electrophoretic mobility of nucleic acids alone, and afterwards we will discuss how the formation of protein-nucleic acid complexes alters these characteristics.

In a non-denaturing agarose or polyacrylamide gel and conventional buffer conditions the nucleic acids, being negatively charged, will migrate towards the anode when electric current is applied. The gel will then act as a sieve selectively impeding the migration in proportion to the nucleic acid molecular weight, which is generally proportional to its charge. Therefore, and as the weight is approximately related to chain length, the length of nucleic acid is estimated by its migration. There is though another property that affects gel migration that is the topology of the nucleic acid (conformation, circularity) making the molecules seem longer or shorter than they really are. Secondary and tertiary structures can be removed using denaturing agents (for example, formaldehyde, formamide and urea) allowing for the electrophoretic mobility to become a simple function of molecular weight. Obviously, this denaturing step cannot be applied in a gel retardation assay as it would impede the interaction between the protein and nucleic acid.

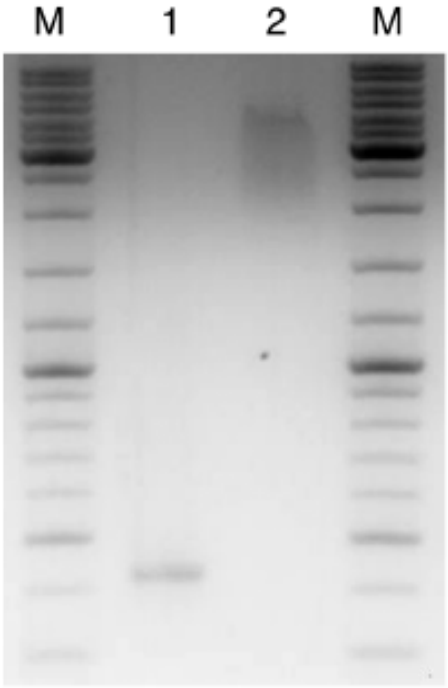


Fig. 1. Example of an electrophoretic mobility shift assay. An unlabeled DNA of 400 base pairs (bp) was incubated in a phosphate buffer (137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4, 1.5mM KH_2PO_4, pH 7.4) in the absence (1) or presence (2) of 2μM of small delta antigen. The samples were loaded onto a 1.5% agarose gel and after electrophoresis in TAE buffer (40mM Tris acetate, 1mM EDTA) the DNA was stained with ethidium bromide. (M) represents the molecular weight marker (GeneRuler DNA Ladder mix, Fermentas).

When a protein is added to the mix and interacts with the nucleic acid forming complexes it results in a change in gel migration relative to that of the free nucleic acid. This shift is mainly due to an obvious increase in the molecular weight, the adjustment of charge and eventual changes in the nucleic acid conformation. In figure 1 we give an example of an EMSA study where the small delta antigen was added to DNA. It is clear that the addition

of the small delta antigen (Fig.1. well 2) to a 400bp DNA fragment results in the formation of a complex with decreased gel mobility when compared with the unbound DNA (Fig.1. well 1). We can conclude that under our *in vitro* binding conditions, the small delta antigen interacts with the given 400bp DNA fragment causing a clear mobility shift.

It is expected that when protein binds a nucleic acid fragment there will be a decrease in relative mobility and if the protein doesn't induce any appreciable bend on the nucleic acid then the conformational contribution to the decrease is small. Although an increase in the protein molecular weight results in reduction of gel migration it has been reported that the increase of the nucleic acid length can have the opposite effect. This was reported for the Lac repressor bound to DNA fragments of increasing sizes, which resulted in an increase of relative mobility (Fried, 1989). This observation indicates that the ratio of protein and nucleic acid weights is more important in the migration than the absolute weight of the complex. Another interesting study reports that the binding of protein to a nucleic acid can accelerate mobility. This was observed for relatively large linear DNA binding to a protein from the hyperthermophilic *Methanothermus fervidus* that was shown to induce nucleic acid condensation (Sandman et al., 1990). In this case the conformational change of the DNA is a stronger factor than the weight increase, causing acceleration rather than a decrease in relative mobility.

Overall, the conformational features that influence gel migration of protein-nucleic acid complexes are not thoroughly studied and questions are only raised when exceptions emerge such as the ones mentioned above. Nowadays, the EMSA method is almost exclusively used to analyze the interaction between proteins and nucleic acids and to a lesser extent its conformations that can influence gel migration. When exceptions arise and the retardation pattern is not exactly as predicted, it can still point out clearly whether the molecules are interacting or not. In the end, the exact location of the resulting gel bands cannot be predicted but the answer is usually unambiguous.

External factors can also influence the separation of the bound or unbound nucleic acid such as the nature of the gel matrix and temperature during electrophoresis. Generally, the best resolution is obtained with the smallest pore diameter that allows the migration of unbound nucleic acid. However, if large complexes are expected there should be a compromise in pore size so that they can enter the gel matrix. As will be discussed below, polyacrylamide gels offer the best conditions for small complexes and nucleic acid fragments. On the other hand, agarose gels are more suitable for larger aggregates.

The detection of a protein-nucleic acid complex within a gel depends critically on the resolution obtained between unbound nucleic acid and the formed complexes as well as its stability within the gel matrix. In most cases, the gel matrix is expected to stabilize the preformed complex as it impedes the diffusion of dissociating components maintaining the concentration of protein and nucleic acid (and complex) at levels as high or higher than those achieved in the equilibrium binding reaction. This of course is compromised if for instance the salt concentration in the binding reaction differs largely from that in the electrophoresis/gel buffer, resulting in an adjustment in salt concentration that could disrupt the complexes formed. As the gel retardation method is an *in vitro* assay, when extrapolating to the *in vivo* conditions one must be careful as the former may provide favorable binding conditions that are not achieved at physiological concentrations.

4. The method

There are five focal steps in a conventional EMSA protocol that involve different variables susceptible to optimization: (1) preparation of protein sample; (2) synthesis and labeling of nucleic acid; (3) binding reaction; (4) non-denaturing gel electrophoresis and (5) detection of the outcome. In this segment we will discuss each step separately mentioning the key variables in each one and the options available for any given situation. Figure 2 represents schematically the regular steps in a gel retardation assay that will be discussed below. Whenever possible we will also refer to examples in the literature.

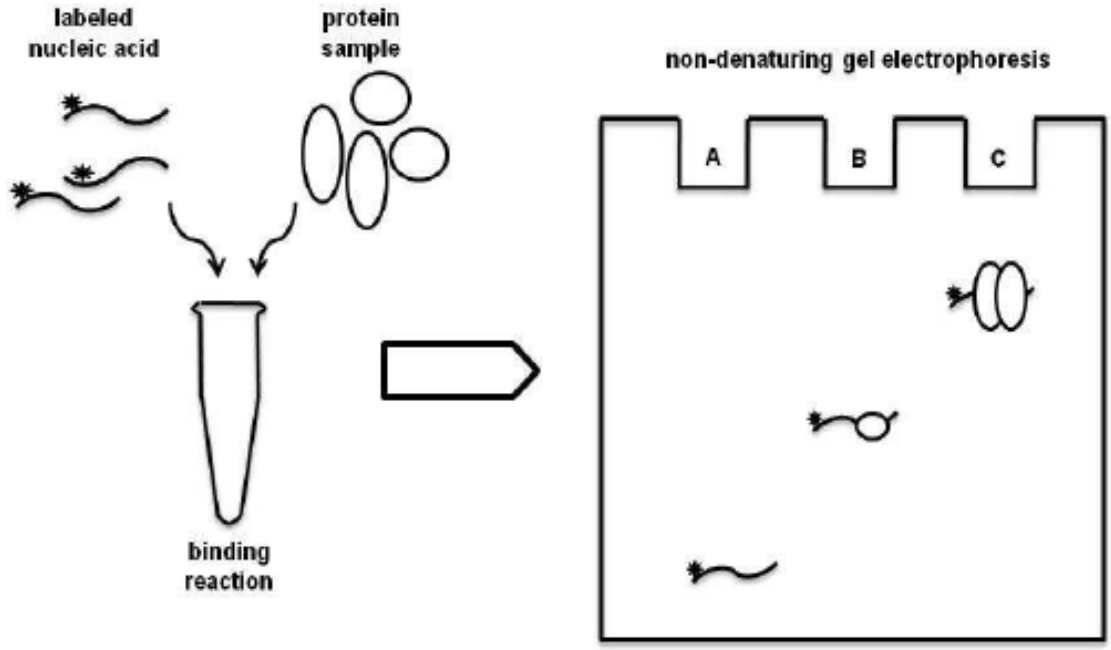


Fig. 2. A schematic representation of a conventional EMSA protocol. The labeled nucleic acid, simplified as lines with a star representing the label, is mixed with the protein sample, represented by the oval shapes, in a binding reaction and then loaded into a non-denaturing gel. After electrophoresis the result is detected according to the label in the nucleic acid. On the schematic gel (A) represents a well on which only the labeled nucleic acid was loaded. The free nucleic acid is expected to have more mobility than the bound molecules. In well (B) is symbolized a labeled nucleic acid binding to one small peptide and in well (C) is binding to two larger proteins. The heavier complex (in C) is expected to display the lowest mobility during electrophoresis and therefore is closer to the beginning of the gel.

4.1 Preparation of the protein sample

Regarding the protein sample, the EMSA can be divided into two categories based on whether the nucleic acid-interacting protein is known or not. Therefore, preparing the protein sample will depend on which category it falls, in order to obtain an optimal performance.

When faced with a putative nucleic acid-binding protein or complex of completely unknown subcellular origin, whole cell extracts must be used. If there is an educated guess

on the nature of the protein, it is advisable to isolate nuclear and cytoplasmic proteins from crude extracts improving the results. Particularly, if the binding protein is thought to be nuclear and in low abundance, the isolation of nuclear extracts will prevent the dilution that would occur if whole cell extracts were used, which could render the concentration too low for the protein to be even detected.

Cell extracts are easy and relatively fast to obtain and the methods are commonly derived from the protocol described by Dignam and collaborators almost three decades ago (Dignam et al., 1983). This method isolates both nuclear and/or cytoplasmic proteins suitable for later analysis using EMSA. One disadvantage in preparing cell extracts is its crudeness; they generally degrade faster than purer preparations due to the presence of cellular proteases. To limit protein degradation or alteration the protocol should be performed on ice or at 4ºC and protease inhibitors should be added. A control test can easily be performed to assess the viability of the extract by using ubiquitous DNA probes (Kerr, 1995). If these fail than the cell extract might be "dead". Despite its disadvantages cell extracts are needed when the interest lies in identifying new nucleic acid-binding proteins or when a complex of different proteins is needed to interact with the target nucleic acid as sometimes one recombinant protein cannot bind by itself. Tissue samples can also be a source of protein sample for these assays. The same care should be taken as in whole cell extracts to minimize the activity of proteases.

If the nucleic acid-binding protein is known then recombinant proteins can be expressed and purified. Recombinant or heterologous proteins are commonly expressed in bacteria or an eukaryotic cell line of interest. Fusion proteins of the target are generally constructed with a tag to facilitate purification. Common tags, such as glutathione-S-transferase (GST), tandem affinity purification tag (TAP tag), maltose binding protein (MBP) or 6xHistidine, are cloned in frame with the protein. Sometimes it is possible to include a protease cleavage site between the protein of interest and the tag so the latter can be easily removed after purification. Even though a tag can be very helpful, it should be taken into account that it can alter the recombinant protein conformation and even disrupt its binding ability. On the other hand they can be helpful in stabilizing the protein terminus they are close to. A careful study is needed when choosing the tag and usually small peptides are preferred to minimize its impact on the recombinant protein of interest.

There are several systems available for the production of heterologous proteins of which bacterial extracts of *Escherichia coli* are one of the most widely used. This Gram-negative bacterium remains an attractive host due to its ability to grow rapidly and with high density using inexpensive substrates. Its genetics has been well characterized for quite some time and there is a wide range of cloning vectors as well as mutant host strains that make it such a versatile system. Typically, the heterologous complementary DNA is cloned into a compatible plasmid which is then transfected into the bacteria to achieve a high gene dosage. This doesn't necessarily guarantee the accumulation of high levels of a full-length active form of the recombinant protein but other efforts can be made to improve that. To achieve high-level production in *E. coli* strong promoters should be used such as the bacteriophage T7 late promoter, and usually the T7 polymerase is also present under IPTG (isopropyl-β-D-1-thiogalactopyranoside)-induction. In the past years several strains have been engineered to improve the recombinant protein yields through efforts to increase mRNA stability as well as improve transcription

termination and translational efficiency (reviewed by Baneyx, 1999 and Makino et al., 2011). However, this extensively used system for protein overexpression has an important drawback when studying eukaryotic proteins. The bacterial systems are not able to perform post-translational modifications that would eventually happen *in vivo* in eukaryotic cells.

When working with recombinant nucleic acid-binding proteins it should be taken into account the importance of post-translational modifications on the protein's binding ability. A careful research of previous reports might hint if it is necessary to perform modifications prior to the binding reaction. In some cases post-translational modifications change the sequence-specificity of the binding. For example, genotoxic stress induces modifications on the C-terminus of the tumor suppressor protein p53 that modulate its DNA-binding specificity (Apella & Anderson, 2001). If the modifications are crucial, rather than using bacterial extracts a more biologically relevant host should be considered. Transient gene expression in mammalian cells has become a routine approach to express proteins in cell lines such as human embryonic kidney cells. The benefits are obvious for the production of eukaryotic proteins in mammalian cells as post-translational modifications will likely be native or near-native, solubility and correct folding are more likely to occur as well as expression of proteins in their proper intracellular compartments. These methods, however, tend to be more expensive as cells need a more complex growth media and there is a lower diversity in cloning vectors. To get out of the latter limitation an alternative approach uses baculovirus-infected insect cells. In this method a recombinant virus is produced either by site-specific transposition of an expression cassette into the shuttle vector or through homologous recombination (reviewed by Jarvis, 2009).

When expressing recombinant proteins, sometimes, the heterologous genes interfere severely with the survival of the host cell. For toxic proteins produced in *E. coli* strains there are some techniques available to get around this problem. A highly toxic gene can be defined as a gene that, when introduced into a cell, causes cell death or severe growth and maintenance defects even prior to expression induction. The best solution for expressing a highly toxic gene is to enable the host to tolerate it during the growth phase, so that after induction an efficient expression ensures a rapid and quantitative production of the toxic protein before the cell dies (reviewed by Saida et al., 2006). This can be achieved by different strategies such as manipulation of the gene's transcriptional and translational control elements, for example, by suppressing basal expression of the toxic protein from leaky inducible promoters. Managing the coding sequence to produce reversible inactive forms or controlling the plasmid copy number is also an option as well as selecting less susceptible *E. coli* strains or adding stabilizing sequences.

Cell-free systems are also available to express recombinant proteins including *in vitro* transcription\translation systems such as rabbit reticulocyte systems, wheat germ based systems or *E. coli* cell-free protein expression systems (reviewed by Endo & Sawasaki, 2006). Here, proteins can be expressed directly from cDNA templates obtained through PCR, avoiding subcloning which makes it a faster method by skipping this step, and eventually cheaper. It can also be used to express proteins that seriously interfere with the cell physiology such as the toxic proteins mentioned above. On the other hand these methods usually achieve smaller yields than for instance bacterial extracts approaches.

4.2 Synthesis and labeling of nucleic acids

One of the key advantages of EMSA is its versatility as it can be performed using a wide range of nucleic acid structures and sizes. This method can characterize both double- and single-stranded DNA as well as RNA, triplex and quadruplex nucleic acids or even circular fragments. The probe design and synthesis depends on the application or purpose of the study and is a significant aspect, as it will influence the detection and therefore the sensitivity of the results. There are two main aspects to consider in this step: the length of the nucleic acid and its labeling.

Unlabeled nucleic acids can be used in a gel retardation assay and be detected by post-electrophoretic staining with chromophores or fluorophores that bind nucleic acids or in the "classical way" using ethidium bromide. However the use of labeled nucleic acids is usually preferred as it can facilitate detection and add sensitivity to the method. The most common choice is radioisotope labeling as it offers the best sensitivity without interfering with the structure of the probe. A higher sensitivity makes it ideal for assays that have a limited amount of starting material. The radioisotope, usually ^{32}P, can be incorporated in the nucleic acid during its synthesis, by the use of labeled nucleotides, or afterwards via end labeling using a kinase or a terminal transferase. With a radioactive label the EMSA results can be easily detected by autoradiography. Even if radioisotope labeling confers high sensitivity to the method it implies handling hazardous radioactive material requiring extra safety measures that may not be available. Other labels can be used as alternatives that, even though are less sensitive, are a lot safer to manipulate and more stable such as fluorophores, biotin or digoxigenin (Holden & Tacon, 2011). When these molecules are used detection is achieved by chemiluminescence or immunohistochemistry. Although, in general radioisotope labeling achieves higher sensitivity there are some reports that similar results can be obtained with other labels such as Cyano dye Cy5 (Ruscher et al., 2000).

Although the most common approach is the labeling of the nucleic acid probe there are protocols available that employ protein labeling at the same time. For example, Adachi and co-workers suggest the use of an iodoacetamide derivative labeling of the thiol residue of cysteins (Adachi et al., 2005). Using radioisotope labeled DNA mixed with a nuclear protein extract they perform a conventional EMSA and after detection by autoradiography the complexes are eluted from excised gel bands and treated with 5-iodoacetamidofluorescein for protein labeling. The sample is then loaded onto a denaturing gel and after electrophoresis is transferred to a membrane and detected with anti-fluorescein antibody. This allows the characterization of the proteins in the complex giving information on how many proteins are present and their molecular weight. However it is not able to detect proteins without cystein residues.

Regarding the length of the nucleic acid probe, it depends on what is being studied. If one is looking for specific binding sites, small probes can be used to assess with each segment the protein will interact. The use of short nucleic acids has several advantages as they are easily synthesized and inexpensive to purchase; a small sequence has less non-specific binding sites (it should be particularly advantageous when a protein has low sequence-specificity); the electrophoretic resolution between complexes and free nucleic acid is higher so shorter electrophoresis times can be used. Nevertheless, in a short sequence the binding sites are closer to the molecular ends which can cause aberrant binding and it can be tricky to resolve the free nucleic acid from the complexes formed if these have a very high molecular weight.

On the other hand, the longer nucleic acid targets avoid these problems but will have more non-specific binding sites and the mobility shift is generally smaller requiring longer electrophoresis times as they run more slowly. A compromise needs to be reached depending on what the EMSA study is trying to achieve.

4.3 Binding reaction

The interaction between proteins and nucleic acid is sensitive to salt concentration and pH as it will influence the protein charge and conformation. However, the experimental conditions are very versatile in that different buffers can achieve good results. The most commonly used are Tris based buffers but other options include 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 3-(N-morpholino)-propanesulfonic acid (MOPS), and glycine or phosphate buffers. Naturally, it is advisable to provide an environment as close as possible to physiological conditions so the data obtained *in vitro* can be related to what happens *in vivo*.

Additives can be included in the binding reaction either if the interactions require the presence of co-factors or stabilizing agents, or as helpful components to minimize non-specific binding. Glycerol or other small neutral solutes, for example sucrose, can be added to the binding mixture to stabilize labile proteins or enhance the stability of the interaction (Vossen et al., 1997). These solutes are used at final concentrations of 2M or less, as higher concentrations might interfere with the sample's viscosity and complicate handling. Other assays may require the presence of co-factors for a correct interaction such as the presence of cAMP for the *E. coli* CAP protein (Fried & Crothers, 1984) or ATP for human recombinase Rad51 (Chi et al., 2006). Non-ionic detergents are used to maximize protein solubility. In this case, the concentrations used depend on the detergent and system under study. Nuclease and phosphatase inhibitors can be useful as well as protease inhibitors, which as mentioned before, are particularly important when the protein sample comes from cell extracts. These inhibitors are commercially available and the concentration depends on the manufacturer's instructions. Some of the additives mentioned, particularly those involved in stabilizing the formed complexes can be included not only in the binding mixture but also in the gel buffers.

To minimize non-specific loss of protein the addition of a carrier protein (less than 0.1mg/mL) such as bovine serum albumin can be very helpful. The addition of unlabeled competing nucleic acids is suitable when there are secondary binding activities that mask the relevant one. Of course this only works if the protein interacts with the target nucleic acid with greater affinity then its competitor and the secondary binding does not discriminate between the sequences. Since the presence of a competing nucleic acid will always reduce the amount of specific binding, testing different competitors and concentrations is needed to optimize the assay. Another option to circumvent the problem of non-specific binding is the addition of salt at concentrations that will disrupt non-specific ionic bonds but leave the more specific interactions unimpaired.

4.4 Non-denaturing gel electrophoresis

After the binding reaction the free nucleic acid is separated from the formed complexes by non-denaturing gel electrophoresis. EMSA can be performed on polyacrylamide or agarose

gels depending mainly on the size of the nucleic acid and desired resolution. The average pore size is estimated to be around 5 to 20nm in diameter for 10 and 4% acrylamide gels respectively (Lane et al., 1992). Typically the higher concentration gels are used for oligonucleotides and small RNAs and the lowest concentration for DNA fragments of around 100bp. A polyacrylamide gradient gel is sometimes preferred over linear gels as the gradient in pore size increases the range of molecular weight fractioned in a single run, which is particularly important when the complex has a much higher weight than the free nucleic acid (Walker, 1994). When complexes of different composition are formed, the gradient gels are also more likely to separate those with close molecular weight.

Agarose gels, on the other hand, have a pore size of around 70 to 700nm (Lane et al., 1992) in diameter and are therefore mostly used in assays with larger nucleic acid fragments or when large protein complexes are expected. Overall, polyacrylamide gels offer a better resolution for nucleic acid-protein complexes with a molecular weight of up to 500,000Da (Fried, 1989 as cited in Hellman & Fried, 2007).

Regarding the electrophoresis buffers, it should be taken into account the fact that the interaction between nucleic acids and proteins involves an ionic component. Therefore, the buffer's ionic strength and pH are important features that play a role in the complex stability. Although this is a very important factor there hasn't been, to our knowledge, any thorough study on the subject. The choice of electrophoresis buffers is varied and generally low ionic strength buffers are preferred and sometimes coincide with the buffer used in the binding reaction. Buffers with a medium salt concentration help stabilize the complexes, generate less heat during electrophoresis and also increase the speed of migration. High salt concentrations not only disrupt the complexes but also interfere with its movement into the gel matrix and lead to significant heating during the electrophoresis. Too low salt concentrations can also disrupt the stability of the preformed complexes as well as separate a double stranded DNA template (Kerr, 1995). The most common buffers are TBE (90mM Tris-Borate, 2mM EDTA, pH 8) and TAE (40mM Tris-Acetate, 1mM EDTA, pH 8). However, there are some complexes that cannot be detected with the classical buffers. For example the complexes formed between phage Mu repressor and its operators have an electrophoresis buffer-dependent stability and require Tris-glycine buffer at pH 9.4. (Alazard et al., 1992 as cited in Lane et al., 1992).

Particularly, in agarose gels it is important to monitor the temperature during electrophoresis to prevent the gel from heating up which could result in dissociation of the nucleic acid-protein complexes. Some cases may require that pre-cooling of the gel or even that the electrophoresis proceeds at lower than room temperatures, which can be achieved with special refrigeration devices.

4.5 Detection

The detection of an EMSA result will naturally depend on the labels used if any has been used. The results uncovered can involve the detection of the mobility shift between free nucleic acid and the complexed form or the detection of the mobility shift of free protein and the complexes.

Looking at the nucleic acid component without any label added the shift in mobility can be detected by staining with molecules that bind nucleic acids. Different products can be used

ranging from the classic but hazardous ethidium bromide to other chromophores or fluorophores such as RedSafe DNA Stain (ChemBio) or SYBR® Safe DNA gel stain (Invitrogen). When the nucleic acid has been previously labeled the detection methods depend on the nature of the label. A ^{32}P radioisotope is one of the easiest and most sensitive methods to detect nucleic acids but it's a hazardous material to work with. Other very common labels are biotin, digoxigenin or fluorophores. These labels are innocuous but usually give less sensitive results and the detection procedure can involve extra steps such as transfer to a membrane and incubation with primary and secondary antibodies as well as intermediate washing steps. The results in these cases can be observed by immunohistochemistry or chemiluminescence approaches.

The detection of protein mobility shift involves less direct methods, meaning, extra steps such as a denaturing step and electrotransfer onto a membrane, may be necessary as they are usually immunodetected. If the protein of interest is known, and a specific antibody is available, it can be used in detection. If not, a method such as the one discussed above, proposed by Adachi and colleagues that involves labeling the thiol group of cysteins and using an antibody against the label. Stepwise, the easier way to detect protein in an EMSA is by labeling it with radioisotope, a method designated by reverse EMSA that will be discussed ahead. This procedure has the disadvantage of working with radioactive material but the mobility shift can be visualized by autoradiography.

5. EMSA applications

The gel retardation assay has been used under different conditions in order to achieve specific results. The method is useful in studying not only the interaction between proteins and nucleic acids but also in assessing nucleic acid conformational characteristics. It can be used to characterize bends in the DNA double helix with polyacrylamide gels and comparative measurements (for an example Crothers & Drak, 1992) or to detect complexes formed with super coiled DNA being sometimes designated as topoisomer gel retardation (for examples see Palecek, 1997; Nordheim & Meese, 1988). In this section we mention how a gel retardation assay can help characterize protein-nucleic acid interactions.

5.1 Binding constants

Although EMSA is most commonly used as a qualitative assay it can, under certain conditions, provide quantitative data for relatively stable complexes. One of its earliest applications was in the measurement of kinetic and thermodynamic parameters. The association rates are determined by mixing the complex components at known concentrations and loading them in a running gel at precise intervals (for an example Spinner et al., 2002). For dissociation rates, a time course experiment is done by addition of competing nucleic acid to the preformed complexes (Fried & Crothers, 1981). The binding constant can be determined by the amount of complex formed as a function of protein concentration at equilibrium or as a ratio of the association and dissociation constants (for an example Demarse et al., 2009). An alternative method to measure kinetic and thermodynamic constants is the nitrocellulose filter binding assay that will be mentioned below.

As an example we show in figure 3 the titration of a DNA with the small delta protein to assess binding constants. The binding reaction was done by incubating the samples in a

phosphate buffer during the same period of time (10 minutes) and then loading them onto an agarose gel for electrophoresis. It is clear that when the protein is present at only 0.25μM it does not interfere with the DNA mobility (Fig.3. well 2) as the band covered the same distance as the first sample, in which the protein was not present (Fig.3. well 1). But when 1.5μM of the small delta antigen are present in the binding reaction there is almost no free DNA present and the majority of the molecules are bound in a complex (Fig3. well 5). In the intermediate concentrations it can be clearly observed the decreasing presence of free DNA and increasing DNA-protein complexes as the protein concentration raises. We can consider that the dissociation constant can be estimated by quantifying the disappearance of the free DNA band (Demarse et al., 2009). From figure 3 we can say that the apparent dissociation constant is between 1 and 1.5μM.

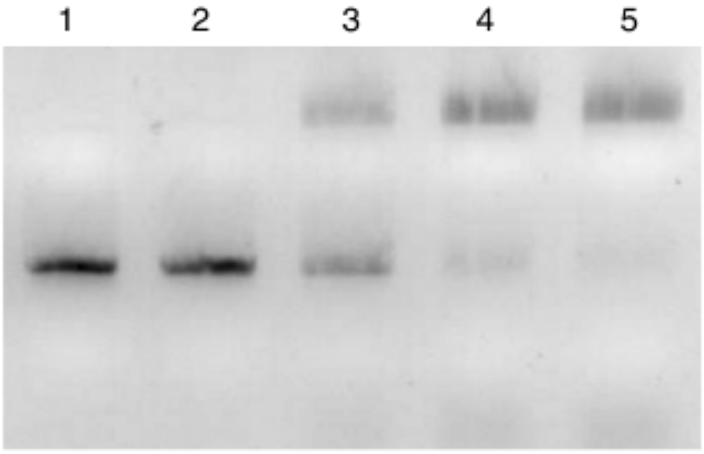


Fig. 3. Titration of a 500bp DNA fragment with the small delta antigen to estimate binding constants. An unlabeled 500bp DNA complementary to part of the HDV RNA was incubated, in a phosphate buffer (137mM NaCl, 2.7mM KCl, 4.3mM Na_2HPO_4, 1.5mM KH_2PO_4, pH 7.4), with increasing concentrations of small delta antigen of 0; 0.25; 0.5; 1; and 1.5μM and samples were loaded onto wells 1, 2, 3,4 and 5, respectively. Electrophoresis was in a 1.5% agarose gel in TAE buffer and the DNA was stained with ethidium bromide.

5.2 Cooperativity

Proteins can bind nucleic acids in a cooperative manner, that is, the complexes formed involve the binding of more than one protein to a specific nucleic acid segment. These multiprotein complexes may be a consequence of direct protein-protein interaction needed for nucleic acid binding, or a protein-induced deformation of the nucleic acid is a prerequisite to facilitate the binding of a second protein, or it may result from the bringing together of molecules bound at distinct sites in the nucleic acid sequence. The cooperativity can be inferred in a gel retardation assay from the underrepresentation of intermediate complexes between the unbound and saturated states. Multiprotein complexes can be comprised of a single protein species forming a homomultimer or of different proteins. The latter can be easily characterized by EMSA by the stability of the complexes formed with one protein in the presence or absence of the other(s).

5.3 Stoichiometry

Determining the important parameter that is stoichiometry is not as easy a task as it seems. The apparent weight changes estimated from the complexes' gel mobility are not applicable in determining the stoichiometry due to complications of charges and conformational effects on gel migration. A different approach is needed. The presence of truncated or extended protein derived from the wild-type but with the same binding and multimerization capacity will originate new bands that can reflect the monomers bound to the nucleic acid (Hope & Struhl, 1987). A similar method that will be discussed in the next segment is the supershift EMSA that uses an antibody specific for the binding protein recognizing an epitope that is accessible while the protein is bound to the nucleic acid. The addition of the antibody to the preformed complex can provide an estimate of the number of proteins bound by the extent of increments in retardation (Michael N & Roizman B, 1991 as cited in Lane & Prentki, 1992).

A more complex approach has been proposed in 1988 to determine a complex's stoichiometry (Granger-Schnarr et al., 1988). After the separation of the free and the complexed nucleic acid on a non-denaturing gel, the proteins are transferred to a membrane after sodium dodecyl sulfate (SDS) denaturation. This then allows the detection of proteins directly or indirectly using a specific antibody. The protein bands as well as the nucleic acids autoradiograph are then quantified by densiometry and the relative stoichiometry can be determined. The need for a specific antibody limits this method to complexes formed by well known proteins with available antibodies.

6. EMSA variants

Over the years variations or coupling of the EMSA protocol with other methods has been proposed to enhance its results or obtain more information from one experiment. Some examples of these EMSA-based approaches will be presented.

6.1 Reverse EMSA (rEMSA)

A reverse EMSA consists in labeling the protein sample rather than the nucleic acid (Filion et al., 2006). This method shows the difference in mobility between the free protein and nucleic acid-bound protein. It is an approach that can facilitate the determination of the protein binding affinity using different nucleic acids. Because the label used is ^{35}S instead of ^{32}P it is less sensitive than the conventional EMSA due to the isotope's energy.

6.2 Supershift EMSA

The supershift EMSA uses the same protocol as a regular EMSA except in that an antibody against the binding protein is added. As a result there is a more marked mobility shift during electrophoresis because the antibody will increase the overall complex molecular weight, hence the term supershift. This method can help identify if the proteins present in the complex have a specific epitope and is also used to validate previously identified proteins. It can also improve resolution when the difference between free nucleic acid and the complex is very small.

6.3 Multiplexed competitor EMSA (MC-EMSA)

The multiplexed EMSA was developed in 2008 by Smith and Humphries to characterize nuclear protein and DNA interactions, namely with transcription factors. In this method the nuclear extract is incubated with a pool of unlabeled DNA consensus competitors prior to adding the labeled DNA probe. An initial EMSA run will determine which cocktail competes with the probe binding to nuclear proteins which will then run individually in another EMSA to determine the precise competitor (Smith & Humphries, 2008). It is a competition-based method to identify uncertain DNA binding proteins requiring only a prior knowledge of transcription factor consensus sequences.

6.4 Two-dimensional EMSA (2D-EMSA)

The two-dimensional EMSA is a process that combines EMSA with proteomic or sequencing techniques to identify the proteins or the nucleic acid sequences that are present in the formed complexes. Two slightly different protocols have been developed to identify the interacting proteins and another method aims at the target nucleic acid sequence.

An initial approach was proposed by Woo and colleagues as they tried to identify and characterize transcription factors (Woo et al., 2002). A crude nuclear extract is partially purified by gel filtration and the resulting fractions are then bound to the nucleic acid probe and analyzed by EMSA. Meanwhile, in parallel, the pI and molecular weight of the putative interacting protein(s) is estimated as the fractions are analyzed by isoelectric focusing or SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) in order to characterize possible candidates. Next, spots with the predetermined pI and molecular weight of the candidates are excised from a two-dimensional array of nuclear proteins and the proteins are eluted, renatured and tested for their binding ability through EMSA and the spots are afterwards analyzed by mass spectrometry for protein identification. This method is limited to proteins that can re-form into functional nucleic acid-binding conformations after the denaturing SDS-PAGE step, although EMSA can still show results even if renaturation efficiency is low. Because the final EMSA step that confirms the binding is performed with protein eluted from single spots it is only possible to identify proteins that interact with the nucleic acids as monomers or homomultimers. Proteins that only interact when complexed with other proteins will give a negative result on the validation EMSA.

A similar 2D-EMSA technique has since then been developed that incorporates EMSA into a two-dimensional proteomics approach by replacing the isoelectric focusing with EMSA as the first dimension of the 2D method (Stead et al., 2006). The protein sample, in the presence or absence of the nucleic acid, is separated by native PAGE as in a conventional EMSA. The protein bands from both conditions are then separated in a second dimension by denaturing SDS-PAGE. The proteins showing the nucleic acid dependent shift in mobility can be extracted from the gel for mass spectrometry identification. This approach does not require any previous knowledge of the chemical or physical properties of the binding protein and does not require protein renaturation after gel excision. It is also not limited to identify proteins that bind by themselves or as homomultimers and allows the characterization of complexes composed of different proteins.

These 2D approaches were developed by the two groups to study transcription factors, therefore, double stranded DNA is used as a nucleic acid probe but they can also adapted to

other nucleic acid probes making them quite versatile methods to identify nucleic acid-interacting proteins.

Chernov and collaborators have developed a similar protocol with two dimensions but instead of aiming to identify the interacting protein(s) it characterizes and maps the specific protein target sites in regions of the human genome (Chernov et al., 2006). This approach is also based on first separating the complexes from the free nucleic acid in a non-denaturing gel and afterwards separating it under denaturing conditions (Vetchinova et al., 2006). The group used a pool of radioisotope-labeled short DNA sequences covering the genome region of interest and mixed it with a nuclear extract from a specific cell line. The formed complexes were separated in a non-denaturing one-dimensional standard EMSA. The complexes were localized by autoradiography and the gel strip containing them was excised and treated with a denaturing agent, SDS, to disrupt the preformed complexes. The strip is then loaded onto the second-dimension denaturing gel and another electrophoresis is performed. The gel is autoradiographed to determine the location of the freed DNAs, which are afterwards cut from the gel to be analyzed. By pairing this method with high-throughput sequencing the authors were able to identify a multitude of specific protein binding sites within a given genomic region.

6.5 EMSA-three-dimensional-electrophoresis (EMSA-3DE)

A three dimensional approach has very recently emerged to purify nucleic acid binding proteins from complexes separated by EMSA (Jiang et al., 2011). This method focuses on recovering the protein in high yield for subsequent analysis and has been developed to study low abundant transcription factors. In this EMSA-based purification procedure the complexes formed are extracted after a native PAGE retardation assay and applied to two-dimensional electrophoresis, isoelectric focusing and SDS-PAGE. The EMSA conditions are systematically optimized to reduce non-specific binding and increase protein yield. After the three electrophoreses the sample can then be electrotransfered onto a nitrocellulose or polyvinylidene difluoride membrane for southwestern and western blotting analysis to further characterize the complexes. Spots of interest can be cut from the gel or the membrane for protein identification by mass spectrometry.

7. Alternatives to EMSA

There are several alternatives to EMSA used in the analysis of nucleic acid-protein interactions with its own advantages and disadvantages when compared to EMSA.

7.1 Footprinting

Footprinting is essentially a protection assay used to characterize the binding site recognized by a given protein. It relies on the fact that a protein bound to the nucleic acid will protect it and interfere with the modification of the sequence it is bound to. The modification can be chemical or enzymatic and it is usually the endonuclease cleavage of radioisotope-labeled nucleic acid previously mixed with the protein(s) of interest. After cleavage the resulting ladder is analyzed on denaturing polyacrylamide gel and visualized by autoradiography. The gaps in the ladder are indicative of sites protected by the protein or proteins in the mixture (reviewed by Hampshire et al., 2007). This method was originally

developed to characterize sequence selectivity but it is also helpful in estimating the binding strength through a footprinting reaction over a range of protein concentrations. For slow binding reactions footprinting can also be applied to assess the reaction kinetics estimating the association and dissociation rates. Although it is a widely used method, there are other approaches that provide higher throughput as the ones described ahead.

A variant on DNA footprinting is the *in vivo* approach, a technique that enables the detection of DNA-protein interactions as they occur in the cell. *In vivo* footprinting also relies on the fact that the bound protein protects the nucleic acid, at its binding site, from cleavage by endonucleases or modification by a chemical agent. The difference is that the cleavage of DNA is carried out within the nucleus following the *in vivo* binding of the proteins to chromatin. Footprints and endonuclease hypersensitive sites that are due to deformations of DNA in chromatin can be detected by this *in vivo* method. This method has been coupled with deep sequencing to identify DNaseI hypersensitive sites in the genome of different cell lines. It enabled the precise identification of a large number of specific cis-regulatory protein binding events with a single experiment (Boyle et al., 2011). Accordingly, the data obtained by this procedure may be more significant and representative of true events when compared with data obtained by the previously described *in vitro* footprinting.

7.2 Nitrocellulose filter binding

Nitrocellulose filter binding assays were developed in the 70s as a rapid enough method to allow kinetic as well as equilibrium studies of DNA-protein interactions (Riggs et al., 1968 and Riggs et al., 1970 as cited in Helwa & Hoheisel, 2010). The manipulation required is rapid enough to allow such measurements. The assay is based on the premise that proteins can bind to nitrocellulose without losing the ability to bind DNA. After the binding reaction the mixture is separated by electrophoresis and then blotted onto a nitrocellulose membrane. Only protein bound DNA remains on the membrane as the free double-stranded DNA will not be retained on nitrocellulose. The amount of DNA on the membrane can be quantified by measuring the label on the nucleic acid. However, this method has its limitations such as the fact that the proteins involved are not identified or the proportion in which they bind DNA. It also provides no information on the DNA sequence the protein interacts with unless well defined nucleic acid fragments are used and is limited to double stranded DNA as single stranded DNA can bind to nitrocellulose under certain conditions resulting in undesirable background.

7.3 Microfluidic mobility shift assay (MMSA)

The capillary microfluidic mobility shift assay (MMSA) is a method that uses fluorescence-based multi-well capillary electrophoresis to characterize protein-nucleic acid interactions. For example, it has been used effectively in characterizing RNA-protein binding in a study of the interaction between human immunodeficiency virus 1 transactivator of transcription and the transactivation-responsive RNA (Fourtounis et al., 2011). This technique requires only nanoliter amounts of sample that are introduced into microscopic channels and separated by pressure-driven flow and application of a potential difference. The free molecules or complexes are visualized by LED-induced fluorescence, discarding the need for hazardous radiolabeling. With the ability to perform 384-well screening this method has an increased capacity over regular EMSA to be compatible with high-throughput screenings.

7.4 Yeast hybrid systems

The yeast one-hybrid is an approach used to identify proteins that bind a given nucleic acid sequence as opposed to the methods that are suited to identify the nucleic acid sequences preferably recognized by a known protein. The protocol is based on a hybrid prey protein fused to a transcription activation domain that allows the expression of a reporter gene when the prey protein interacts with the DNA bait (reviewed by Deplancke et al., 2004). This method allows for a proteome-scale analysis depending on the prey protein library but only detects monomers that bind the target nucleic acid. Although it is an *in vivo* approach it is performed in yeast (*Saccharomyces cerevisiae*), which may not be the endogenous context, and is limited to DNA-protein interactions.

RNA-protein interactions can be studied with a yeast three-hybrid system that involves the expression in yeast cells of not one but three chimerical molecules, which assemble in order to activate two reporter genes (Kraemer et al., 2000). It represents a modification of the yeast two-hybrid system, widely used to identify protein-protein interactions, that was designed to allow high sensitivity *in vivo* detection of RNA-protein interactions. The yeast three-hybrid system includes: a fusion protein consisting of a DNA binding protein and a RNA-binding protein; a hybrid protein consisting of a transcription activating domain and a peptide thought to interact with a particular RNA; a RNA intermediate that promotes the interaction of the two hybrid proteins, this RNA includes the RNA that interacts with the system's RNA-binding protein and the RNA molecule to be investigated. The successful interaction of these 3 components allows the reconstitution of a transcription factor and subsequent activation of reporter genes (Hook et al., 2005 and Wurster & Maher, 2010)

7.5 ChiP assays

Chromatin immunoprecipitation (ChiP) is a commonly used method to study DNA-binding proteins *in vivo* and a standard method for the identification of transcription binding sites and histone modification locations (reviewed by Massie & Mills, 2008). In this method a cross-linking agent (e.g. formaldehyde) is added to cells to covalently bind proteins and chromatin that are in direct contact. Afterwards, the cells are lysed and chromosomal DNA is isolated and fragmented. Specific antibodies are used to immunoprecipitate the targeted proteins with the cross-linked DNA. The bound nucleic acid is released by reverting the cross-linking and then analyzed. Classically, the DNA was characterized by polymerase chain reaction (PCR) which required some previous knowledge of the candidate DNA regions. Nowadays, the DNA bound to protein is more commonly characterized through more powerful tools either coupled with microarrays that represent the genome (ChIP-chip) or state-of-the-art high-throughput sequencing (ChIP-seq). The improvements in DNA sequencing technology allow tens of millions of sequence reads, therefore ChIP-seq has a major advantage of increased sensitivity and resolution to add to the fact that it is not limited to predetermined probe sets as ChIP-chip. The major strength of the ChIP-based approaches is that they capture complexes *in vivo* and the binding reactions can be studied under different cellular conditions and at different time points. However it also has important limitations. The method requires high-quality antibodies that are available only for a limited number of proteins. To circumvent this, epitope-tagged proteins could be used although it usually implies the introduction of modified genes into the endogenous locus in order to obtain expression at physiological levels. This method does not distinguish between

proteins that bind directly to the genomic DNA and those that only interact with other proteins that do bind.

7.6 SELEX

The Systematic Evolution of Ligands by Exponential Enrichment (SELEX) is a well established method that enables the selection of enriched sequences from a random library that bind recombinant proteins. This procedure starts with the synthesis of the oligonucleotide library and then incubating the generated sequences with the putative interacting protein(s). The sequences that bind are eluted, amplified by PCR and subjected to more rounds of selection with increasing stringency conditions. This allows the identification of the tightest-binding sequences. It is a widely used approach to obtain transcription factors binding motifs as it requires low amounts of purified proteins (Matys et al., 2006). This approach becomes very complicated to use when large numbers of nucleic acid-binding proteins are analyzed as it then requires multiple rounds of selection. Another limitation is the fact that it is aimed at the identification of the best binding DNA targets *in vitro* and does not allow the characterization of the exact *in vivo* selectivity.

7.7 Protein microarray

A protein microarray is a method that allows high-throughput analysis in which labeled nucleic acids are queried against proteins immobilized on a chip (reviewed by Hu et al., 2011). In a functional protein microarray, thousands of purified recombinant proteins can be immobilized in a glass slide in discrete locations forming a high-density protein matrix, providing a flexible platform to characterize different protein activities. It is a very versatile method as it can perform a semi-quantitative analysis of protein binding to a wide range of molecules (nucleic acids, other proteins, antibodies, lipids, glycans…). In theory, it is feasible to print arrays of all the annotated proteins of a given organism originating a whole proteome microarray. However, it implies the expression and purification of each individual protein and several conditions need to be optimized to render the proteins apt for this method. Since the protein is immobilized it is crucial to guarantee that its structural integrity remains intact especially the binding domains that are to be studied.

7.8 Nucleic acid microarrays

Nucleic acid microarrays can also be used for a direct analysis of protein-nucleic acid interactions. In this case it is the nucleic acid that is immobilized and not the protein. Nucleic acid chips are a powerful and versatile tool in biological research. They consist of high-density arrays of oligonucleotides or complementary DNA that can cover a whole genome (reviewed by Stoughton, 2005). For protein-interaction studies, the protein(s) of interest is expressed usually with an epitope tag, and purified. The tag serves two purposes; it helps to isolate the protein through affinity purification, and allows detection by an epitope-specific reporter antibody. After incubation of the protein with the nucleic acid chip the signal intensities at the several array spots can be measured.

7.9 Ribonucleoprotein Immunoprecipitation – Microarray (RIP-chip)

RNA immunoprecipitation and chip hybridization (RIP) is a protocol very similar to ChIP-chip except that it targets RNA-protein interactions rather than DNA-protein (Keene et al.,

2006). RIP-chip is an approach that consists on a microarray profiling of RNAs obtained from immunoprecipitated RNA-protein complexes. Genome-wide arrays are used to identify messenger RNAs (mRNAs) that are present in endogenous messenger ribonucleoprotein complexes making it a great tool to identify the physiological substrates of mRNAs. The endogenous complexes are immunoprecipitated from cell lysates which limits this study to kinetically stable interactions. Even though it can identify RNA-protein complexes with heteromultimers, at least one of the proteins has to be previously known to be the basis of immunoprecipitation and "fish out" the whole complex.

7.10 Crosslinking and Immunoprecicipation (CLIP) and Photoactivable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP)

The RIP-chip method that has just been described is limited to studies of very stable RNA-protein complexes; to remediate this problem another method is available to study RNA-binding proteins. The crosslinking and immunoprecipitation (CLIP) approach uses *in vivo* UV crosslinking prior to the complexes immunoprecipitation to identify less stable interactions (Ule et al., 2003). After immunoprecipitation RNA molecules are separated and cDNA sequencing is carried on. However, this method is not perfect as the commonly used UV 254nm RNA-protein crosslinking has low efficiency and it is difficult to distinguish between crosslinked RNAs from background non-crosslinked fragments that can be detected in the sample due to the presence of abundant cellular RNAs.

A more recent approach tries to further improve the CLIP method using photoreactive ribonucleoside analogs such as 4-thiouridine or 6-thioguanosine (Hafner et al., 2010). In this photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation (PAR-CLIP) protocol the photoreactive nucleosides are incorporated into nascent transcripts within living cells. The irradiation is performed with UV light of 365nm, which induces an efficient crosslink of the labeled cellular RNA to its interacting proteins. The labeled RNAs are isolated after co-immunoprecipitation, and converted into cDNA for deep sequencing. The precise crosslinking position can be identified by mutations in the sequenced cDNA making it possible to distinguish the crosslinked fragments from background.

7.11 High-Throughput Sequencing – Fluorescent Ligand Interaction Profiling (HisT-FLIP)

Very recently a new method was developed to characterize DNA-protein interactions using second-generation sequencing instruments (Nutiu et al., 2011). This method allows high throughput and quantitative measurement of DNA-protein binding affinity. This High-Throughput Sequencing – Fluorescent Ligand Interaction Profiling (HiTS-FLIP) uses the optics of a high-throughput sequencer to visualize *in vitro* binding of a protein to the sequenced DNA in a flow cell. The new method was initially used on a *Saccharomyces cerevisiae* transcription factor. The fluorescently tagged protein was added at different concentrations to a flow cell containing around 88 million DNA clusters, the equivalent of over 160 yeast genomes. The traditional EMSA was used as an independent validation of the dissociation constants obtained and found a high correlation with values obtained with the new method and those from EMSA as reported in literature. This high-throughput method has an obvious advantage in the fact that it can provide hundreds of millions of measurements but is limited to DNA-protein interactions and requires expensive

equipment. Another advantage is that the sequencing instrument can measure multiple fluorescent wavelengths allowing hetero and homodimeric forms to be measured in the same run, using distinct tags on individual proteins.

8. Conclusion

Since the first report, 30 years ago, EMSA became one of the most popular methods for detection and characterization of protein-nucleic acid interactions. Hundreds of protocols have been published accommodating modifications in virtually every parameter influencing the experimental outcome. Improvements were made in all EMSA steps including the methods for preparation of protein samples and purification, synthesis and labeling of nucleic acids, and detection. This allowed enlarging and diversifying the applications of EMSA and resulting in a number of variants of the method.

However, despite the large amount of available literature and protocols trial and error will ultimately be the way to optimize the EMSA conditions for the nucleic acid-protein complex to be analyzed. The guidelines discussed above help to provide an initial protocol adjusted to each study but slight changes may be needed to improve binding and detection of the complexes.

In recent years, the use of highthrouhput approaches to detect biologically relevant interactions, including those between proteins and nucleic acids, was reported. Development of these approaches was made possible, at least in part, by the availability of more sensitive and specific equipment and tools. Although EMSA cannot achieve a high throughput level it remains a valuable tool to confirm the detected interactions.

9. Acknowledgements

We are grateful to Dr. Cristina Branco for constructive comments. Work in the authors' laboratory is supported by Fundação para a Ciência e Tecnologia (PTDC/SAU-MII/098314/2008). CA is a recipient of a FCT PhD grant.

10. References

Adachi, Y.; Chen, W.; Shang, W. & Kamata, T. (2005). Development of a direct and sensitive detection method for DNA-binding proteins based on electrophoretic mobility shift assay and iodoacetamide derivative labelling. *Analytical Biochemistry,* Vol.342, pp. 348-351

Alves, C.; Cheng, H.; Roder, H. & Taylor, J. (2010). Intrinsic disorder and oligomerization of the hepatitis delta virus antigen. *Virology,* Vol. 370, pp 12-21

Apella, E. & Anderson, C. (2001). Post-translational modifications and activation of p53 by genotoxic stresses. *European Journal of Biochesmistry,* Vol.268, No.10, pp. 2764-2772

Baneyx, F. (1999). Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology*, Vol.10, pp. 411-421

Boyle, A.; Song, L.; Lee, B.; London, D.; Keefe, D.; Birney, E.; Iyer, V.; Crawford, G. & Furey, T. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Research,* Vol.21, No.3, pp. 456-464

Chernov, I.; Akopov, S.; Nikolaev, L. & Sverdlov, E. (2006). Identification and mapping of DNA binding proteins target sequences in long genomic regions by two-dimensional EMSA. *BioTechniques,* Vol.41, pp. 90-96

Chi, P.; Van Komen, S.; Sehorn, M.; Sigurdsson, S. & Sung, P. (2006). Roles of ATP binding and ATP hydrolysis in human Rad51 recombinase function. *DNA Repair,* Vol.5, No.3, pp. 381–391

Crothers, D. & Drak, J. (1992). [3] Global features of DNA structure by comparative gel electrophoresis. *Methods in Enzymology*, Vol 212, pp. 46-71

Demarse, N.; Ponnusamy, S.; Spicer, E.; Apohan, E.; Baatz, J.; Ogretmen, B. & Davies, C. (2009). Direct binding of glutharaldehyde 3-phosphate dehydrogenase to telomeric DNA protects telomeres against chemotherapy-induced rapid degradation. *Journal of Molecular Biology,* Vol.394, No.4, pp. 789-803

Deplancke, B.; Dupuy, D.; Vidal, M. & Walhout, A. (2004). A gateway-compatible yeast one-hybrid system. *Genome Research,* Vol.1, 2093-2101

Dignam, J.; Lebovitz, R. & Roeder, R. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Research,* Vol.11, No.5, pp. 1457-1489

Endo, Y. & Sawasaki, T. (2006). Cell-free expression systems for eukaryotic protein production. *Current Opinion in Biotechnology,* Vol.17, pp. 373-380

Filion, G.; Fouvry, L. & Defossez, P.-A. (2006). Using reverse electrophoretic mobility shift assay to measure an compare protein-DNA binding affinities. *Analytical Biochemistry,* Vol.357, pp. 156-158

Fourtounis, J.; Falgueyret, J.-P. & Sayegh, C. (2011). Assessing protein-RNA interactions using microfluidic capillary mobility shift assays. *Analytical Biochemistry,* Vol.411, pp. 161-163

Fried, M. (1989). Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis*, Vol. 10, pp. 366-376

Fried, M. & Crothers, D. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acid Research,* Vol.9, No.23, pp. 6505-6525

Fried, M. & Crothers, D. (1984). Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. *Journal of Molecular Biology,* Vol.172, No.3, pp. 241-262

Fried, M. & Daugherty, M. (1998). Electrophoretic analysis of multiple protein-DNA interactions. *Electrophoreis,* Vol.19, pp. 1247-1253

Garner, M. & Revzin, A. (1981). A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Research,* Vol.9, No.13, pp. 3047-3060

Goh, A.; Coffil, C. & Lane, D. (2011). The role of mutant p53 in human cancer. *The Journal of Pathology,* Vol.223, No.2, pp. 116-126

Granger-Schnarr, M.; Lloubes, R.; de Murcia, G. & Shnarr, M. (1988). Specific protein-DNA complexes: immunodetection of the protein component after gel electrophoreis and Western blotting. *Analytical Biochemistry,* Vol.174, No.1, pp. 235-238

Greco-Stewart, V. & Pelchat, M. (2010). Interaction of host cellular proteins with components of the hepatitis delta virus. *Viruses,* Vol.2, No.1, pp 189-212

Haffner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M.; Jungkamp, A.-C.; Munschauer, M.; Ulrich, A.; Wardle, G.; Dewell, S.; Zavola, M. & Tuschi, T. (2010). PAR-CliP – A method to identify transcriptome-wide binding sites of RNA binding proteins. *Journal of Visualized Experiments,* Vol.41, pp.2034

Hampshire, A.; Rusling, D.; Broughton-Head, V. & Fox, K. (2007). Footprinting: a method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands. *Methods,* Vol.42, pp. 128-140

Hellman, L. & Fried, M. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature Protocols,* Vol.2, No.8, pp. 1849-1861

Helwa, R. & Hoheisel, J. (2010). Analysis of DNA-protein interactions: from nitrocellulose filter binding assays to microarray studies. *Analytical & Bioanalytical Chemistry,* Vol.398, pp. 2551-2561

Holden, N. & Tacon, C. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of Pharmacological and Toxicological Methods,* Vol.63, pp. 7-14

Hope, I. & Struhl, K. (1987). GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. *The EMBO Journal,* Vol.6, No.9, pp. 2781-2784

Hook, B.; Bernstein, D.; Zhang, B. & Wickens, M. (2005). RNA-proteins interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *RNA,* Vol.11, No.2, pp. 227-233

Hu, S.; Xie, Z.; Qian, J.; Blackshaws, S. & Zhu, H. (2011). Functional protein microarray technology. *WIREs Systems Biology and Medicine,* Vol.3, pp. 255-268

Jarvis, D. (2009). Baculovirus-insect cell expression systems. *Methods in Enzymology,* Vol.463, pp. 191-222

Jiang, D.; Jia, Y. & Jarrett, H.W. (2011). Transcription factor proteomics: Identification by a novel gel mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography-electrospray-mass spectrometry analysis. *Journal of Chromatography A,* Vol.1218, No.39, pp. 7003-7015

Jiang, S.; Macias, M. & Jarrett, H. (2010). Purification and identification of a transcription factor, USF2, binding to E-box elements in the promoter of human telomerase reverse transcriptase (hTERT). *Proteomics,* Vol.10, No.2, pp. 203-211

Jimenez, J. (2010). Protein-DNA interaction at the origin of neurological diseases: a hypothesis. *Journal of Alzheimer's Disease,* Vol.22, No.2, pp. 375-391

Keene, J.; Komisarow, J. & Friedersdorf, M. (2006). RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nature Protocols,* Vol.1, pp. 302-307

Kerr, L. (1995). [42] Electrophoretic mobility shift assay. *Methods in Enzymology,* Vol.254, pp. 619-632

Kraemer, B.; Zhang, B.; SenGupta, D.; Fields, S. & Wickens, M. (2000). Using the yeast three-hybrid system to detect and analyze RNA-protein interactions. *Methods in Enzymology,* Vol.328, pp. 297-321

Lane, D.; Prentki, P. & Chandler, M. (1992). Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiological Reviews,* Vol.56, No.4, pp. 509-528

Makino, T.; Skretas, G. & Georgiou, G. (2011). Strain engeneering for improved expression of recombinant proteins in bacteria. *Microbial Cell Factories,* Vol.10, No.1, pp. 32-42

Massie, C. & Mills, I. (2008). ChIPping away at gene regulation. *EMBO Reports,* Vol.9, No.4, pp. 337-343

Matys, V.; Kel-Margoulis, O.; Fricke, E.; Liebich, I.; Land, S.; Barre-Dirrie, A.; Reuter, I.; Checkmenev, D.; Krull, M.; Hornischer, K.; Voss, N.; Stegmaier, P.; Lewicki-Potapov, B.; Saxel, H.; Kel, A. & Wingender, E. (2006). TRANSFAC and its module TRANSComple: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research,* 1 34 Database issue D108-110

Nordheim, A. & Meese, K. (1988). Topoisomer gel retardation: detection of anti-Z-DNA antibodies bound to Z-DNA within supercoiled DNA minicircles. *Nucleic Acids Research,* Vol.16, No.1, pp. 21-37

Nutiu, R.; Friedman, R.; Luo, S.; Khrebtukova, I.; Silva, D.; Li, R.; Zhang, L.; Schroth, G. & Burge, C. (2011). Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nature Biotechnology,* Vol.29, No.7, pp. 659-664

Palecek, E.; Vlk, D.; Stankova, V.; Brazda, V.; Vojtesek, B.; Hupp, T.; Schaper, A. & Jovin, T. (1997). Tumor suppressor protein p53 binds preferentially to supercoiled DNA. *Oncogene,* Vol.15, No.18, pp. 2201-2209

Puzio-Kuter, A. (2011). The role of p53 in metabolic regulation. *Genes & Cancer,* Vol.2, No.4, pp. 385-391

Rizzetto, M. (2009). Hepatitis D: Thirty years after. *Journal of Hepatology,* Vol.50, pp. 1043-1050

Ruscher, K.; Reuter, M.; Kupper, D.; Trendelenburg, G.; Dirnagl, U. & Meisel, A. (2000). A fluorescence based non-radioactive electrophoretic mobility shift assay. *Journal of Biotechnology,* Vol78, pp. 163-170

Saida, F.; Uzan, M.; Odaert, B. & Bontems, F. (2006). Expression of highly toxic genes in E. coli: special strategies and genetic tools. *Current Protein and Peptide Science,* Vol.7, pp. 47-56

Sandman, K.; Krzycki, J.; Dobrinski, B.; Lurz, R. & Reeve, J. (1990). HMf, a DNA-binding protein isolated from the hyperthermophilic archaeon Methanothermus fervidus, is most closely related to histone. *Proceedings of the National Academy of Sciences of the United States of America,* Vol.87, No.15, pp. 5788-5791

Simicevic, J. & Deplancke, B. (2010). DNA-centered approaches to characterize regulatory protein-DNA interaction complexes. *Molecular BioSystems,* Vol.6, No.3, pp. 462-468

Spinner, D.; Liu, S.; Wang, S-W. & Schmidt, J. (2002). Interaction of the myogenic determination factor myogenin with E12 and a DNA target: mechanism and kinetics. *Journal of Molecular Biology,* Vol.317, No.3, pp. 431-445

Stead, J.; Keen, J. & McDowall, K. (2006). The identification of nucleic acid-interacting proteins using a simple proteomics-based approach that directly incorporates the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics,* Vol.5, No.9, pp. 1697-1702

Stoughton, R. (2005). Applications of DNA microarrays in biology. *Annual Review of Biochemistry,* Vol.74, pp. 53-82

Ule, J.; Jensen, K.; Ruggiu, M.; Mele, A.; Ule, A. & Darnell, R. (2003). CLIP identifies Nova-regulated RNA networks in the brain. *Science,* Vol.302, No.5648, pp. 1212-1215

Vetchinova, A.; Akopov, S.; Chernov, I.; Nikolaev, L. & Sverdlov, E. (2006). Two-dimensional electrophoretic mobility shift assay: identification and mapping of transcription factor CTCF target sequences within an FXYD5-COX7A1 region of human chromosome 19. *Analytical Biochemistry,* Vol.354, pp. 85-93

Vossen, K.; Wolz, R.; Daugherty, M. & Fried, M. (1997). Role of macromolecular hydration in the binding of the Escherichia coli cyclic AMP receptor to DNA. *Biochemistry,* Vol.36, No.39, pp. 11640-11647

Walker, J. (1994). Gradient SDS polyacrylamide gel electrophoresis of proteins. *Methods in Molecular Biology,* Vol.32, pp. 35-38

Woo, A.; Dods, J.; Susanto, E.; Ulgiati, D. & Abraham, L. (2002). A proteomics approach for the identification of DNA binding activities observed in the electrophoretic mobility shift assay. *Molecular & Cellular Proteomics,* Vol.1, No.6, pp. 472-478

Wurster, S. & Maher, L. (2010). Selections that optimize RNA display in the yeast three-hybrid system. *RNA,* Vol.16, No.2, pp. 253-258

Review
Future Virology

Order and disorder in viral proteins: new insights into an old paradigm

Carolina Alves[1] & Celso Cunha*[1]
[1]Medical Microbiology Unit, Center for Malaria and Tropical Diseases, Institute of Hygiene and Tropical Medicine, Nova University, Lisbon, Portugal
*Author for correspondence: Molecular Microbiology Unit, Center for Malaria and Tropical Diseases, Institute of Hygiene and Tropical Medicine, Rua Junqueira 100, 1349-008 Lisbon, Portugal • Tel.: +351 213 652 620

The conventional dogma stating that proteins must fold into a well-defined structure in order to display biological function is being challenged everyday as new data emerges on the relevance of disordered regions and intrinsically disordered proteins. Viral proteins in particular can benefit greatly from the conformational flexibility granted by partially folded or unfolded protein segments. It enables them to adapt to hostile and changing environmental conditions, interact with the required host machinery while evading host defence mechanisms, and tolerate the high mutation rates viral genomes are prone to. In this review, we will summarize and discuss the importance of the recent research field of protein disorder that is proving vital to gain better understanding of the roles and functions of viral proteins.

Intrinsically disordered proteins: a paradigm shift

The notion that a rigid 3D structure is a prerequisite for a protein to be functional has been accepted for a long time. The origin of this structure–function paradigm can be dated back to the lock-and-key hypothesis, proposed in 1894 by Emil Fischer, to explain enzymatic specificity [1]. This concept was later validated as the crystal structure of proteins was beginning to be solved by X-ray diffraction [2]. Thus, for a long time, the conventional view was that a functional protein folds into a unique and stable 3D structure, perfectly matching the substrate to which it should bind.

Until a couple of decades ago, the idea that disordered proteins could have specific functions was almost outlandish. Nevertheless, occasionally, flexible but functional proteins were discovered or rediscovered (reviewed in [2]). Information on these flexible proteins was scarce and shallow, since they did not fit the structure–function paradigm; they were considered mere exceptions to the rule. However, throughout the years, different terms appeared to designate these nonconventional proteins, such as partially folded, flexible, pliable, chameleon, vulnerable and natively unfolded, among others [2].

Only in the late 1990s did researchers start to realize that these unstructured proteins were representative of a broad class of rather important proteins [3–7]. In recent years, the term 'intrinsically disordered proteins' (IDPs) has become the most widely accepted, along with 'intrinsically disordered regions' (IDRs), to define proteins or protein segments that are biologically functional, although they exist as collapsed or extended mobile conformational ensembles [2]. However, the coined terms have attracted some controversy, as criticism arises from the fact that for many proteins disorder does not last, as they become ordered when bound to partners and even crystallize [8]. Critics say the concept should be discarded, which is not realistic as the collection of highly disordered proteins grows and it has been shown that proteins can remain disordered even when bound to their partners [9].

The increasing number of experimentally characterized IDPs led to the creation of DisProt in 2007, a database of disordered proteins [10].

Characterization of IDPs: disorder prediction & experimental methods

Like ordered proteins, the structures of which can be inferred from their amino acid sequence, IDPs and IDRs have signature characteristics that allow the prediction of disorder based on sequence data alone. A mark of probable intrinsic disorder is a low hydrophobic amino acid content, which usually form the core in folded proteins, and high levels of polar amino acids, conferring high net charge to the disordered protein [4,11]. Low hydrophobicity is thought to lead to a low driving force for protein compaction, and high net charge may result in strong electrostatic repulsion. Hence, these features contribute to structural disorder. In 1978, it had already been suggested that IDPs have amino acid compositions that differ from ordered proteins and therefore disorder could

Keywords
conformational flexibility • intrinsic disorder prediction • intrinsically disordered protein • intrinsically disordered region • viral proteins

be predicted by the abnormally high ratio of charged residues to hydrophobic residues [12]. However, this early mode of intrinsic disorder prediction was based on a very small set of proteins and never tested for other sets. Later, it was shown that IDPs are deficient in what has been called 'order-promoting' amino acids such as Ile, Leu, Val, Trp, Tyr, Phe, Cys and Asn; and are enriched in 'disorder-promoting' amino acids like Ala, Arg, Gly, Gln, Ser, Glu, Lys and Pro [7].

IDP and IDR sequence characteristics were used to design algorithms to predict intrinsic disorder. In 1997 and 2000 the first well-tested predictors of intrinsic disorder were published [4,13]. Since then, more than 50 predictors have been developed to evaluate disordered regions based on amino acid sequences on a per-residue analysis [14]. Many predictors can be assessed via public servers such as the PONDR® predictors [13,15], FoldIndex [16], Dis-EMBL [17], GlobPlot [18] and PrDOS [19], to name a few. There are also meta-predictors, such as PONDR-FIT [20], which combines the output of six individual disorder predictors, and metaPrDOS [21], which is claimed to have an improved prediction accuracy. There is also a web metaserver, MeDor, which permits a rapid simultaneous analysis by multiple predictors of a given sequence, retrieving a integrated view of the outputs [22].

Disorder predictors have been applied to entire genomes to assess the extent of intrinsic disorder, with predictions spanning the three kingdoms; prokaryota, archea and eukaryota. Studies have shown that IDPs and IDRs are not rare exceptions, but are highly abundant in all species. In fact, almost 70% of proteins in the Protein DataBase by 2003 had IDRs [23]. IDPs are more common in eukaryota than in prokaryota and archea, with up to 30% of eukaryotic proteins being mostly disordered [24]. However, in a recent study comparing 194 eukaryotic and 87 prokaryotic proteomes, researchers found an overlap in the frequency of predicted disorder, which spans a wide range in both kingdoms [25]. Although prokaryotes were found to have a lower average disorder than eukaryotes, both groups have a very broad range of predicted disorder, with scores of 0.12–0.35 for prokaryotes and 0.1–0.41 in eukaryotes [25]. The same study also found the highest levels of predicted intrinsic disorder in single-celled protists, which were often higher than complex eukaryotic organisms. Hence, a new theory was proposed by the authors correlating intrinsic disorder with lifestyle and not only with the complexity of the organism [25]. They suggest that a decreased level of disorder reflects an adaptation to the environment, as low values were found in intracellular parasites and endosymbionts, while organisms such as host-changing parasites, which lead a varied lifestyle in changing habitats have higher levels of disorder [25].

The amount of predicted intrinsic disorder in viral proteins has been compared with that of eukaryotes, even though eukaryotic proteomes contain more proteins with long disordered sequences and viral proteomes have more short disordered segments [26]. Nevertheless, viruses have also been shown to have the widest spread of proteome disorder, ranging from 7.3% of disordered residues in human coronavirus to up to 77.3% in avian carcinoma virus [27]. Viral proteins display a high propensity for intrinsic disorder, as they tend to have reduced amounts of hydrophobic residues and a high proportion of polar amino acid residues [26]. In particular, RNA viruses, which display the highest mutation rates, also have a high incidence of disordered regions and significantly lower van der Waals contact densities, reflecting the intensity of the 3D interaction network of a protein [28]. As will be examined later, viral proteins profit at different levels from the flexibility that results from intrinsic disorder.

IDPs exist in an ensemble of conformational states, so the determination of one unique structure is not possible and a multiparametric approach is required. The dynamic structural characterization of an IDP relies on different parameters from a variety of physiochemical methods to obtain information on different aspects of the protein, such as overall compactness, conformational stability, presence of residual secondary structure, transient long-range contacts and regions of more or less mobility [29]. Most of the techniques used were initially developed to analyze ordered proteins, so data analysis needs to be performed carefully as results for intrinsic disorder are usually due to lack of signals characteristic for ordered sequences. That is also why several methods should be used and results compared with one another in order to minimize ambiguity [29]. Thorough reviews and books on this topic can be found in the literature [29–31]. Some techniques will be briefly summarized below, of which the most commonly used to obtain data describing IDPs' structures are nuclear magnetic resonance (NMR) and small angle X-ray scattering [32].

X-ray crystallography is the 'classic' technique for determining a protein's crystal structure.

With this method the higher flexibility of the atoms in disordered regions results in noncoherent X-ray scattering, making them invisible. The outcome is a region with missing electron density corresponding to the disordered segment [7]. This method depends on the ability of proteins to form crystals, and highly disordered proteins, in general, are not able to crystallize. As an example, there is the hepatitis δ antigen, a viral IDP which at least three different research groups have tried to crystallize, but failed [33]. Only a small peptide including the less disordered region of the antigen was crystallized [33,34]. An NMR spectroscopy approach is commonly used; it provides information on a residue-by-residue basis, supplying little information on the overall shape of the protein. 1D NMR is limited to small molecules, therefore, heteronuclear multidimensional NMR is preferable for larger proteins. The latter method can be used to obtain precise data on 3D structures and also provides direct measurements of IDRs' mobility [35]. Circular dichroism (CD) in the far-UV region provides estimates of secondary structure and near-UV CD displays sharp peaks for aromatic groups when the protein is ordered. CD lacks residue-specific information and the data obtained is not very clear when the proteins contain ordered and disordered segments [7]. The level of protein compaction or hydrodynamic dimension can be assessed by different techniques, such as gel filtration, viscometry, small angle X-ray scattering or sedimentation [29]. Small angle X-ray scattering can also be used to determine a protein's degree of globularity, providing information on the presence or absence of a tightly packed core [29]. Proteolytic degradation is also a method to identify disordered segments in a protein, since flexibility is a major determinant for digestion of possible cut sites [36]. A structured protein needs a segment of more than ten residues to be unfolded so that typical proteases can cut the protein [7].

Note that many IDPs and IDRs adopt a well-defined structure when bound to their partners [37]. In these cases, the structure of the bound protein can be readily solved using methods to characterize ordered proteins.

Intrinsically disordered viral proteins

Viral proteins may be considered to be a very peculiar group of proteins. They rarely have homologs amongst modern cells, which suggests that viruses are very antique [38]. Viruses must adapt to fast-changing surroundings, survive in their host's environment as well as inside the host, while evading defense mechanisms. In order to adapt, viral genomes are subject to very high mutation rates, ranging from 10^{-5} to 10^{-3} nucleotide exchanges per generation for RNA viruses and 10^{-8} to 10^{-5} for DNA viruses [39]. The higher mutation rate found in RNA viruses probably reflects the lack of RNA repair mechanisms. In comparison, bacteria and eukaryotes have a mutation rate of 10^{-9} on average [39]. Since viruses have highly compact genomes, often with overlapping reading frames, a single mutation can have an impact on more than one viral protein [40]. Finally, viral proteins usually need to perform numerous interactions with host cell components during the different steps of the virus lifecycle, from entry, to replication, to the formation and exit of new infectious particles. Viral proteins must interact with host membranes, host proteins and in some cases host nucleic acids, even though viral proteins are often phylogenetically separated from their host proteins [40,41].

All these features make it extremely interesting to assess whether they are associated with the unique biophysical characteristics of viral proteins. As will be discussed later, intrinsic disorder may be a way for viral proteins to cope with these distinctive circumstances, as the resulting plasticity can confer a number of exceptional functional advantages. IDPs are more flexible, and without a rigid compact structure viral proteins can be highly promiscuous and take part in several interactions with multiple partners. IDRs in particular can act as flexible linkers between functional domains, enabling mechanisms that facilitate binding and promiscuity. These flexible linkers can also help viral proteins to elude the host cell's immune system by making it difficult for the epitopes to be properly recognized. Disorder can also be a way to cope with the high mutation rates mentioned above, which are characteristic of viruses. High flexibility, resulting from low interactions between amino acids, can be linked to high adaptability and represent a strategy to buffer the deleterious effects of mutations. An already-unstructured protein has less to lose from a substitution than a highly structured one, as it is already unfolded.

Although viral proteins can clearly benefit from the conformational flexibility granted by intrinsic disorder, not all viral proteins are IDPs or even have IDRs. Intrinsic disorder predictors have been used in a comparative analysis of viral proteins, which has shown a relationship between

the level of predicted disorder and the location of a protein within the virion [42]. The study began with the construction of a database including viral proteins from influenza A and HIV-related viruses followed by the comparison of protein sequence, predicted structure, function and location within the virion. The results showed, particularly for influenza virus, a correlation between protein disorder and proximity to the RNA core of the virion. The closer a protein was located to the core, the higher the percentage of disorder. This finding can be related to the fact that proteins at the core are more likely to interact with the viral RNA. Nucleic acid-binding proteins are commonly disordered proteins or at least have an IDR at the nucleic acid binding site [43]. The association between intrinsic disorder and protein location in the virion is not observed in HIV, mostly due to the presence of enzymes close to the core, which are usually predominantly structured proteins [44–46]. The matrix proteins of both influenza A and HIV viruses appear to be relatively disordered. The HIV matrix protein was predicted to be highly disordered, whereas influenza virus matrix protein was predicted to be somewhat ordered or less disordered [42]. Regarding the surface proteins, although they found that the surface protein of HIV, gp120, had a low predicted disorder value across all the analyzed strains, the transmembrane protein gp41 had high levels of disordered residues [42]. Influenza A surface proteins, hemagglutinin (HA) and neuraminidase, were predicted to be mostly ordered proteins [42]. However, the study found that the levels of predicted disorder vary among the subtypes, which may be linked with different levels of virulence as will be discussed later [42,47].

Promiscuity of flexible viral proteins

Intrinsic disorder or conformational flexibility enable the disordered proteins to adapt to and interact with several distinct partners. One IDR can bind multiple partners by gaining very different structures [48]. IDPs can perform different interactions: they can be involved in highly stable complexes or in signaling interactions in which they transit between the bound and unbound state as a dynamic and sensitive on–off switch [49]. IDPs' ability to have different conformations depending on environmental conditions allows them to exercise different functions in different contexts.

Binding promiscuity is a key characteristic for viral proteins; even though some viruses have a genome that encodes several proteins, it is not sufficient on it's own, and they require host cell machinery to complete their life cycles. As mentioned before, viruses have very compact genomes and by having viral proteins with IDRs or even the entire protein disordered, a single protein can be involved in different tasks by interacting with different partners. We will describe three viral proteins that exemplify how intrinsic disorder relates to binding promiscuity.

A clear example of the importance of intrinsic disorder is the hepatitis δ virus (HDV), which has the smallest RNA genome of any animal virus, encoding only one protein, the δ antigen (HDAg) [50]. HDAg is a small, 195 amino acid-long protein essential for viral replication although it has no known enzymatic activity [51]. It has been shown to be an IDP both by using a meta-predictor of intrinsic disorder as well as by CD measurements [33]. The lack of viral proteins with enzymatic activity implies that the HDV replication cycle relies entirely on its only viral protein and host cell components. Therefore, HDAg is a highly promiscuous protein with multiple partners identified in the host proteome through different approaches, although the role of these interactions is mostly unclear and still being studied [52,53]. HDAg also seems to lack nucleic acid-binding specificity, as the full-length HDAg binds HDV RNA as well as other RNAs and even dsDNA *in vitro* [33].

Another example is HCV NS5A, which has a key role in viral replication and is also involved in particle assembly [48,54] NS5A is not fully disordered, it is a membrane-associated protein with an anchor on its N-terminal, but its cytoplasmic portion, which is divided into three domains, has a disordered region. Domain 1 of NS5A is a highly conserved and ordered sequence [55], whereas domains 2 and 3 are less conserved and have highly disordered regions [56,57]. NS5A is a well-studied promiscuous protein and some of the interactions involving its disordered domains have been identified [58]. Specifically, distinct interaction modes have been described for the binding motifs present in domain 2, which seem to disturb host regulation processes like signaling pathways and apoptosis [59].

Our third example comes from the measles virus and its nucleoprotein (N) that forms the viral nucleocapsid. The N protein has an IDR in the C-terminal [60], which performs functions essential for transcription and replication by interacting with the phosphoprotein of the viral polymerase complex [61]. Beyond this crucial interaction, the C-terminal tail of the N protein has been shown to interact with several host

components, including a cellular receptor, and even components of the cellular cytoskeleton [62]. The above-mentioned phosphoprotein of the measles virus is an important polymerase cofactor and also contains long disordered regions. The phosphoprotein IDR is thought to be involved in recruiting the transcriptional machinery [62]. It has been shown that when the IDRs of both N protein and phosphoprotein bind, some flexibility persists, with most of the N protein IDR remaining disordered within the complex [63]. Based on this finding, it has been suggested that the disordered nature of the N protein IDR serves as a platform for interaction with other partners [63]. Structural disorder was successively shown to be a common feature present in the nucleoproteins of paramyxoviruses [64]. Unstructured segments were also found in abundance in the nucleoproteins and phosphoproteins of the Nipah and Hendra virus [63].

Intrinsic disorder to cope with high mutation rates

The abundance of intrinsically disordered residues in viral proteins can also be related to the high mutation rates characteristic of viral genomes, as a strategy to buffer the possible deleterious effects of mutations. In addtion, the effects of mutations on viral genomes should have a particularly high impact, due to the overlapping reading frames common in compact viral genomes [40]. Two hypotheses have been proposed to explain why overlapping genes tend to encode disordered proteins [65]. First, the new protein of each overlap is disordered because it is less likely to generate a novel fixed 3D structure at birth [65]. Second, proteins with intrinsic disorder are subject to less structural constraint [66], therefore, disorder could be a mode adopted to alleviate evolutionary constraints imposed by the overlap [65].

A study by Tokuriki and colleagues has analyzed 26 proteins from RNA viruses and 19 DNA viral proteins and compared them with 26 thermophilic proteins, 26 mesophilic eukaryotic, and 26 mesophilic prokaryotic proteins to assess the effects of mutations on protein conformational stability by comparing $\Delta\Delta G$ values [28]. The average $\Delta\Delta G$ values for all possible mutations are a measure of how the protein reacts to destabilizing mutations. Viral proteins, particularly from RNA viruses, displayed the lowest $\Delta\Delta G$ values, being on average 0.20 kcal/mol lower than other proteins of the same size and 0.26 kcal/mol lower than thermophilic proteins, which, being more compact, seems to lose more stability with individual mutations and therefore have higher $\Delta\Delta G$ values [28]. The robustness of thermophilic proteins is related to the well-packed and very compact and stable hydrophobic core. The high thermodynamic stability seems to be related to a high mutational tolerance. On the other hand, viral proteins appear to benefit from being in the other end of the spectrum, in other words possessing high levels of intrinsic disorder. In this case, the low contact densities and high disorder was proposed to be the reason behind the low destabilizing effects of mutations.

However, it should be noted that other parameters may be relevant to buffering mutation effects. These can include genetic redundancy [67], host chaperones [68] or the existence of coevolving quasispecies, as is the case with poliovirus and HCV [69,70].

Eluding defence mechanisms

Intrinsically disordered residues in viral surface proteins may be involved in mechanisms to evade host immune response and even drugs. Conformational flexibility could present a strategy to evade drug binding with little loss of function [71].

As an example, conformational diversity, structural plasticity and rearrangements have been suggested to play a central role in HIV's immune evasion [72]. Although, as mentioned before, HIV surface protein gp120 is substantially ordered, it has been observed that it has some outstanding structural plasticity due to the presence of small IDRs [73]. The gp120 protrudes out of the virus lipid bilayer and is implicated in important roles such as attachment and penetration of host cells [73]. Moreover, a structural study on the resistance of influenza virus neuraminidase to an inhibitor, oseltamivir, has shown that the formation of a compact hydrophobic pocket is a prerequisite for the tight binding of the inhibitor with the hydrophobic side chain. When the compact formation is not obtained, the inhibitor is unable to engage the active site and consequently binds less tightly [71].

The antigenicity of a protein usually resides in a restricted number of antigenic determinants, epitopes, exposed on its surface. These have been shown, long ago, to correspond to surface segments with high mobility, probably to help the determinant to adjust to the antibody [74]. On the other hand, it has also been discovered that an effective epitope, despite being mobile, should have a predisposition to become ordered

and form a defined structure [75]. In fact, it has been reported in some cases that long disordered regions promote weak immune response or can even be nonimmunogenic [42].

Anti-HIV neutralizing antibodies are likely to represent a key element in future vaccine development, and the potential targets are located in viral external envelope proteins. For HIV-1 the main determinant targeted is the third variable loop region of gp120 [76]. It has a highly variable sequence and is conformationally heterogeneous, which contributes to the virus' ability to escape the host immune system [77]. Consequently, the third variable loop is likely to escape detection by antibodies designed to recognize only specific conformations. It has been thus proposed that this elusion mechanism relies on the fact that the small disordered region, with high flexibility, 'confuses' the immune system, weakening its response [42].

Predicting disorder, predicting virulence

In 2009, a research group applied a disorder predictor to try to understand the different levels of virulence between subtypes of influenza A viruses [47]. They gave particular attention to the mysterious disappearance of the deadly 1918 H1N1 virus and the decrease in virulence in the ensuing H1N1 strains. The protein of interest in this study was HA, a surface glycoprotein with an important role in viral entry. Using an intrinsic disorder predictor, PONDR®VLXT, to compare several subtypes and strains, they found a number of differences between HA proteins of virulent and nonvirulent strains of particular interest. The authors reported a specific region, comprising amino acids 68–79 of HA2, which is in contact with the receptor chain HA1 and is likely to influence the motion of the exposed portion of HA. Predicted disorder of this small region was observed for the virulent strains (1918 H1N1 and H5N1), but is absent from less virulent strains (1930 H1N1 and H3N2). In this case, this IDR is a linker between ordered regions and it is likely that the motions caused by the disordered region can impair recognition by the host immune system, therefore increasing the virulence of 1918 H1N1 and H5N1. This reinforces the idea that IDRs can be used as predictors of influenza A virulence, even though other factors may contribute as well.

Another example, which links intrinsic disorder with a more severe disease course relates to the E6 and E7 proteins of HPV. There are over 100 different types of HPVs that cause benign papillomas and are risk factors for the development of carcinomas [78]. HPVs can be grouped into low- and high-risk types with respect to cancer, and two of its seven nonstructural proteins, E6 and E7, are known to function as oncoproteins in the high-risk HPVs. Their role in the oncogenic process is due to their ability to interact with tumor suppressors retinoblastoma and p53 [78]. In high-risk HPVs these two proteins have an increased amount of intrinsic disorder when compared with the ones expressed by low-risk HPVs [79]. Regarding E7, it has been shown that high-risk HPV16 E7 and low-risk HPV6 E7 share only 50% of their amino acid sequence and there is also a divergence in functionality [79]. HPV16 E7 is also highly promiscuous, with extreme functional plasticity demonstrated by interacting with a great number of host cell proteins [80]. The E6 protein from low-risk HPVs also lacks functions that the high-risk HPV E6 can perform, mostly correlating with oncogenic activity [79]. E6 is a 150 amino acid-long protein and, although it is less disordered than in E7, in high-risk HPVs an alternative transcript encoding only the first 50 amino acids of E6 appears, which is highly disordered and incredibly promiscuous [81]. This fragment, which only occurs in high-risk HPVs, forms low-molecular-weight species with only residual structure that can oligomerize into different conformations and hence interact with a wide range of partners [81]. It can play a role, directly or indirectly, in several cellular processes deregulating the metabolism of the host cell and causing tumorigenesis [81].

Future perspective

In the past couple of decades the old dogma in structural biology, claiming that a functional protein must have a well-defined 3D structure, has been challenged as researchers are realizing the importance of intrinsic disorder in proteins. This lack of conformation in proteins is particularly relevant for virus research, as viral proteins have been found to contain an abundance of disordered domains. The flexibility granted by intrinsic disorder may represent a 'strategy' for viral proteins to cope with their peculiarities and their many roles played during replication.

More viral proteins need to be characterized, regarding their (lack of) structure to better understand the possible significance of intrinsic disorder and its importance in viral biology.

As an initial approach, predictors of intrinsic disorder can be used to screen viral genomes, but eventually, predicted viral IDPs need to be physiochemically analyzed. A set of complementary methods should be used so that the conformational ensembles can be properly characterized.

Intrinsically disordered binding sites and the resulting binding promiscuity and weak interactions led researchers to believe that targeting IDPs for drug treatments was unviable. However, data is starting to emerge, which shows that selective blocking of interactions between IDPs and their partners is possible [82]. Two mechanisms can be used to block protein–protein interactions involving an IDP; either using a small molecule that binds the IDP or a small molecule that binds the partner. The advantage is that an IDP or IDR, being promiscuous, can probably bind several chemically different molecules and also have several different binding sites. Progress in this area can be crucial to develop drugs targeting viral IDPs, mainly viral surface proteins, as these are crucial in the first step of the virus lifecycle, cell attachment, and are involved in recognition by host cell defenses.

In conclusion, this emerging field of research will probably be crucial to improving our knowledge of viruses, their replication strategies, evolution, and interaction with host cells and organisms. Hopefully, the increasing knowledge and data concerning IDP characterization and function may enable us to develop new treatment approaches and solutions leading to an improved control of viral diseases in the near future.

Financial & competing interests disclosure

Work in the authors' lab is supported by Fundação para a Ciência e Tecnologia (PTDC/SAU-MII/098314/2008). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Intrinsically disordered proteins: a paradigm shift

- The past and future of structural biology: intrinsically disordered proteins challenge the old 'structure to function' paradigm.

Characterization of IDPs: disorder prediction & experimental methods

- IDP sequence characteristics allow intrinsic disorder prediction.
- Comparison of IDP prevalence in the three kingdoms.
- Physiochemical methods to characterize IDPs and IDRs.

Intrinsically disordered viral proteins

- Characteristics that make viral proteins so peculiar.
- How viral proteins can profit from intrinsic disorder.
- Different levels of intrinsic disorder amongst viral proteins through two examples: HIV and influenza A.

Promiscuity of flexible viral proteins

- Relating intrinsic disorder and binding promiscuity.
- Examples of promiscuous viral proteins: hepatitis δ antigen; hepatitis C NS5A; measles virus nucleoprotein.

Intrinsic disorder to cope with high mutation rates

- Intrinsic disorder as a strategy to buffer proteins from the effects of mutations.
- Effects of mutations in proteins with different disorder levels: comparing sets of viral proteins with thermophilic, eukaryotic and prokaryotic proteins.

Eluding host defence mechanisms

- Intrinsic disorder in viral proteins as a strategy to avoid the host immune system.
- Example: HIV gp120 surface protein.

Predicting disorder, predicting virulence

- IDR present in hemagglutinin is linked with increased virulence in influenza A.
- HPV nonstructural proteins E6 and E7: increased intrinsic disorder leads them to function as oncoproteins in high-risk HPVs.

Future perspective

- Thorough characterization of viral proteins: predicting disorder and experimental characterization.
- IDRs as targets for future drug development.

References

Papers of special note have been highlighted as:
▪ of interest
▪▪ of considerable interest

1. Lemieux RU, Spohr U. How Emil Fischer was led to the lock and key concept for enzyme specificity. *Adv. Carbohydr. Chem. Biochem.* 50, 1–20 (1994).
2. Uversky VN. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* 43, 1090–1103 (2011).

▪ **An updated review on the general characteristics of intrinsically disordered proteins (IDPs).**

3. Wright P, Dyson H. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331 (1999).
4. Uversky V, Gillespie J, Fink AL. Why are 'natively unfolded' proteins unstructured under physiologic conditions? *Proteins* 41, 415–427 (2000).
5. Tompa P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27(10), 527–533 (2002).
6. Romero P, Obradovic Z, Kissinger C *et al.* Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput.* 437–448 (1998).
7. Dunker AK, Lawson JD, Brown CJ *et al.* Intrinsically disordered protein. *J. Mol. Graph. Model.* 19, 26–59 (2001).
8. Chouard T. Structural biology: breaking the protein rules. *Nature.* 471 (7337), 151–153 (2011).
9. Mittag T, Orlicky S, Choy W *et al.* Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *PNAS* 105, 17772–17777 (2008).
10. Sickmeier M, Hamilton JA, LeGall T *et al.* DisProt: the database of disordered proteins. *Nucleic Acids Res.* 35, D786–D793 (2007).
11. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 42, 38–48 (2001).
12. Williams RJ. The conformational mobility of proteins and its functional significance. *Biochem. Soc. Trans.* 6, 1123–1126 (1978).
13. Romero P, Obradovic Z, Kissinger C, Villafranca JE, Dunker AK. Identifying disordered regions in proteins from amino acid sequence. *Proc. Int. Conference Neural Networks (ICNN'97)* 1, 90–95 (1997).
14. He B, Wang K, Liu Y, Xue B, Uversky VN, Dunker AK. Predicting intrinsic disorder in proteins: an overview. *Cell Res.* 19, 929–949 (2009).

▪ **Extensive review on disorder prediction.**

15. Peng K, Radivojac P, Vucetic S, Dunker AK, Obradovic Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7, 208–224 (2006).
16. Prilusky J, Felder CE, Zeev-Ben-Mordehai T *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* 21, 3435–3438 (2005).
17. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure* 11, 1453–1459 (2003).
18. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31, 3701–3708 (2003).
19. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464 (2007).
20. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* 1804(4), 996–1010 (2010).
21. Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 24, 1344–1348 (2008).
22. Lieutaud P, Canard B, Longhi S. MeDor: a metaserver for predicting protein disorder. *BMC Genomics* 9(Suppl. 2), S25 (2008).
23. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. *Proteins* 53(Suppl. 6), 566–572 (2003).
24. Oldfield C, Cheng Y, Cortese M, Brown C. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44, 1989–2000 (2005).
25. Pancsa R, Tompa P. Structural disorder in eukaryotes. *PLoS One* 7(4), e34687 (2012).
26. Xue B, Williams RW, Oldfield CJ, Goh GK-M, Dunker AK, Uversky VN. Viral disorder or disordered viruses: do viral proteins possess unique features? *Protein Pept. Lett.* 17, 932–951 (2010).
27. Xue B, Dunker AK, Uversky VN. Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J. Biomol. Struct. Dyn.* 30(2), 137–149 (2012).
28. Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, Tawfik DS. Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* 34, 53–59 (2009).

▪▪ **How intrinsically disordered viral proteins cope with high mutation rates in comparison with highly structured proteins.**

29. Uversky VN, Dunker AK. Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal. Chem.* 84, 2096–2104 (2012).

▪ **A review on the technical approaches used to structurally characterize IDPs.**

30. Receveur-Bréchot V, Bourhis J-M, Uversky V, Cannard B, Longhi S. Assessing protein disorder and induced folding. *Proteins* 62(1), 24–45 (2006).
31. *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation.* Uversky V, Longhi S. (Eds) John Wiley & Sons Inc, NJ, USA (2010).
32. Tompa P. Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21, 419–425 (2011).
33. Alves C, Cheng H, Roder H, Taylor J. Intrinsic disorder and oligomerization of the hepatitis delta virus antigen. *Virology* 407, 333–340 (2010).
34. Zuccola H, Rozzelle J, Lemon S, Erickson B, Hogle J. Structural basis of the oligomerization of hepatitis delta antigen. *Structure* 6, 821–830 (1998).
35. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochim. Biophys. Acta* 1804(6), 1231–1264 (2010).

▪ **Extensive review on intrinsic disorder covering the advantages granted by disorder, different functions involving IDPs, and disorder in disease.**

36. Johnson D, Xue B, Sickmeier M *et al.* High-throughput characterization of intrinsic disorder in proteins from the Protein Structure Initiative. *J. Struct. Biol.* doi:10.1016/j.jsb.2012.05.013 (2012) (Epub ahead of print).
37. Dyson HJ, Wright PE. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12, 54–60 (2002).
38. Koonin EV, Senkevich TG, Dolja VV. The ancient virus world and evolution of cells. *Biol. Direct* 1–27 (2006).
39. Drake JW, Charlesworth B, Charlesworth D, Crow J. Rates of spontaneous mutation. *Genetics* 148, 1667–1686 (1998).
40. Reanney DC. The evolution of RNA viruses. *Ann. Rev. Microbiol.* 36, 47–73 (1982).
41. Forterre P. The origin of viruses and their possible roles in major evolutionary transitions. *Virus Res.* 117, 5–16 (2006).
42. Goh G, Dunker A, Uversky V. Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics* 9, S4–S18 (2008).

▪▪ **Using disorder prediction to compare viral proteins and how intrinsic disorder may relate to virulence.**

43. Dyson HJ. Roles of intrinsic disorder in protein–nucleic acid interactions. *Mol. Biosyst.* 8, 97–104 (2012).
44. Xie H, Vucetic S, Iakoucheva LM *et al.* Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome Res.* 6, 1882–1898 (2007).
45. Vucetic S, Xie H, Iakoucheva LM *et al.* Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions. *J. Proteome Res.* 6, 1899–1916 (2007).
46. Xie H, Vucetic S, Iakoucheva LM *et al.* Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins. *J. Proteome Res.* 6, 1917–1932 (2007).
47. Goh GK-M, Dunker AK, Uversky VN. Protein intrinsic disorder and influenza virulence: the 1918 H1N1 and H5N1 viruses. *Virology J.* 6, 1–12 (2009).
48. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14–13–3 with their partners. *BMC Genomics* 9(Suppl. 1), S1-S13 (2008).
49. Uversky VN. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* 40, 1623–1634 (2011).
50. Taylor JM. Chapter 3. Replication of the hepatitis delta virus RNA genome. *Adv. Virus Res.* 74, 103–121 (2009).
51. Rizzetto M. Hepatitis D: thirty years after. *J. Hepatol.* 50, 1043–1050 (2009).
52. Greco-Stewart V, Pelchat M. Interaction of host cellular proteins with components of the hepatitis delta virus. *Viruses* 2, 189–212 (2010).
53. Casaca A, Fardilha M, da Cruz e Silva E, Cunha C. The heterogeneous ribonuclear protein C interacts with the hepatitis delta virus small antigen. *Virology J.* 8, 358–370 (2011).
54. Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR, Harris M. All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *J. Virol.* 84, 9267–9277 (2010).
55. Love RA, Brodsky O, Hickey MJ, Wells PA, Cronin CN. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol.* 83, 4395–4403 (2009).
56. Hanoulle X, Verdegem D, Badillo A, Wieruszeski J-M, Penin F, Lippens G. Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem. Biophys. Res. Commun.* 381, 634–638 (2009).
57. Hanoulle X, Badillo A, Verdegem D, Penin F, Lippens G. The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept. Lett.* 17, 1012–1018 (2010).
58. Macdonald A, Harris M. Hepatitis C virus NS5A: tales of a promiscuous protein. *J. Gen. Virol.* 85, 2485–2502 (2004).
59. Feuerstein S, Solyom Z, Aladag A *et al.* Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J. Mol. Biol.* 420, 310–323 (2012).
60. Longhi S, Receveur-Bréchot V, Karlin D *et al.* The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and unfolds upon binding to the C-terminal moiety of the phosphoprotein. *J. Biol. Chem.* 278, 18638–18648 (2003).
61. Curran J, Kolakofsky D. Replication of paramyxoviruses. *Adv. Virus Res.* 54, 403–422 (1999).
62. Bourhis J, Canard B, Longhi S. Structural disorder within the replicative complex of measles virus: functional implications. *Virology* 344, 94–110 (2006).
63. Habchi J, Longhi S. Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol. Biosyst.* 8, 69–81 (2012).
64. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in Paramyxovirinae N and P. *J. Gen. Virol.* 84, 3239–3252 (2003).
65. Rancurel C, Khosravi M, Dunker AK, Romero PR, Karlin D. Overlapping genes produce proteins with unusual sequence properties and offer insight into *de novo* protein creation. *J. Virol.* 83, 10719–10736 (2009).
66. Brown CJ, Takayama S, Campen AM *et al.* Evolutionary rate heterogeneity in proteins with long disordered regions. *J. Mol. Evol* 55, 104–110 (2002).
67. Sanjuan R, Forment J, Elena S. *In silico* predicted robustness of viroids RNA secondary structures. I. The effect of single mutations. *Mol. Biol. Evol.* 23, 1427–1436 (2006).
68. Elena SF, Carrasco P, Daròs J-A, Sanjuán R. Mechanisms of genetic robustness in RNA viruses. *EMBO Rep.* 7, 168–173 (2006).
69. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* 439, 344–348 (2005).
70. Farci P. New insights into the HCV quasispecies and compartmentalization. *Semin. Liver Dis.* 31(4), 356–374 (2011).
71. Smith BJ, McKimm-Breshkin JL, McDonald M, Fernley RT, Varghese JN, Colman PM. Structural studies of the resistance of influenza virus neuramindase to inhibitors. *J. Med. Chem.* 45, 2207–2212 (2002).
72. Kong L, Huang C-C, Coales SJ *et al.* Local conformational stability of HIV-1 gp120 in unliganded and CD4-bound states as defined by amide hydrogen/deuterium exchange. *J. Virol.* 84, 10311–10321 (2010).
73. Xue B, Mizianty MJ, Kurgan L. Protein intrinsic disorder as a flexible armor and a weapon of HIV-1. *Cell. Mol. Life Sci.* 69, 1211–1259 (2009).
74. Westhof E, Altschuh D, Moras D *et al.* Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* 311, 123–126 (1984).
75. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J. Mol. Recognit.* 18, 343–384 (2005).
76. White-Scharf ME, Potts BJ, Smith LM, Sokolowski KA, Rusche JR, Silver S. Broadly neutralizing monoclonal antibodies to the V3 region of HIV-1 can be elicited by peptide immunization. *Virology* 192, 197–206 (1993).
77. Weliky DP, Bennett AE, Zvi A, Anglister J, Steinbach PJ, Tycko R. Solid-state NMR evidence for an antibody-dependent conformation of the V3 loop of HIV-1 gp120. *Nat. Struc. Mol. Biol.* 6, 141–145 (1999).
78. Woodman CBJ, Collins SI, Young LS. The natural history of cervical HPV infection: unresolved issues. *Nat. Rev. Cancer* 7, 11–22 (2007).
79. Uversky VN, Roman A, Oldfield CJ, Dunker AK. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J. Proteome Res.* 5, 1829–1842 (2006).
80. Alonso LG, García-Alai MM, Nadra AD *et al.* High-risk (HPV16) human papillomavirus E7 oncoprotein is highly stable and extended, with conformational transitions that could explain its multiple cellular binding partners. *Biochemistry* 41, 10510–10518 (2002).
81. Heer A, Alonso LG, de Prat-Gay G. E6, the 50 amino acid product of the most abundant spliced transcript of the E6 oncoprotein in high-risk human papillomavirus, is a promiscuous folder and binder. *Biochemistry* 50, 1376–1383 (2011).
82. Dunker AK, Uversky VN. Drugs for 'protein clouds': targeting intrinsically disordered transcription factors. *Cur. Op. Pharma.* 10, 782–788 (2010).