

# Hepatitis B virus peptide inhibitors



# Organic & Biomolecular Chemistry

Accepted Manuscript






ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/obc

# Hepatitis B Virus Peptide Inhibitors: Solution Structures and Interactions with the Viral Capsid

Azira Muhamad[a#], Kok Lian Ho[b], Mohd. Basyaruddin Abd. Rahman[c], Bimo A. Tejo[d], Dušan Uhrín[e], Wen Siang Tan[*a,f]

[a]Institute of Bioscience, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[b]Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[c]Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[d]Department of Biotechnology, Surya University, Scientia Business Park, Tangerang 15810, Indonesia

[e]EastChem School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3JJ, United Kingdom

[f]Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia

[#]Current address: Malaysia Genome Institute, Kajang, Selangor, Malaysia

*Corresponding author

Email: wstan@upm.edu.my; Tel: +60389466715; Fax: +60389430913

**Abstract**

Hepatitis B virus (HBV) infection remains a health problem globally despite the availability of effective vaccines. In the assembly of the infectious virion, both the preS and S regions of HBV large surface antigen (L-HBsAg) interact synergistically with the viral core antigen (HBcAg). Peptides **preS** and **S** based on the L-HBsAg were demonstrated as potential inhibitors to block the viral assembly. Therefore, the objectives of this study were to determine the solution structures of these peptides and study their interactions with HBcAg. The solution structures of these peptides were solved using $^{1}H$, $^{13}C$, and $^{15}N$ NMR spectroscopy. Peptide **preS** has several structured region of β-turns at Ser7-Pro8-Pro9, Arg11-Thr12-Thr13 and Ser22-Thr23-Thr24 sequences whereas peptide **S** has only one structured region observed at Ser3-Asn4-His5. Both peptides contain bend-like structures surrounding the turn structures. Docking studies revealed both peptides interacted with the immunodominant region of HBcAg located at the tip of the viral capsid spikes. Saturation Transfer Difference (STD) NMR experiments identified several aromatic residues in peptides **preS** and **S** that interact with HBcAg. This study provides insights into the contact regions of L-HBsAg and HBcAg at atomic resolution which can be used to design antiviral agents that inhibit HBV morphogenesis.

**Introduction**

Hepatitis B virus (HBV) is one of the causative agents of liver cirrhosis and hepatocellular carcinoma. Currently, it is estimated that there are more than 360 million chronic HBV carriers worldwide[1], despite the existence of effective vaccines. The emergence of vaccine escape mutants[2] poses a threat to the worldwide vaccination program. An effective therapy for chronic hepatitis B is currently unattainable. Clinical treatments involving nucleoside analogues and interferon have limited success due to development of drug resistance variants and limited efficacy. The increased number of people co-infected with human immunodeficiency virus (HIV) and HBV, particularly in Africa and East Asia[3] has complicated the processes of drug and vaccine developments against HBV. Therefore, HBV infection remains a major worldwide health problem.

HBV infectious particle is enveloped by a lipid bilayer containing three related surface antigens (HBsAg) known as small (S), middle (M) and large (L)-HBsAg. These proteins are translated from a single open reading frame using three different in-frame start codons and a common stop codon; they therefore share the same 226 amino acids (the whole S-HBsAg) at their C-terminal ends. The M-HBsAg contains the S-HBSAg and an N-terminal extension of 55 amino acids known as the preS2 region. The L-HBsAg is composed of the entire M-HBsAg with an additional 108 or 119 residues (depending on the HBV subtype) called the preS1 region. Together, the preS1 and preS2 regions are known as the preS region. Both the preS and S regions interact with the viral capsid during the virus assembly[4]. The capsid is made of many copies of hepatitis core antigen (HBcAg), a single polypeptide containing 183 amino acids[5]. The C-terminal end of HBcAg is highly rich in positively charged amino acids which interact with the viral genome of 3.2 kb[6]. C-terminal truncated HBcAg

mutants without the positively charged residues still assembled into icosahedral capsids in *Escherichia coli* with triangulation number $T = 3$ and $T = 4$[7]. However, electron cryomicroscopic studies of native HBV virions revealed some differences between the capsids containing nucleic acids and the empty capsids produced in *E. coli*, particularly in the hydrophobic pocket located at the base of capsid spikes[8,9]. The differences in the capsid structures are believed to be important for the envelopment of the viral capsids.

Interactions between L-HBsAg and the viral capsid are critical for the assembly of the virion in liver cells[10]. The preS region of L-HBsAg is exposed on the cytosolic side of endoplasmic reticulum, signifying a crucial role for this region in virion assembly by forming docking sites for the viral capsid[11]. A 22-amino acid polypeptide in this region has been identified to play a critical role in HBV morphogenesis[12]. Deletion mutagenesis of the L-HBsAg revealed that a contiguous amino acid sequence from Arg92 to Ser113 in the preS region is essential for the interaction with the HBcAg[13]. In addition, a 25-amino acid polypeptide located from Pro56 to Phe80 in the S-HBsAg was demonstrated to bind efficiently to HBcAg[4]. Synthetic peptides correspond to these minimum contact regions in the L-HBsAg and S-HBsAg were shown to inhibit the binding of L-HBsAg with HBV capsid with $IC_{50}$ of 30 μM and 35 μM, respectively[14]. Up till now, there is no three dimensional structure of L-HBsAg, hence the focus of this study on two independent peptides in the preS and S regions would provide insights into the contact regions between L-HBsAg and HBcAg.

A synthetic peptide with a core motif LLGRMKG which was isolated from a phage displayed peptide library against HBV capsids formed by C-terminal truncated HBcAg was demonstrated to inhibit the binding of L-HBsAg to HBV capsid with an $IC_{50}$ of 78 μM[15]. This peptide was later shown to prevent the production of HBV in human hepatoma cells transfected with HBV DNA[16]. The binding of the peptide was located at the immunodominant region of HBcAg. By using cryo-electron microscopy and image reconstruction, Seitz *et al*[17] postulated that this peptide blocks the contact region between L-HBsAg and HBcAg in HBV virion. These observations suggest that the interaction sites of L-HBsAg and HBcAg could serve as a target for anti-HBV inhibitors.

In order to obtain more information about the contact regions of L-HBsAg and HBcAg, the solution structures of the two independent peptides corresponding to the preS and S regions were determined by NMR. Interactions of these peptides with HBV capsids formed by C-terminal truncated HBcAg were studied by using molecular docking and STD NMR. High resolution structural analysis of these peptides would furnish clues to the interactions of these peptides with HBcAg, which eventually provides information for designing therapeutic agents that block the assembly of HBV.

## Results and discussion

### Purification of hepatitis B virus core antigen (HBcAg)

HBcAg was expressed in *E. coli* and purified using a size exclusion chromatography column attached to an FPLC system (Akta Purified, GE Healthcare, USA). The fractions containing HBcAg were pooled and concentrated. The concentrated protein was analyzed by SDS-PAGE where an intense band of approximately 17 kDa corresponding to the HBcAg monomer was observed (Fig. S1). The purity of the HBcAg was 95%. Dynamic Light Scattering revealed that HBcAg self-assembled into capsids as described in Yoon *et al*[18].

### Chemical shift assignments and identification of secondary structure elements

Peptide **preS** has the amino acid sequence RQPTPLSPPLRTTHPQAMHWNSTTF while peptide **S** has the sequence PISNHSPTSCPPTCPGYRWMCLRRF. $^{1}$H resonance assignment of peptides **preS** and **S** was almost completed based on the analysis of 2D proton-proton and proton-carbon correlated NMR spectra. NH protons of all residues were identified except the N-terminal Arg1 of peptide **preS** and the Pro residues that lack an amide proton. The chemical shifts of the rest of the protons were assigned based on 2D TOCSY (Fig. 1), $^{1}$H-$^{13}$C HSQC and $^{1}$H-$^{13}$C HSQC-TOCSY spectra. Analysis of these spectra also yielded the resonance assignment of protonated heteroatoms (Tables S1 and S2).

All $^{15}$N chemical shifts for peptide **preS** were assigned, however for peptide **S**, the last five C-terminal amino acids could not be assigned because of their very similar $H^N$ chemical shifts (Fig. S2). In addition, some of the cross peaks of these residues were weak. Strong cross peaks were observed in the NOESY spectra corresponding to $H^{\alpha}$ of the preceding residues to Pro $H^{\delta}$. These indicated *trans* peptide bonds were present in both **preS** and **S** peptides.

Based on the CSI analysis, the structures of both peptides appear to be close to a random coil or a turn structure. Fig. 2A and 2B show that no four successive residues have negative or positive CSI values for either peptide, indicating that their structures could not be classified as adopting an α-helical or β-strand motifs. However, only positive CSI values were returned for peptide **preS**, which implicates that this peptide has a propensity for β-strand like structures. The same applies to the N-terminal part of peptide **S**, while its C-terminal part shows a tendency for α-helical structure. The analysis of the backbone $^{3}J(H^N\text{-}H^{\alpha})$ coupling constants confirmed the presence of a random coil structure as most of the $^{3}J(H^N\text{-}H^{\alpha})$ values for both peptides **preS** and **S** were in the range of 5-8 Hz (Table S3 and S4).

At 300 K, all $H^{\alpha}$ to $H^N$ NOE of cross peaks were clearly visible allowing sequential resonance assignment. Nearly all residues showed strong sequential NOE except for Cys21 of peptide **S** and the proline residues (Fig. 2). However, the sequence of these residues was assigned based on other NOEs, including side chain to side chain and $H^N$ to $H^N$ NOEs. Medium to strong $H^N$ to $H^N$ NOEs were

observed in three or more consecutive residues for peptide **preS**, these NOEs are indicative of a type I/I' β-turn. Nevertheless, long range NOEs were not observed for either peptide, indicating that both are unstructured.

**Structure determination using restrained molecular dynamics**

A total of 312 and 347 NOE restraints were used in the structure calculation of peptides **preS** and **S**, respectively. The $^{3}J(H^{N}\text{-}H^{\alpha})$ coupling constants obtained from the 2D $^{1}H\text{-}^{15}N$-HSQC spectra were also used as restraint. The structures obtained from the MD trajectories were consistent with the criteria chosen for structure calculations, thus these structures were further analyzed and the structural ensembles of peptides **preS** and **S** were constructed. An ensemble of **preS** and **S** consists of ten structures that were chosen from each of the top ten most populated clusters. The statistics of the two peptides structure ensembles are given in Table 1 with an overlay of the structures shown in Fig. 3A and 3B. The φ angles of the peptide ensembles were within the allowed limit from the ideal φ angles calculated based on the Karplus curve. Only the φ angles of Leu6 and Ser7 of peptide **preS** model with the most populated cluster deviated 6° and 4° more than the accepted limit, respectively (Table S3). While for peptide **S**, only Ile2 deviated 6° more than the allowed limit (Table S4). However, other conformations in **preS** and **S** ensembles were within the 30° limit. The Ramachandran plot suggests that majority of the residues are in the most favored regions for peptide **preS** (54%) while the residues of peptide **S** are generally in the additionally allowed region (43%). However, none of the residues are in the disallowed regions. Peptide **preS** sequences that were observed to have medium to strong $H^{N}$ to $H^{N}$ NOEs were further analyzed, and they were identified to contain a turn-like structure at residues Ser7-Pro8-Pro9 (turn A), Arg11-Thr12-Thr13 (turn B) and Ser22-Thr23-Thr24 (turn C) sequences. Based on the phi and psi angles for these amino acids, type I/I', type IV and type IV turns were determined for turn A, turn B and turn C, respectively[19]. The turns A and B were also categorized as potential to form type $3_{10}$ helical turn as characterized by the NOEs observed at $H^{\alpha}\text{-}H^{N}$ for sequence Pro9-Thr12. Furthermore, the turns B and C were in good agreement with the suggested $H^{N}\text{-}H^{N}$ NOE connectivities observed at these sequences. In addition, residues flanking the turn structures can be seen as bend-like, and these bends can be identified clearly in Fig. 3A and 3C. On the other hand, the ensemble of peptide **S** contained only one turn structure at Ser3-Asn4-His5, where a type IV turn was determined. $H^{N}\text{-}H^{N}$ NOE connectivities were not observed at this sequence, however medium strength $H^{\alpha}\text{-}H^{N}$ NOEs were found to connect between Ile2-Ser3 and Asn4-His5. The sequences from Ser6-Pro15 and Trp19-Arg23 were classified as bend-like structures since they did not fit into a turn-like structure (Fig. 3B and 3D).

To date, the three dimensional structure of HBsAg, which is the main component of HBV envelope, at atomic resolution has not been resolved. On the other hand, the three dimensional structures of HBV capsids have been determined using cryo-electron microscopy and X-ray crystallography[9,20,21,22]. The interaction between L-HBsAg and HBcAg, the monomer of HBV capsid,

is mediated by two distinct binding sites which are located in the preS and S regions. Therefore, the three dimensional structures of peptides **preS** and **S**, which correspond to the docking sites in L-HBsAg and their association with HBcAg capsid may provide insights into the viral morphogenesis and development of therapeutic agents that block this process.

In previous experiments, a comprehensive multi-dimensional NMR study revealed that the preS1 domain of L-HBsAg belongs to the family of so-called intrinsically unstructured or intrinsically disordered proteins (IDPs). The preS1 domain was shown to contain several local pre-structured motifs that are likely to be essential for viral infection but it could not form a stable tertiary structure[23]. The discovery of IDPs has shifted the classical hypothesis towards a more dynamic paradigm where a protein can be fully functional even in the absence of a stable, folded structure. Particularly when several studies found that many short antigenic peptides undergo a transition from an unstructured state to form a β-turn or an α-helix upon binding[24,25].

In this study, the solution structures of peptides **preS** and **S** were determined by NMR spectroscopy at pH 4. NMR structural studies of proteins and peptides are typically performed at slightly acidic conditions (pH 3-6), which enable a better observation of resonances from labile protons for instance the backbone amide NHs by minimizing their exchange rates[26]. Interestingly, in our study, both peptides **preS** and **S** adopt a random coil 'turn-like' structure. Peptide **preS** containing 17 C-terminal residues of preS1 region and 8 residues of N-terminal end of preS2 region, was observed to have three β-turn structures. Peptide **S** consisting of amino acids 56-80 of S region was found to contain one β-turn.

**STD NMR**

HBcAg monomer self-assembles into capsids or virus-like particles (VLPs) is a good candidate for STD NMR experiments. STD has no upper limit on the size of the receptor, and favors heavier receptor molecules at stronger magnetic fields[27]. Due to the typically large molecular weight, spin diffusion is able to spread the saturation throughout the receptor spectrum and the broad protein background signals can be eliminated. Peptide resonances appear in an STD spectrum as a result of the transfer of magnetization from the large-molecular-weight protein to a bound small peptide; efficient transfer is possible only when the peptide and macromolecular protons are spatially close, it also depends on the exchange regime[28]. In the present study, STD experiments were acquired using the parameters described by Rademacher *et al*[29]. This technique was applied to study the calicivirus-ligand interactions.

To confirm that the STD spectra peaks were not affected by a systematic error, STD experiments using the peptides in the absence of HBcAg were also performed and resulted in no signals detected. All STD experiments were carried out with buffers at pH 6 due to the stability of the HBcAg capsid. In order to identify the $^{1}$H resonances in the STD spectra, $^{1}$H-$^{1}$H 2D TOCSY experiments were performed at pH 6 for both peptides **preS** and **S**. Almost complete assignment of $^{1}$H

resonances of the free peptides was achieved at pH 6, furthermore the $^{1}$H resonances were sufficiently resolved to allow unambiguous assignment of the STD spectra. However, quantitative analysis of the STD spectra could not be carried out because some of the signals arising from the STD experiments were broad and overlapping.

The STD spectrum of peptide **preS** with HBcAg capsid revealed significant signal enhancements for several $^{1}$H of the peptide. The aromatic side chain of residues His14, His19, Trp20 and Phe25 showed significant STD effects. Furthermore, Arg1H$_{\varepsilon}$, Met18H$_{\varepsilon}$, Leu6 and Leu10 H$_{\delta}$, and all Thr H$_{\gamma}$ and H$_{\beta}$ similarly showed some signal enhancements (Fig. 4A). The STD spectrum of peptide **S** in the presence of HBcAg capsid also showed STD effects. However, the signals were broader than those of **preS** peptide and overlapping, therefore unambiguous assignment of resonances was challenging. Nevertheless, a few resonances showing significant STD effects were assigned (Fig. 4B). These $^{1}$H included aromatic side chains of Tyr17, Trp19 and Phe25. In addition, Leu22H$_{\delta}$, Ile2H$_{\gamma}$ and H$_{\delta}$, Met20H$_{\varepsilon}$ and Arg24Hη were in close contact with the HBcAg capsid.

STD spectra obtained from peptides **preS** and **S** in the presence of HBcAg capsids indicate that both peptides bind to the capsid. Residues Arg1, His14, His19, Trp20, Phe25 of peptide **preS** and residues Ile2, Tyr17, Trp19, Met20, Leu22, Arg24, Phe25 of peptide **S**, showed signal enhancements. These residues appear to play important roles in their interactions with HBcAg capsid, and inhibit L-HBsAg-HBcAg interactions during the assembly of the infectious virions[4,12]. Arg1 of peptide **preS** which corresponds to Arg92 of the preS1 domain showed a significant STD signal enhancement. Amino acid mutations revealed that Arg92 plays a major role in L-HBsAg-HBcAg interactions[13].

**Docking studies**

Peptides **preS** and **S** were docked to the immunodominant region of HBcAg which is located at the tip of HBcAg capsid spikes. Only one peptide was considered for each docking experiment, even though both peptides **preS** and **S** may interact synergistically with the HBV nucleocapsid. The docking of peptides **preS** and **S** was centered on the tip of HBV capsid spikes formed by HBcAg dimer because previous studies showed that the binding of antiviral peptides at the tip inhibited L-HBsAg-HBcAg interactions[16,30]. Furthermore, a cryo-EM map of HBV infectious virions showed that the tips of the capsid spikes lay close to the HBsAg[9,17]. Three structures that formed the ensemble of structures for both peptides **preS** and **S** were docked to HBcAg dimers with the PDB code 3KXS[31] and 1QGT[21] using HADDOCK. Two 3D structures of HBcAg were chosen in order to compare docking of the structure of dimers that do not assemble into a capsid (3KXS) and the structure of dimers that form a capsid (1QGT). The three representative structures of the peptides were obtained from the clusters with the highest number of trajectories. Docking of the first model of peptide **preS** ensemble to HBcAg AB dimer of PDB 3KXS produced 6 clusters while peptide **S** produced 11 clusters. A Z-score of -1.4 and -2.3 was obtained for peptides **preS** and **S,** respectively, from the best HADDOCK cluster. Docking of the first model of peptide **preS** to CD dimer produced 10 clusters and

a Z-score of -1.6 while peptide **S** produced 6 clusters and a Z-score of -1.3 (Fig. 5). A Z-score of -1.6 and -1.7 was obtained for peptides **preS** and **S**, respectively, when docked to EF dimer. The peptides appear to interact with both monomers of each dimer, AB, CD and EF at the tips of HBcAg capsid spikes.

Hydrogen bonds and hydrophobic contacts between the peptides and HBcAg were analyzed by Ligplot. Analysis of all binding modes of peptide **preS** docked to the three dimers of PDB 3KXS showed that residues Arg1, Gln2, Leu6, Thr13, Met18, Thr24 and Phe25 contributed to the interaction with HBcAg. In addition, both His14 and His19 were involved in hydrogen bonding. Similarly, a number of aromatic and polar amino acids of peptide **S** were found to be hydrogen bonded to the HBcAg. These residues included His5, Ser6, Ser9, Trp19 and Phe25. Residues Arg18, Arg23 and Arg24 were found to contribute to the hydrogen bonding interaction as well. Several binding modes produced by HADDOCK for peptide **S** showed that residues Ile2, Met20 and Leu22 were also hydrogen bonded to HBcAg.

Docking of the first model of peptide **preS** ensemble to HBcAg AB dimer of PDB 1QGT produced 6 clusters while peptide **S** produced 12 clusters. A Z-score of -1.4 and -1.7 was obtained for peptides **preS** and **S,** respectively. Ligplot analysis showed that almost similar residues of peptides **preS** and **S** were involved in hydrogen bonding and hydrophobic contacts with the HBcAg dimer of 1QGT when compared to the docking of the peptides with HBcAg dimers of PDB 3KXS. These residues include Arg1, Leu6, Thr12, Met18, His19, Trp20 and Thr23 for peptide **preS**, and 2Ile, His5, Thr8, Ser9, Cys10, Tyr17, Arg18, Trp19 and Cys21 for peptide **S**.

The immunodominant region (amino acids 74-84) of HBcAg seems to play an important role in the interaction of the peptides with HBcAg. Amino acids Asn75, Glu77, Asp78 and Ser81 in the immunodominant region were involved in hydrogen bonding with both peptides in all of the complexes analyzed using both the HBcAg dimers of PDB 3KXS and 1QGT. Whereas, residues Ala80 and Asp83 from HBcAg dimers AB and EF of PDB 3KXS interact with peptide **preS**, while residue Thr74 from dimers AB and CD was found to interact with peptide **S** by hydrogen bond interaction. Residues Leu76, Glu77, Asp78, Pro79, Ala80, Ser81 and Leu84 were involved in hydrophobic interaction with both peptides **preS** and **S** in the majority of the binding modes analyzed. In another study, Ponsel and Bruss[32] demonstrated that mutations of 11 amino acids (Ser17, Phe18, Leu60, Leu95, Lys96, Phe122, Ile126, Arg127, Asn136, Ala137, and Ile139) which are located around the base of the capsid spike and in a small area close to the pores in the capsid, did not inhibit nucleocapsid formation but blocked particle envelopment and virion formation. These residues were therefore postulated to interact with HBsAg during virion morphogenesis. By using an *in vitro* assay, Tan *et al*[13] revealed that at least two distinct elements within a large area of L-HBsAg interacted synergistically with HBcAg. One element is located between residues 24-191 and the other between residues 191-322 of L-HBsAg. In the present study, peptide **preS** which lies in the first element and peptide **S** which is located in the N-terminal end of the second element (residues 191-263) were found

to dock at the tip of capsid spike. However, binding of the C-terminal end of the second element (residues 263-322) with HBcAg was not investigated in the present study. The residues in the C-terminal end of the second element may interact with other regions on the surface of the capsid and mediate the assembly of the virion.

Docking studies showed that residues Arg1, Gln2, Leu6, Thr13, Met18, Thr24 and Phe25 of peptide **preS,** which correspond to residues Arg92, Gln93, Leu97 and Thr104 of the preS1 domain and Met1, Thr7 and Phe8 of the preS2 domain interacted with residues Asn75, Glu77, Asp78 and Ser81 of HBcAg. Whereas residues His5, Ser6, Ser9, Arg18, Trp19, Arg23, Arg24 and Phe25 of peptide **S** which correspond to residues His60, Ser61, Ser64, Arg73, Trp74, Arg78, Arg79 and Phe80 of S-HBsAg, were found to interact with the HBcAg capsid spike. These findings are in good agreement with the STD results where the residues that form the hydrogen bonding network are spatially close to the HBcAg based on the STD spectra obtained. The similar residues are Arg1, Thr13, Met18, Thr24 and Phe25 of peptide **preS,** and residues Ile2, Trp19, Met20, Leu22, Arg24 and Phe25 of peptide **S**. In general, most of the residues in peptides **preS** and **S** interacted at the tip of capsid spikes. This is in good agreement with earlier findings showing different segments and protrusions of HBsAg interacted at the tips of capsid spikes formed by two conformations of HBcAg dimers[17].

## Conclusions

The conformations of peptides **preS** and **S** can be categorized as disordered structures, stabilized by the presence of β-turn motifs. The adaptability of the disordered structures may play an important role in L-HBsAg as a dual-functional protein that involves in virion assembly and entry into host cells. STD NMR experiments of the peptides in the presence of HBcAg capsid identified residues in close contact with the tips of capsid spikes. The STD results are supported by the computational model of the complex formed between the solution structures of the peptides with the HBcAg dimer. Favorable interactions uncovered by the docking studies include hydrogen bonds and hydrophobic contacts observed between the peptides and the tips of capsid spikes provide useful information for designing inhibitors that block HBV morphogenesis.

## Experimental

### Purification of hepatitis B virus core antigen

*E. coli* strain harboring the plasmid pR1-11E which encodes HBcAg (residues 3-148) was grown in Luria Bertani broth as previously described[30]. The cells were harvested and lysed by sonication. HBcAg in lysate was precipitated with ammonium sulfate and the sample was dialyzed with two changes in 50 mM Tris-HCl (pH 8), 100 mM NaCl buffer. The dialyzed HBcAg was purified by size exclusion chromatography (SEC)[18] using Sephacryl-500 (GE Healthcare) packed in

an XK 16/100 column (Amersham, USA). Fractions containing the HBcAg were collected, pooled and concentrated using 300 kDa cut-off Vivaspin (Sartorius Stedim Biotech, Germany) concentrator by centrifugation. The concentrated proteins were analyzed on SDS-PAGE[33] and the final concentration of the purified HBcAg was determined using the Bradford assay[34]. The formation of HBcAg capsid was confirmed by dynamic light scattering (Dyna Pro-801$^{TM,}$ Protein Solution Ltd. High Wycombe, UK) as described by Yoon *et al*[18].

**Peptide synthesis**

Two peptides, **preS** and **S**, derived from the sequence of HBsAg (*adyw* subtype) were synthesized using the Fmoc solid phase peptide chemistry (Genemed Synthesis Inc., USA). Peptide **preS** has the amino acid sequence RQPTPLSPPLRTTHPQAMHWNSTTF (17 residues of the C-terminal preS1 region + 8 residues of the N-terminal preS2 region) while peptide **S** has the sequence PISNHSPTSCPPTCPGYRWMCLRRF (residues 56-80 of the S region of HBsAg). The peptides were analyzed by HPLC and mass spectrometry.

**NMR spectroscopy**

All NMR experiments were performed at 300 K using a Bruker Avance III 800 MHz spectrometer equipped with a TCI cryoprobe. The peptides were dissolved in 550 μL of sodium phosphate (20 mM) in 90% $H_2O$/10% $D_2O$, and the pH was adjusted from 4.0 to 6.0 with the peptide concentrations from 3 mM to 6 mM. Proton chemical shifts were referenced relative to the $H_2O$ offset frequency (4.705 ppm at 27 °C) and heteronuclear chemical shifts were referenced relative to the proton frequency using the method of Wishart *et al*[35]. $^1$H-$^1$H 2D NMR TOCSY and NOESY experiments were carried out using the DPFGSE for water-suppression[36]. TOCSY spectra were acquired using mixing times of 20 ms and 80 ms, while NOESY spectra were recorded using mixing times of 150 ms. The spectral width was typically 9500 Hz with acquisition time of 0.21 s. 2D $^1$H-$^{13}$C-HSQC and $^1$H-$^{13}$C HSQC-TOCSY spectra were also collected to assist in $^1$H chemical shift assignments. Acquisition data were sampled with 2048 x 1536 complex points in the $t_2$ and $t_1$ dimensions. In addition, 2D $^1$H-$^{15}$N-HSQC spectrum was acquired with 8192 x 256 complex points in the $t_2$ and $t_1$ dimensions to assist in backbone torsion angle calculations. The torsions were derived from the $^3J(H^N\text{-}H^\alpha)$ coupling constants extracted from the 2D $^1$H-$^{15}$N-HSQC spectrum acquired without the refocusing of signals and $^{15}$N decoupling. Dihedral angles (φ) were calculated using the Karplus equation[37].

**Spectral assignment**

NMR data were processed in the TopSpin3.2 Bruker BioSpin GmbH, Germany and analyzed using the CCPNmr Analysis 2.3[38]. $^1$H, $^{13}$C, and $^{15}$N chemical shifts of the backbone and side-chains nuclei were obtained by analyzing the 2D spectra. The assignment of the chemical shifts for peptides

**preS** and **S** was performed using the spectra obtained with peptides in pH 4 and 6. A combination of 2D $^{1}$H-$^{13}$C HSQC and $^{1}$H-$^{13}$C HSQC-TOCSY spectra was required to assign unambiguously all resonances. The 2D $^{1}$H-$^{13}$C HSQC-TOCSY spectrum was particularly useful in confirming the chemical shifts of $^{1}$H of the peptides. All aliphatic $^{1}$H resonances were assigned unambiguously since most of the cross peaks were well resolved. Sequential assignment of the amino acids of both peptides was performed using the NOESY spectra. The fingerprint region of the spectra was particularly helpful since NOEs between NH(i) and Hα(i-1) were easily assigned. In addition, the assignment of the NOESY spectrum provided distance restraints information based on the volume of the cross peaks calculated by CCPNmr Analysis and each NOE cross peak was examined manually. The CSI values were also calculated by CCPNmr Analysis using the difference of the recorded chemical shift to random coil chemical shift of Cα, Cβ, C' and Hα atoms[39,40]. These secondary chemical shifts were combined to give the CSI values indicating the type of secondary structure, where -1 indicates alpha-helix and +1 indicates beta-strand.

**Structure calculations**

NMR-restraint molecular dynamics (MD) simulations were performed using the AMBER 12[41] program by incorporating NOE restraints (Tables S5 and S6) and backbone $^{3}J(H^{N}\text{-}H^{\alpha})$ coupling constants (Tables S3 and S4). The input files of the restraints were obtained from the assignment of the spectra for peptides **preS** and **S** in pH 4. NOE cross-peaks were converted into distances based on an r-6 dependence and calibration was performed using the default setting of an average peak intensity that corresponded to a distance of 3.2 Å. The initial starting structure was randomly drawn using the Avogadro software[42] and heated to 600 K then cooled down to 300 K for 20 ps *in vacuo*, in order to relax the structure. The structures were then soaked in a waterbox and were first minimized and heated to 300 K, followed by a MD simulation for 500 ps. A total of 5000 structures were generated. The stability of the simulation was monitored by observing the backbone RMSD values. The clustering facility[43] in the Chimera software was initiated during 100 ps until 300 ps of the MD simulation. The clusters were obtained based on the backbone atoms of the peptide in a pairwise manner. Top ten clusters that contained the highest number of trajectories were further analyzed by selecting the trajectories with the following conditions (i) the structure has $\varphi$ angles within 30$^{o}$ of the $\varphi$-values calculated from $^{3}J(H^{N}\text{-}H^{\alpha})$, and (ii) the structure has an interproton distance error < 0.5 Å compared to the upper boundaries of distances derived from the NOE restraints[44]. One conformation from each of the top ten clusters that suited the above criteria was then selected to form a structural ensemble and analyzed using the PROCHECK-NMR[45]. The PyMOL (Version 1.3 Schrödinger, LLC) and the Discovery Studio Visualizer 3.5 (Accelrys) were used to view the structural ensembles. PyMOL was also utilized to measure the distances of the protons included in the NOE restraints and the dihedral angles of the peptide structure ensembles.

**Saturation transfer difference (STD) experiments**

All STD spectra were recorded at 300 K using a Bruker Avance III 800 MHz spectrometer equipped with a TCI cryprobe. The solvent used was 20 mM sodium phosphate (pH 6), 150 mM NaCl. The peptides and the HBcAg capsid samples were prepared in 9:1 $H_2O$:$D_2O$ buffer using 200 µM peptides, while the HBcAg concentration ranged from 16 nM to 32 nM. The final concentration of the HBcAg capsid was adjusted to 32 nM. A 6.25 molar excess of peptides was used for all STD experiments when compared to the monomeric HBcAg. Typically, STD NMR experiments are performed in 100% $D_2O$ to achieve sufficient magnetization transfer from protein to ligand. However, in this study a ratio of 9:1 $H_2O$:$D_2O$ buffer at pH 6 was used, thus suppression of residual water signal was done using the Watergate 3-9-19. 1D STD spectra were acquired by subtracting an on-resonance FID, with selective saturation of protein resonance at -4 ppm, from an off-resonance FID, with saturation at -300 ppm. Subtraction of the two spectra by phase cycling leads to the difference spectrum that contains signals arising from the saturation transfer. The acquisition and relaxation times were 0.68 and 6 s, respectively. The saturation was achieved by five Gaussian pulses with duration of 80 ms. The total number of scans was fixed at 32 with 4 dummy scans and a typical spectra width of 16 ppm.

**Molecular docking**

Docking of the HBcAg dimer with peptides **preS** and **S** was accomplished using HADDOCK[46]. The WeNMR grid-enabled server[47] was utilized to perform the docking simulations. Crystal structures of HBcAg trimer of dimer with the PDB code 3KXS[31] and a dimer of 1QGT[21] were docked with the peptides. Three dimers, AB, CD and EF of PDB 3KXS structure and AB dimer of PDB 1QGT were each employed for docking with the peptides. The active residues of the protein were fixed around the spike of the HBcAg (aa 72 – 88) dimer whereas all of the amino acids of the peptides were selected as active residues. The docked complex structure produced by the HADDOCK was based on the Z-score, therefore the cluster with the lowest Z-score value was chosen as the most reliable docked structure of the peptides to the protein. The LigPlot program[48] was employed to analyze the docking results.

**Acknowledgments**

Azira Muhamad is supported by the National Science Fellowship of Malaysia. We would like to thank Yoon Kam Yee and Gan Bee Koon for their help in the preparation of HBcAg capsids. This study was supported by the Ministry of Higher Education, Malaysia.

**Notes and references**


1 C. W. Shepard, E. P. Simard, L. Finelli, A. E. Fiore and B. P. Bell, *Epidemiol. Rev.*, 2006, **28**, 112-125.

2 A. J. Zuckerman, *Lancet*, 2000, **355**, 1382-1384.

3 W. S. Tan and K. L. Ho, *World J. Gastroenterol.*, 2014, **20**, 11650-11670.

4 F. Poisson, A. Severac, C. Houioux, A. Goudeau and P. Roingeard, *Virology*, 1997, **228**, 115-120.

5 W. S. Tan, M. R. Dyson and K. Murray, *Biol. Chem.*, 2003, **384**, 363-371.

6 D. Ganem, J. R. Pollack and J. Tavis, *Infect. Agents Dis.*, 1994, **3**, 85-93.

7 A. Zlotnick, N. Cheng, S. J. Stahl, J. F. Conway, A. C. Steven and P. T. Wingfield, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 9556–9561.

8 A. M. Roseman, J. A. Berriman, S. A. Wynne, P. J. G. Butler and R. A. Crowther, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15821–15826.

9 K. A. Dryden, S. F. Wieland, C. Whitten-Bauer, J. L. Gerin, F. V. Chisari and M. Yeager, *Mol. Cell*, 2006, **22**, 843-850.

10 V. Bruss and R. Thomssen, *J. Virol.*, 1994, **68**, 1643-1650.

11 V. Bruss and D. Ganem, *Proc. Natl. Acad. Sci. U. S. A.*, 1991, **88**, 1059-1063.

12 V. Bruss, *Virus Res.*, 2004, **106**, 199-209.

13 W. S. Tan, M. R. Dyson and K. Murray, *J. Mol. Biol.*, 1999, **286**, 797-808.

14 W. S. Tan, *J. Gen. Appl. Microbiol.*, 2002, **48**, 103-107.

15 M. R. Dyson and K. Murray, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 2194-2198.

16 B. Böttcher, N. Tsuji, H. Takahashi, M. R. Dyson, S. Zhao, R. A. Crowther and K. Murray, *EMBO J.*, 1998, **17**, 6839-6845.

17 S. Seitz, S. Urban, C. Antoni and B. Bottcher, *EMBO J.*, 2007, **26**, 4160-4167.

18 K. Y. Yoon, W. S. Tan, B. T. Tey, K. W. Lee and K. L. Ho, *Electrophoresis*, 2013, **34**, 244-253.

19 E. G. Hutchinson and J. M. Thornton, *Protein Sci*, 1994, **3**, 2207-2216.

20 B. Böttcher, S. A. Wynne and R. A. Crowther, *Nature*, 1997, **386**, 88-91.

21 S. A. Wynne, R. A. Crowther and A. G. Leslie, *Mol. Cell*, 1999, **3**, 771-780.

22 W. S. Tan, I. W. McNae, K. L. Ho and M. D. Walkinshaw, *Acta Crystallogr. Sect. F*, 2007, **63**, 642-647.

23 S. W. Chi, D. H. Kim, S. H. Lee, I. Chang and K. H. Han, *Protein Sci.*, 2007, **16**, 2108-2117.

24 D. T. Nair, K. Singh, Z. Siddiqui, B. P. Nayak, K. V. S. Rao and D. M. Salunke, *J. Immunol.*, 2002, **168,** 2371-2382.

25 R. L. Stanfield and I. A. Wilson, *Curr. Opin. Struc. Biol*., 1995, **5**, 103-113.

26 K. Wüthrich, *NMR of proteins and nucleic acids*, 1986, John Wiley & Sons, Inc. New York.

27 M. Mayer and B. Meyer, *Angew. Chem. Int. Ed.*, 1999, **38**, 1784-1788.

28 B. Meyer and T. Peters, *Angew. Chem. Int. Ed.*, 2003, **42**, 864-890.

29 C. Rademacher, N. R. Krishna, M. Palcic, F. Parra and T. Peters, *J. Am. Chem. Soc*., 2008, **130**, 3669-3675.

30 K. F. Tang, M. P. Abdullah, K. Yusoff and W. S. Tan, *J. Med. Chem.*, 2007, **50**, 5620-5626.

31 C. Packianathan, S. P. Katen, C. E. Dann III and A. Zlotnick, *J. Virol*., 2010, **84**, 1607-1615.

32 D. Ponsel and V. Bruss, *J. Virol*., 2003, **77**, 416-422.

33 U. K. Laemmli, *Nature*, 1970, **227**, 680-685.

34 M. Bradford, *Anal. Biochem*., 1976, **72**, 248-254.

35 D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes, *J. Biomol. NMR*, 1995, **6**, 135-140.

36 T. L. Hwang and A. J. Shaka, *J. Magn. Reson. Ser. A*, 1995, **112**, 275-279.

37 R. Brüschweiler and D. A. Case, *J. Am. Chem. Soc*., 1994, **116**: 11199-11200.

38 W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Foqh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides and E. D. Laue, *Proteins*, 2005, **59**, 687-696.

39 D. S. Wishart and B. D. Sykes, *J. Biomol. NMR*, 1994, **4**,171-180.

40 D. S. Wishart and B. D. Sykes and F. M. Richards, *Biochemistry*, 1992, **31**, 1647-1651.

41 D. A. Case, T. E. Cheatham 3rd., T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem*., 2005, **26**, 1668-1688.

42 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *J. Cheminform.*, 2012, **4**, 17.

43 L. A. Kelley, S. P. Gardner and M. J. Sutcliffe, *Protein Eng*., 1996, **9**, 1063-1065.

44 A. Muhamad, K. L. Ho, M. B. A. Rahman, D. Uhrín and W. S. Tan, *Chem. Biol. Drug Des*., 2013, **81**, 784-794.

45 R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J. Biomol. NMR*, 1996, **8**, 477-486.

46 S. J. De Vries, M. van Dijk and A. M. J. J. Bonvin, *Nature Protoc*., 2010, **5**, 883-897.

47 Wassenaar T, van Dijk M, Loureiro-Ferreira N, van der Schot G, de Vries S, *et al*., *J. Grid Comp*., 2012, **10**, 743-767.

48 A. C. Wallace, R. A. Laskowski and J. M. Thornton, *Protein Eng.*, 1995, **8**, 127-134.

**Figures**

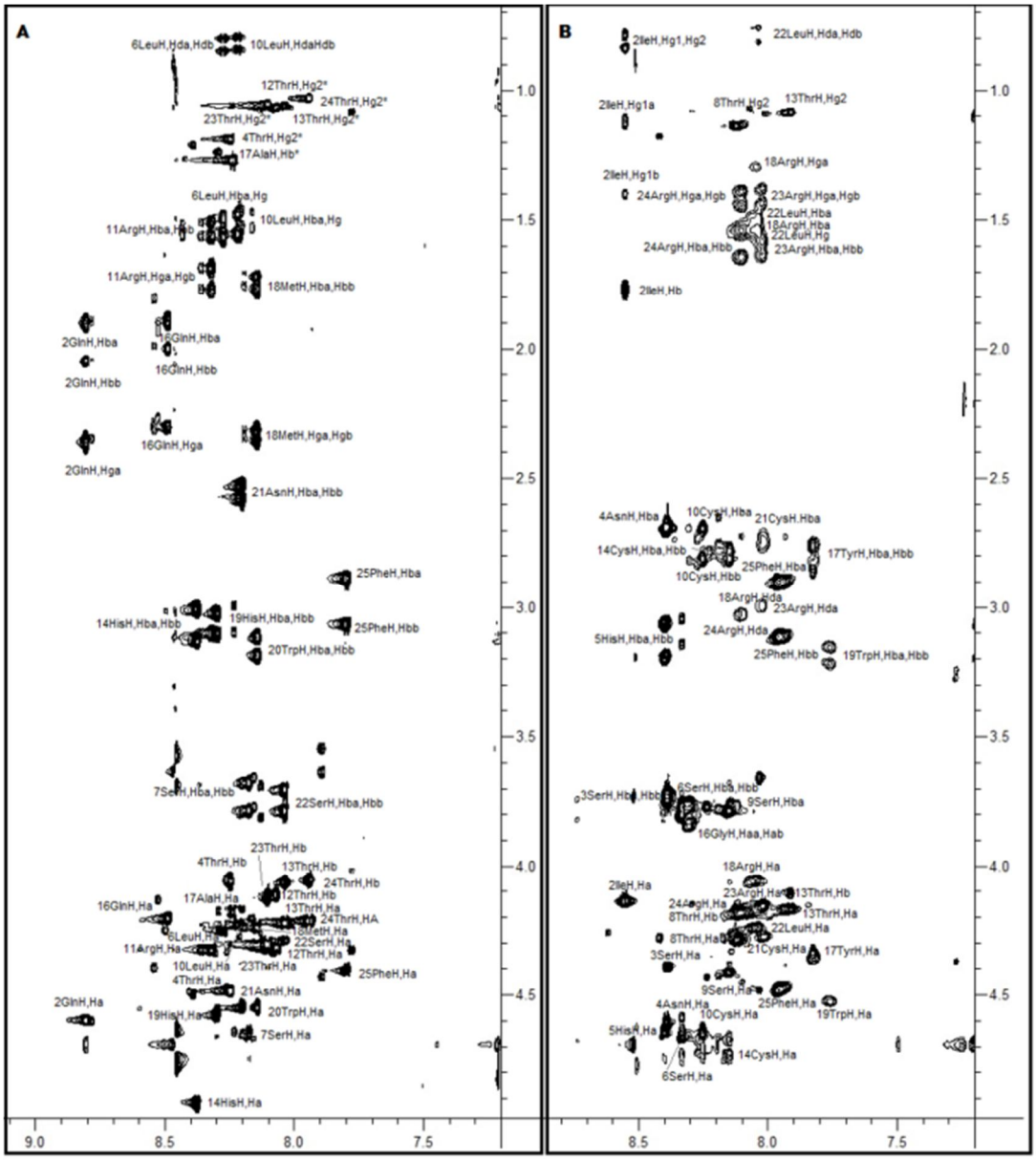


**Fig. 1** 2D $^{1}$H-$^{1}$H TOCSY spectra of peptides **preS** and **S**. The NH regions of **preS** (A) and **S** (B) of the spectra are shown. All NH protons were assigned for both peptides except the N-terminal NH and Pro residues.

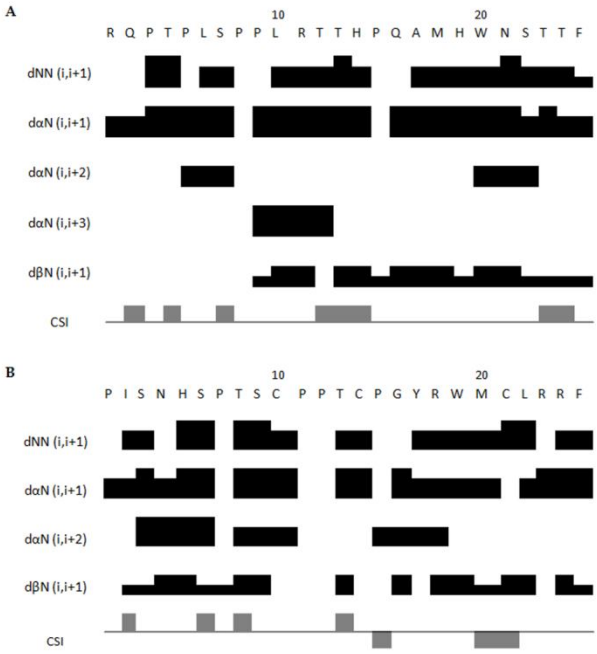


**Fig. 2** A summary of the inter-residue NOEs and chemical shift index (CSI). The CSI values for peptides **preS** (A) and **S** (B) shown as thin straight lines were equal to 0 or were not calculated. The CSI is based on the analysis of $^{1}H_{\alpha}$, $^{13}C_{\alpha}$ and $^{13}C_{\beta}$ chemical shifts. Thickness of the bars for the NOEs represents relative strength (strong, medium, weak).

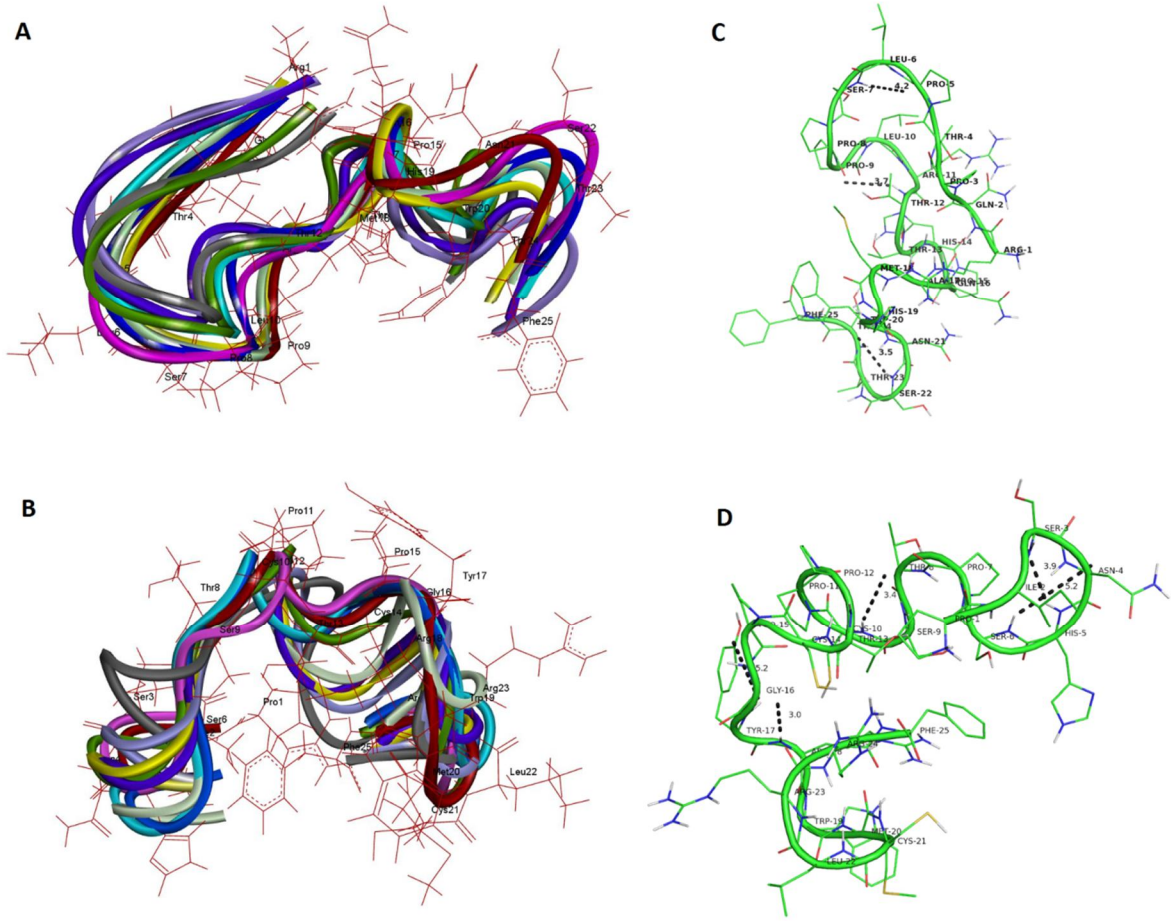


**Fig. 3** Overlay of ten NMR restrained MD structures and the representative structures of peptides. Backbone of the ten structure ensembles of peptides **preS** (A) and **S** (B) are rendered as ribbons. Representative structures of peptides **preS** (C) and **S** (D) rendered as sticks and the backbone as cartoons. The dotted lines indicate the medium range NOEs and their distance.

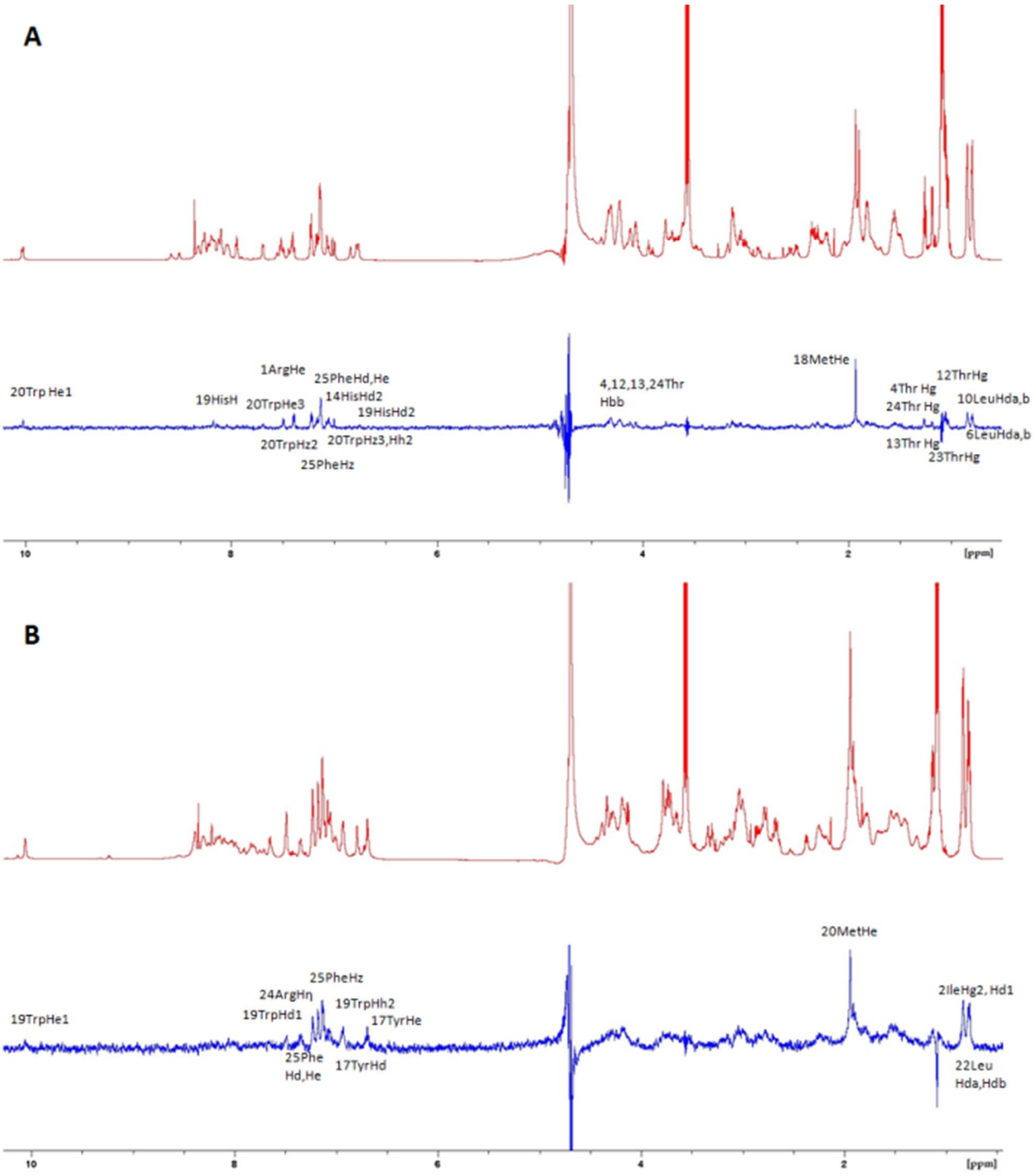


**Fig. 4** STD spectra of peptides preS and S in the presence of HBcAg capsid. The 1D $^{1}$H reference spectra of peptides **preS** (A) and **S** (B) in the presence of HBcAg capsid are shown in red. The STD spectra of the peptides in the presence of HBcAg capsid with on-resonance irradiation at -4 ppm are shown in blue. $^{1}$H resonances arising from the STD experiments are labeled.

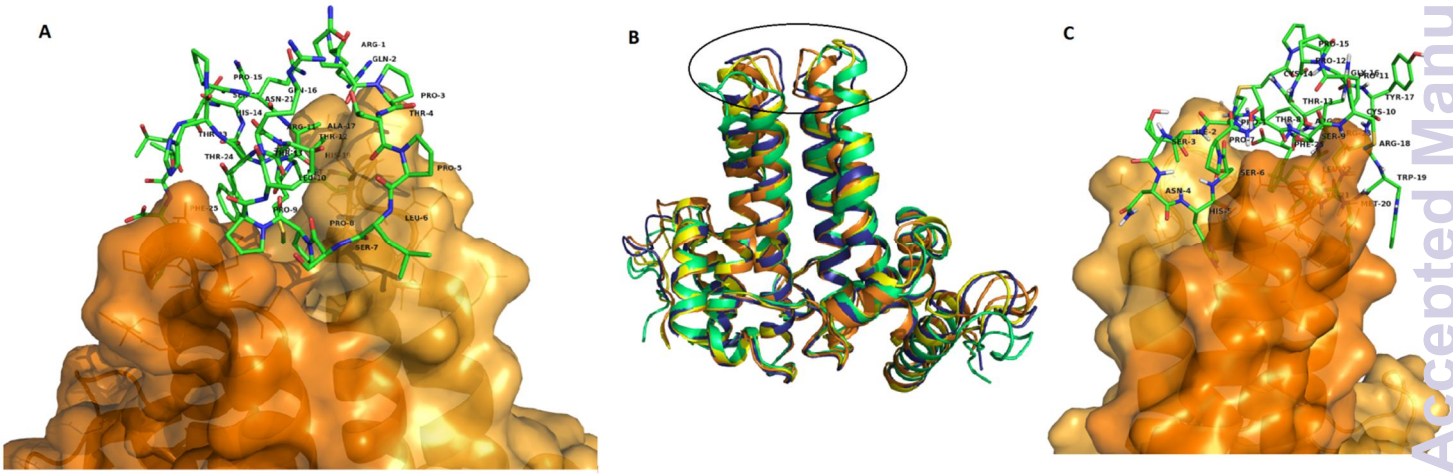


**Fig. 5** Peptides in complex with HBcAg. A close-up of peptides **preS** (A) and **S** (C) docked at the tip of the HBcAg capsid spike. The complex generated by HADDOCK with the lowest Z-score cluster was chosen as representation of docked structures. Peptides and the immunodominant region of HBcAg are rendered as sticks while transparent surface with secondary structure element rendered as ribbon indicating the HBcAg CD dimer of PDB 3KXS. HBcAg monomer C is in orange whereas monomer D is in light orange. Figure 5B shows the structures of HBcAg AB (blue), CD (yellow) and EF (green) dimers of PDB 3KXS, and HBcAg AB (orange) dimer of PDB 1QGT as secondary structure elements. The circled region is the immunodominant region of HBcAg.

**Tables**

**Table 1.** Structural statistics of the simulated peptide ensembles of **preS** and **S**

| Parameter | **preS** | **S** |
|---|---|---|
| *Distance restraints* | | |
| All | 312 | 348 |
| Intra-residue | 199 | 240 |
| Inter-residue | 113 | 108 |
| Sequential (*i,i+1*) | 96 | 87 |
| Long range (*i>=i+2*) | 13 | 19 |
| *RMSD* | | |
| Average heavy atom RMSD to most populated cluster conformation | 1.1 | 0.8 |
| Average backbone RMSD to most populated cluster conformation | 0.9 | 0.4 |
| *Ramachandran Plot* | | |
| Residues in most favored region | 54.44% | 43.33% |
| Residues in additional allowed regions | 45.56% | 54.45% |
| Residues in generously allowed regions | 0% | 2.22% |
| Residues in disallowed regions | 0% | 0% |