# Novel Chimeric Protein as Therapeutic Vaccine against Hepatitis B Virus

Nadeem A. Kizilbash, Abdul Hai, and Jamal Alruwaili

*Abstract*—**A significant percentage of the human population does not respond well to commercially available recombinant Hepatitis B Virus (HBV) surface vaccine. So it must be replaced by modified vaccines. A chimeric protein comprising both the core and surface portions of the viral envelope was designed on the premise that if the HBV surface protein is fused with the core protein of the viral envelope, it can produce B-cell as well as T-cell immune response. A 23 kDa molecular weight protein, comprising 216 amino acids and consisting of the core and surface regions of the viral envelope protein, was designed. NNPREDICT and PSIPRED programs have provided the secondary structure elements of the protein. The tertiary structure of the protein was predicted by the use of 3D-JIGSAW program. In the predicted tertiary structure, α-helices form a helical bundle domain and the β-strands form another separate domain.**

*Index Terms*—**Chimeric protein, hepatitis B virus, homology modeling, therapeutic vaccine.**

## I.    INTRODUCTION

Viral hepatitis causes chronic liver disease, cirrhosis, and hepatocellular carcinoma [1]. Out of the five known hepatitis viruses: hepatitis B virus (HBV) [2], hepatitis C virus (HCV), [3] and the hepatitis delta (or hepatitis D) virus (HDV) [4]; three can cause infection and chronic hepatitis. Hepatitis B virus (HBV) has reportedly infected 2 billion people worldwide. Out of these, 350 million suffer from chronic HBV infection. WHO estimates that 5% of the population of the world is infected with HBV, thereby making this virus one of the most important public health problems in both developed and non-developed countries [5]. The distribution of hepatitis B infection varies greatly throughout the world. The highest prevalence is in Southeast Asia, China, and Africa [6].

Vaccines against HBV have been available for over 20 years. World Health Organization recommends inclusion of hepatitis B vaccination worldwide. Various immunization strategies have been developed for routine infant vaccination, prevention of perinatal transmission and catch-up vaccination for older people. Significant percentage of the human population does not respond well to recombinant hepatitis B surface vaccine, so, it must be replaced by some modified vaccines [7].

Knowledge about the structure and genomic organization of the hepatitis B virus (HBV) has led to the development of immunogenic vaccines. The HBsAg proteins include a major polypeptide of 226 amino acids (S) previously also called major surface protein, both in non-glycosylated (p24) and glycosylated (gp27) forms. The middle-sized protein (M) shares 226 amino acids of the p24 region at the C terminus and has an additional 55 amino acids at the N terminus, is called pre-S2. The large HBs protein (L) contains, in addition to the S and pre-S2 domains, the pre-S1 domain of 119 amino acids [8]-[11] (Fig. 1).
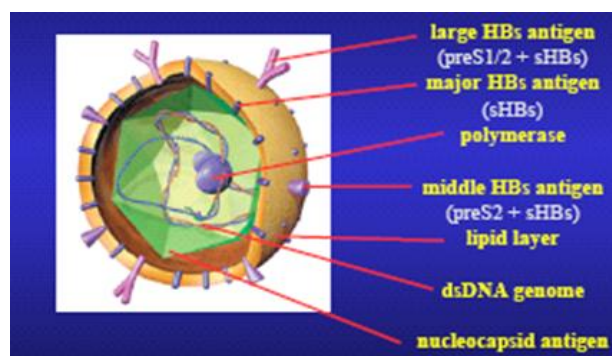


Fig. 1. Structure and organization of the hepatitis B virus (HBV).

A chimeric protein comprising both the core and surface portions of the viral envelope that can produce both B-cell and T-cell immune responses has been designed (Fig. 2).

| HBV Core Protein amino acid 1-78 | HBV Surface Protein amino acid 111-180 | HBV Core Protein amino acid 80-144 |
|---|---|---|

<span style="color:red">E F M D I D P Y K E F G A T V E L L S F L P S D F F P
S V R D L L G T A S A L Y Q V A L E S P E H C S P H
H T A L R Q A I L C W G E L M T L A T W V G G N L
E D</span> <span style="color:blue">V D P G S S T T S A G P C R T C T T S A Q G T S
K Y P P C C C T K P S D G N C T C I P I P S T W A F
G K F L W E W A S A R F S W L S L K L I</span> <span style="color:red">S R D L V V
S Y V N T N M G L K F R Q L L W F H I S C L T F G
G E T V I E Y L V S F G V W I R T P P A Y R P P N A P
I L S T L P L E</span>

Fig. 2. Amino acid sequence of the chimeric protein constructed from core and surface proteins of HBV. Amino acids 1-78 from HBV Core Protein form the N-terminus (red), amino acids 111-180 from HBV Surface Protein form the middle portion (blue) and amino acids 80-144 from HBV Core Protein form the C-terminus (red).

This novel chimeric protein has been designed on the premise that if the Hepatitis B virus (HBV) surface protein is fused with the core protein of the viral envelope, it can produce B-cell as well as T-cell response (Fig. 3 and Fig. 4).

B-cell epitopes involve interaction with the antibody on

---

the surface of the B-cells only, whereas T-cell epitopes interact with both the T-cell receptors and the class II MHC molecules (Fig. 5). T-cells tend to recognize internal peptides that are exposed by processing within antigen-presenting cells or altered self-cells. Homology modeling has been used to acquire the secondary structure elements and tertiary fold of the protein.
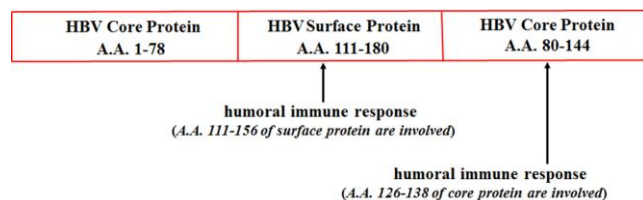
| HBV Core Protein A.A. 1-78 | HBV Surface Protein A.A. 111-180 | HBV Core Protein A.A. 80-144 |
|---|---|---|

**humoral immune response**
*(A.A. 111-156 of surface protein are involved)*

**humoral immune response**
*(A.A. 126-138 of core protein are involved)*

Fig. 3. Schematic showing the design of the chimeric protein and the immunogenic regions producing humoral (B-cell) immune response.

| HBV Core Protein A.A. 1-78 | HBV Surface Protein A.A. 111-180 | HBV Core Protein A.A. 80-144 |
|---|---|---|

**CTL epitopes**
*(A.A. 18-27 of core protein are involved)*

**Th epitopes**
*(A.A. 1-20, 50-69 of core protein are involved)*

**CTL epitopes**
*(A.A. 88-96, 141-144 of core protein are involved)*

**Th epitopes**
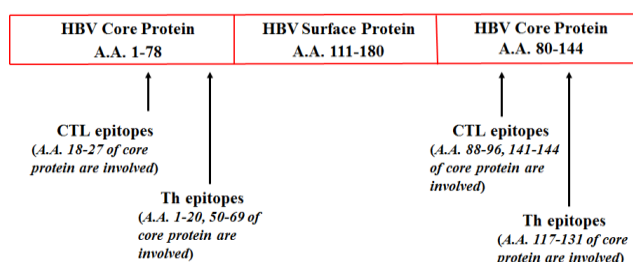*(A.A. 117-131 of core protein are involved)*

Fig. 4. Schematic showing the design of the chimeric protein and the immunogenic regions producing cellular (T-cell) immune response.
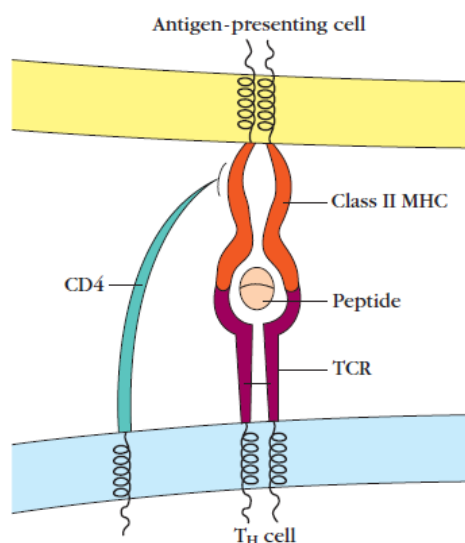


Fig. 5. Schematic diagram of the ternary complex formed between a T-cell receptor (TCR) on a helper T-cell cell (Th), an antigen, and a class II MHC molecule.

## II. METHODS

The amino acid sequence of the chimeric protein was first aligned with the known structure of human hepatitis B virus capsid, chain D (Fig. 6) [12]. The template protein was found through sequence alignment with the proteins in PDB using BLAST [13] followed by alignment corrections. Using this as the template protein, secondary and tertiary structures were determined.

### A. Secondary Structure Prediction

NNpredict program [14] predicts the secondary structure type for each residue in an amino acid sequence using a neural network. The network weights are determined by a modification of a separate program by Rumelhart and McClelland [15]. PSIPRED program also uses neural networks to perform an analysis on output obtained from PSI-BLAST (Position Specific Iterated - BLAST) [16].

### B. Tertiary Structure Determination by Homology Modeling

3D-JIGSAW (version 2.0) server has been used to build a 3-D model of this protein based on homologues of known structures. In the automatic mode, the program looks for homologous templates in the sequence databases and splits the query amino acid sequence into domains. This process can take up to an hour, depending on the load of the system [17]-[19].
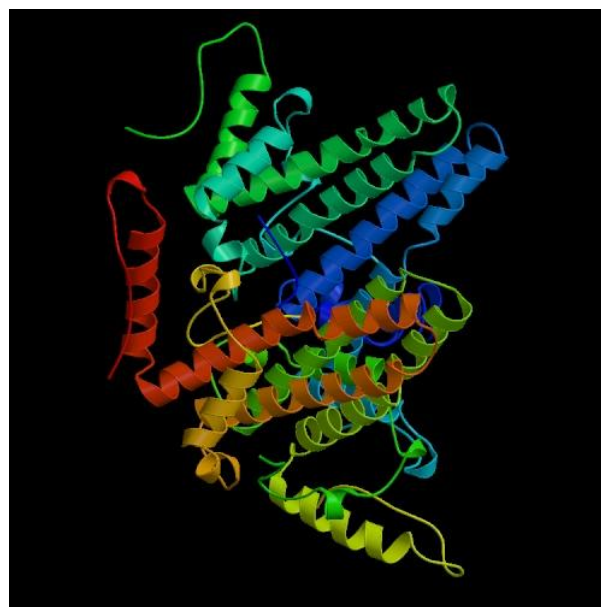


Fig. 6. The crystal structure at 3.30 Å resolution of the human hepatitis B virus capsid (chains C, D, B & A) (S. A. Wynne, R. A. Crowther, A. G. Leslie, (1999) *Mol. Cell* **3**: 771-780).

## III. RESULTS

The secondary structure predicted by the NNPREDICT program comprises six α-helices and four *β*-strands (Fig. 7). The first α-helix, $\alpha_I$, is well defined since it does not contain any Glycine or Proline amino acid residues. At the beginning of $\alpha_I$ is Glycine-12 and at the other end is Leucine-17. $\alpha_{II}$ has Glycine-34 in the middle which produces a "kink" in the α-helix. $\alpha_{III}$ has Glycine-65 in the middle which produces a "kink". $\alpha_V$ also has Glycine-131 in the early part of the helix which can produce a "kink" in the secondary structure. The remaining α-helix ($\alpha_{VI}$) and the two *β*-strands are well defined.

PSIPRED predicted ten α-helices for the protein (Fig. 7).

The first α-helix ($\alpha_I$) is well defined. It starts just after Proline-7 and ends just before Glycine-12. The second α-helix ($\alpha_{II}$) has Proline-22 in the middle. $\alpha_{III}$ has Glycine-34 in the middle. $\alpha_{IV}$ has Glycine-65 in the middle. $\alpha_V$, $\beta_A$ and $\beta_B$ are well defined. $\alpha_{VI}$ starts with Proline-125 and has Glycine-131 in the middle. Two of the α-helices, $\alpha_{VII}$ and $\alpha_{VIII}$ are well defined. The last two α-helices, $\alpha_{IX}$ and $\alpha_X$, have Glycine-181 and Glycine-193 respectively at the beginning.

Fig. 7. Comparison of secondary structures predicted for chimeric protein by the NNPREDICT (structural elements shown at the top) and PSIPRED (structural elements shown at the bottom) programs.

The tertiary structure predicted by 3-D Jigsaw shows the predicted $\alpha$-helices and the $\beta$-strands. The $\alpha$-helices form a helical domain and the $\beta$-strands together form another domain (Fig. 8).
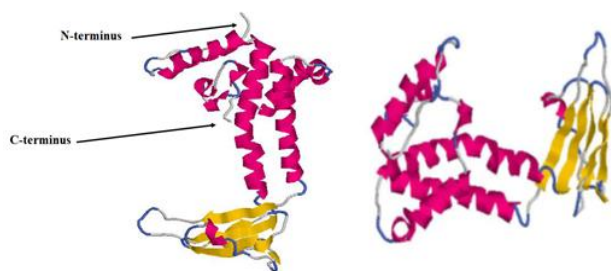


Fig. 8. Tertiary structure of the chimeric protein predicted by 3-D Jigsaw (front and side views). The N-terminus and C-terminus of the protein are shown.

## IV. DISCUSSION

A vaccine uses weakened or killed forms of a microbe or one of the surface proteins from the microbe coat to stimulate the immune system to recognize the microbe as foreign and to destroy it. Vaccines can be prophylactic or therapeutic. Therapeutic vaccines treat a disease or condition by inducing or strengthening a pre-existing immune response.

Homology modeling is useful for predicting structures of proteins for which NMR and X-ray models are not available. Homology modeling is not a pure quantum mechanical or molecular mechanics-based method. It is commonly performed using molecular modeling programs which employ force field methods in the optimization step for the construction of the protein model. Homology modeling is based on two observations: 1) amino acid sequence determines the structure of a protein [20]; 2) during evolution the structure of a protein is conserved rather than its sequence [21]. This suggests that two proteins with low sequence identity still can fold similarly.

The study has provided the secondary structure elements and tertiary fold of the protein by homology modeling. The secondary structure elements predicted by the NNPREDICT program comprise six $\alpha$-helices and four $\beta$-strands (Fig. 7). PSIPRED has shown the protein to comprise ten $\alpha$-helices and two $\beta$-strands. Both of the $\beta$-strands are in the middle of the protein whereas five $\alpha$-helices are towards the N-terminus and five $\alpha$-helices are towards the C-terminus (Fig. 7). The tertiary structure provided by 3-D Jigsaw also shows $\beta$-strands to be in the middle. There are $\alpha$-helices at both the N- and C-terminus and they together form a helical bundle (Fig. 8).

Both PSIPRED and 3-D Jigsaw programs have corroborated the information provided by each other. It is possible that both these programs given a comparably more accurate picture about the protein. The data has provided enough information to be able to create a model of this protein for binding to antibodies on the B-cell surface and MHC-II and T-cell receptor involved in T-cell immune response. Knowledge about the structure and the binding mode of the chimeric protein can allow the use of software tools that can confirm the epitope-binding regions of the protein available to antibodies on the surface of B-cells; and to class II MHC molecule and TCR involved in T-cell response. This information can be useful for refining the design further and allow this protein to be used as a therapeutic vaccine against HBV.

## V. CONCLUSION

The data produced by PSIPRED and 3-D Jigsaw programs have shown that the chimeric protein comprises $\alpha$-helices and $\beta$-strands. Both these programs have corroborated each other's results by placing $\beta$-strands in the middle and $\alpha$-helices at both the N- and C-terminus. The 3-D Jigsaw program has further shown that the $\beta$-strands form a sheet-like structure and the $\alpha$-helices form a helical bundle.

This information can be used to construct a model for the chimeric protein that can explain binding to antibodies found on the surface of B-cells, class II MHC molecules and T-cell receptors (TCR). A future direction can be to conduct a docking study to investigate the binding orientation of the chimeric protein to class II MHC molecule and TCR.

## REFERENCES

[1] M. J. Alter and E. E. Mast, "The epidemiology of viral hepatitis in the United States," *Gastroenterol. Clinics of North America*, vol. 23, no. 3, pp. 437–455, 1994.

[2] J. Y. N. Lau and T. L. Wright, "Molecular virology and pathogenesis of hepatitis B," *The Lancet*, vol. 342, no. 8883, pp. 1311-1340, Nov. 1993.

[3] S. Cooper, A. L. Erickson, E. J. Adams, J. Kansopon, A. J. Weiner, D. Y. Chien, M. Houghton, P. Parham, and C. M. Walker, "Analysis of a Successful Immune Response against Hepatitis C Virus," *Immunity*, vol. 10, no. 4, pp. 439-449, Apr. 1999.

[4] J. H. Hoofnagle, "Type D (Delta) Hepatitis," *JAMA*, vol. 261, no. 9, pp. 1321-1325, March 1989.

[5] D. Lavanchy, "Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures," *J. of Viral Hepatitis*, vol. 11, no. 2, pp. 97-107, March 2004.

[6] M. Kane, "Global programme for control of hepatitis B infection," *Vaccine*, vol. 13, suppl. 1, pp. S47–S49, 1995.

[7] M. V. Chengalvala, R. A. Bhat, B. M. Bhat, S. K. Vernon, and M. D. Lubeck, "Enhanced immunogenicity of hepatitis B surface antigen by insertion of a helper T cell epitope from tetanus toxoid," *Vaccine*, vol. 17, no. 9-10, pp. 1035-1041, March 1999.

[8] D. Ganem and H. E. Varmus, "The molecular biology of the hepatitis B viruses," *Ann Rev Biochemistry*, vol. 56, pp. 651-693, 1987.

[9] A. R. Neurath and S. B. H. Kent, "Antigenic structure of human hepatitis viruses," *Immunochemistry of viruses: The basis for serodiagnosis and vaccines*, M. H. V. van Regenmantel and A. R. Neurath, eds. New York: Elsevier, pp. 325-365, 1985.

[10] D. R. Milich, G. B. Thornton, A. R. Neurath, S. B. Kent, M. L. Michel, P. Tiollais, and F. V. Chisari, "Enhanced immunogenicity of the pre-S region of hepatitis B surface antigen," *Science*, vol. 228, pp. 1195-1199. June 1985.

[11] K. H. Heerman, U. Goldmann, W. Schwartz, T. Seyffarth, H. Baumgarten, and W. H. Gerlich, "Large surface proteins of hepatitis B virus containing the pre-S sequence," *J. Virol.* vol. 52, pp. 396-402, Nov., 1984.

[12] S. A. Wynne, R. A. Crowther, and A. G. Leslie, "The crystal structure of the human hepatitis B virus capsid," *Mol. Cell* vol. 3, no. 6, pp. 771-780, 1999.

[13] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol. Biol.*, vol. 215, pp. 403-410, Oct. 1990.

[14] L. H. Holley and M. Karplus, "Protein secondary structure prediction with a neural network," in *Proc. Nat. Acad. Sci.*, USA, vol. 86, pp. 152-156, Jan. 1989.

[15] J. L. McClelland, "Parallel distributed processing: Implications for cognition and development," in *Parallel Distributed Processing*: *Implications for Psychology and Neurobiology*, R. Morris, Ed., New York: Oxford University Press, pp. 8-45, 1989.

[16] L. J. McGuffin, K. Bryson, and D. T. Jones, *Bioinform.,* vol. 16, pp. 404-405, Apr. 2000.

[17] P. A. Bates, L. A. Kelley, R. M. MacCallum, and M. J. E. Sternberg, "Enhancement of Protein Modelling by Human Intervention in Applying the Automatic Programs 3D-JIGSAW and 3D-PSSM," *Proteins*: *Struc. Func. Gen.*, Suppl. 5, pp. 39-46, 2001.

[18] P. A. Bates and M. J. E. Sternberg, "Model Building by Comparison at CASP3: Using Expert Knowledge and Computer Automation," *Proteins*: *Struc. Func. Gen.*, Suppl. 3, pp. 47-54, 1999.

[19] B. Contreras-Moreira and P. A. Bates "Domain Fishing: a first step in protein comparative modeling," *Bioinformatics*, vol. 18, pp. 1141-1142, 2002.

[20] C. J. Epstein, R. F. Goldberger, and C. B. Anfinsen, "The genetic control of tertiary protein structure: Studies with model systems," *Cold Spring Harb. Symp. Quant. Biol.*, vol. 28, pp. 439-449, 1963.

[21] C. Sander and R. Schneider, "Database of homology-derived protein structures and the structural meaning of sequence alignment," *Proteins*, vol. 9, pp. 56-68, 1991.

**Nadeem Kizilbash** was born in Karachi on the 25th of December, 1963. He received his Bachelor's degree in chemistry in 1989 from Longwood College in Farmville, U.S.A. His Master's degree was also in chemistry in 1994 from Washington University in St. Louis, U.S.A. His doctoral degree was awarded in biophysics in 2003 from Boston University in U.S.A.

He is presently working as an assistant professor of Biochemistry at Northern Border University in Arar, Saudi Arabia. He has previously worked as an Assistant Professor at Quaid-i-Azam University in Pakistan. His work experience also includes service as a Senior Scientist at National Institute for Biotechnology and Genetic Engineering in Pakistan. He has published ten papers in various journals on topics as diverse as protein structure, drug delivery and gene polymorphism.

Dr. Kizilbash is currently a member of editorial board of two international research journals and he has also served as a referee for various journals and a doctoral dissertation in Egypt. He is listed as a Ph.D. supervisor by the Higher Education Commission of Pakistan. At Quaid-i-Azam University, he has supervised the theses of two M. Phil. students.