

BFluena: A Proteomic Database on Bird Flu

Parveen Salahuddin¹ & Asad U Khan^{1,2*}

¹Distributed Information Sub-Centre, A. M. U. Aligarh - 202002, India; ²Interdisciplinary Biotechnology Unit, A. M. U. Aligarh - 202002, India; Asad U Khan - Email: asad.k@rediffmail.com; Phone: +91 571 2723088; Fax: +91 571 2721776; *Corresponding author

Received September 22, 2011; Accepted September 23, 2011; Published September 28, 2011

Abstract:

Influenza A virus subtype H5N1, also known as "bird flu" has been documented to cause an outbreak of respiratory diseases in humans. The unprecedented spread of highly pathogenic avian influenza type A is a threat to veterinary and human health. The BFluena is a relational database which is solely devoted to proteomic information of H5N1 subtype. BFluena has novel features including computed physico-chemical properties data of H5N1 viral proteins, modeled structures of viral proteins, data of protein coordinates, experimental details, molecular description and bibliographic reference. The database also contains nucleotide and their decoded protein sequences data. The database can be searched in various modes by setting search options. The structure of viral protein could be visualized by JMol viewer or by Discovery Studio.

Availability: <http://www.bfluena.info>

Keywords: Bird flu, H5N1, hemagglutinin (HA), neuraminidase (NA), matrix protein M1, ion-channel M2.

Background:

Influenza A virus subtype H5N1, also known as "bird flu", A (H5N1) or simply H5N1 has been documented to cause an outbreak of respiratory diseases in humans. This was initially reported in poultry farms in Hong Kong in 1997 and was eradicated by culling of all chicken. Thereafter, the disease reemerged in Hong Kong prior to widespread throughout Asian countries and subsequently in other countries of the world. There is no evidence of human to human transmission to date because they cannot easily jump among humans. But in future, this strain might acquire ability to spread infection among humans because of RNA segmentation into 8 separate molecules. This segmentation allows frequent genetic exchange by segment reassortment in hosts, coinfecting with two different influenza A viruses. Alternatively, this strain can spread infection among humans *via* direct adaptation. H5N1 influenza is a relatively novel disease with poorly understood pathology and pathogenesis. Various factors are thought to be involved in the pathogenesis of H5N1 influenza virus. For example dysregulation of cytokines and chemokines may cause pathogenesis of H5N1 influenza. Viral replication is also implicated in the pathogenesis of virus. Moreover, H5N1 viral

proteins have various functions ranging from viral RNA synthesis to receptor binding. Several of these proteins have additional functions that may enhance pathogenicity of H5N1 influenza viruses. For example cleavability of HA have relation with virulence [1, 2]. In NA histidine to tyrosine substitution at position 274 confers resistance to oseltamivir [3, 4]. Similarly, Glutamic acid at position 92 of NS1 confers resistance to TNF- α and interferons and thereby increases the pathogenicity of virus [5]. Moreover, Glu-Pro-Glu-Val (EPEV) and Glu-Ser- Glu-Val (ESEV) motifs at the C-terminus of NS1 may disrupt important cell signaling pathways [6]. The NS gene also contributes to dysregulation of cytokines and chemokines. In M2 serine to asparagine substitution at position 31 confers resistance to amantadine. In PB2 protein lysine at position 627 enhances pathogenicity and promotes replication in the cells of upper respiratory tract. Further, serine at position 66 of PB1-F2 increases the virulence of virus.

An increasing number of H5N1 viral genome has been sequenced and structures of several influenza A viral proteins have been solved in recent years. Therefore, there is an urgent need to construct birdflu database that hosts sequences data

along with their computed physico-chemical properties data, structural data and web-based analysis tools in one place for the development of vaccines and drugs in order to prevent a human pandemic, to predict pathogenicity of the virus and to pursue in depth study on the virus. Here, we report the development and web implementation BFluenza database which is a comprehensive, web-accessible, relational database containing novel proteomic informations on H5N1 viral proteins.

Methodology:

Database Design and Contents:

The BFluenza was developed as a relational database using MYSQL software on the Linux Server. The database is searched by PHP script. The open source program Apache HTTP Server was used to build HTTP Server. The sequence data of individual gene was retrieved from NCBI database. These sequences were further computed at expasy server choosing Prot Param tool. Thus, sequence and first database hosts over 5,000 sequences data and their computed proteomic information. The structure database contains 55 structures of H5N1 virus proteins as well as details of structural related information. This information was retrieved from RCSB Protein Data Bank. The modeled structure of viral protein was constructed using SWISS-PDB Server. The structure of H5N1 viral protein can be viewed by Jmol viewer. Alternatively, modeled viral protein structure could be visualized by Discovery Studio.

The BFluenza shares some similarities with other databases [7, 8], in that it contains large number of nucleotide and protein sequences data as well as structures of H5N1 proteins [7, 9]. The BFluenza also shows other similarities to databases [9, 7] like tools for aligning sequences, visualizing three-dimensional structure [9] and for predicting antigenic segments. The BFluenza has several novel features not available in other databases [7-9] like it contains various physico-chemical properties informations of solely H5N1 subtype such as protein molecular weight, protein total length, isoelectric point, extinction coefficient, No. of thiols, N-terminal and C-terminal. Further, database contains several structural informations not present in other database [9] such as data on protein coordinates, molecular descriptions of viral protein, experimental details, and bibliography of H5N1 viral proteins and structures of modeled H5N1 viral proteins [9]. Additionally, database has tools for constructing molecular model of viral proteins and for predicting various types of protein motifs in viral proteins not available in other databases [7-9].

The BFluenza provides a powerful search engine for user to retrieve different types of data hosted in database. There are various ways to search the database. In one such search option viral protein is searched by protein name (**Figure S1, see supplementary material**) and the result file shows sequence and physico-chemical properties data. The Structure Database (**Figure S2, see supplementary material**) can be searched by selecting search option protein name and result page displays data on crystal structures and modeled structures of H5N1 protein, bibliography, website for viewing viral protein

molecule in Jmol viewer, PDB text file containing protein coordinates, molecular information on viral protein and experimental details comprising information on experimental parameters. For a more advance search (**Figure S3, see supplementary material**) user may search by specific host and country. In this search option user can select various search options simultaneously like Protein Name, Specific host and Country from drop down menus. The search result displays information on S.No, Protein Name, Specific Host, and Details. The Details field contains information on sequence data, strain, specific host, country and subtype as well as other proteomic data.

Analysis of database:

The proteomic data as well as sequences data present in BFluenza database were subjected to analysis. The results show that, in hemagglutinin (HA) protein total length varied in the range of 543-568. This range was found most prevalent in the database. Other physico-chemical properties data of HA showed similar behavior. In neuraminidase (NA) protein total length occurred mostly in the range of 429-449. Analysis of HA and NA proteomic data suggest that these data were highly variable among all H5N1 viral proteins present in the database. Additionally, analysis of HA sequences in the database showed the presence polybasic cleavage site. This suggests that H5N1 virus was highly virulent. The nucleoprotein (NP) displayed less variations in physico-chemical properties data than HA and NA proteins. The PB1-F2 protein shows variation in amino acid sequences. The NS1 sequences contained Glutamic acid at position 92 which confers resistance to TNF- α and interferons and thereby increased the pathogenicity of virus [5]. Further, presence of Glu-Pro-Glu-Val (EPEV) and Glu-Ser- Glu-Val (ESEV) motifs at the C-terminus of NS1 may disrupt important cell signaling pathways [6]. RNA polymerase proteins (PA, PB1, PB2,) essential for viral replication showed variations in proteomic data. This observation validates lack of proof-reading activity of RNA polymerases.

Utility to Biological Community:

The important benefits of BFluenza are that from sequences and structures data and from web-based tools built in database; researcher can predict pathogenicity of virus, design efficient drugs and gain an insight into the structure-function relationship of viral protein.

Future Developments:

The database contents in BFluenza will be updated routinely. Using molecular docking tool like "Autodock" drug binding site will be mapped in neuraminidase and hemagglutinin proteins and binding site information along with binding energy and other relevant informations will be included in the structure database. Thus, these data will aid in designing new potential drugs and antibodies with higher efficacies than currently available drugs.

Acknowledgement:

Authors acknowledge the facilities of Distributed Information Sub-centre, Interdisciplinary Biotechnology Unit, A.M.U., Aligarh. Authors are also thankful to Prof. Alok Bhattacharya for his kind suggestions on the data base.

References:

- [1] Horimoto T & Kawaoka Y. *J Virol.* 1994 **68**: 3120 [PMID: 8151777]
- [2] Hatta M *et al.* *Science* 2001 **293**: 1840 [PMID: 11546875]
- [3] de Jong MD *et al.* *N Engl J Med.* 2005 **353**: 2667 [PMID: 16371632]
- [4] Le QM *et al.* *Nature* 2005 **437**: 1108 [PMID: 16228009]
- [5] Seo SH *et al.* *Nat Med.* 2002 **8**: 950 [PMID: 12195436]
- [6] Obenauer JC *et al.* *Science* 2006 **311**: 1576 [PMID: 16439620]
- [7] Bao Y *et al.* *J Virol.* 2008 **82**: 596 [PMID: 17942553]
- [8] Chang S *et al.* *Nucleic Acids Res.* 2007 **35**: D376 [PMID: 17065465]
- [9] Kulkarni-Kale U *et al.* *Nucleic Acids Res.* 2004 **32**: D289 [PMID: 14681415]

Edited by P Kanguane

Citation: Salahuddin & Khan. *Bioinformation* 7(3): 147-151 (2011)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

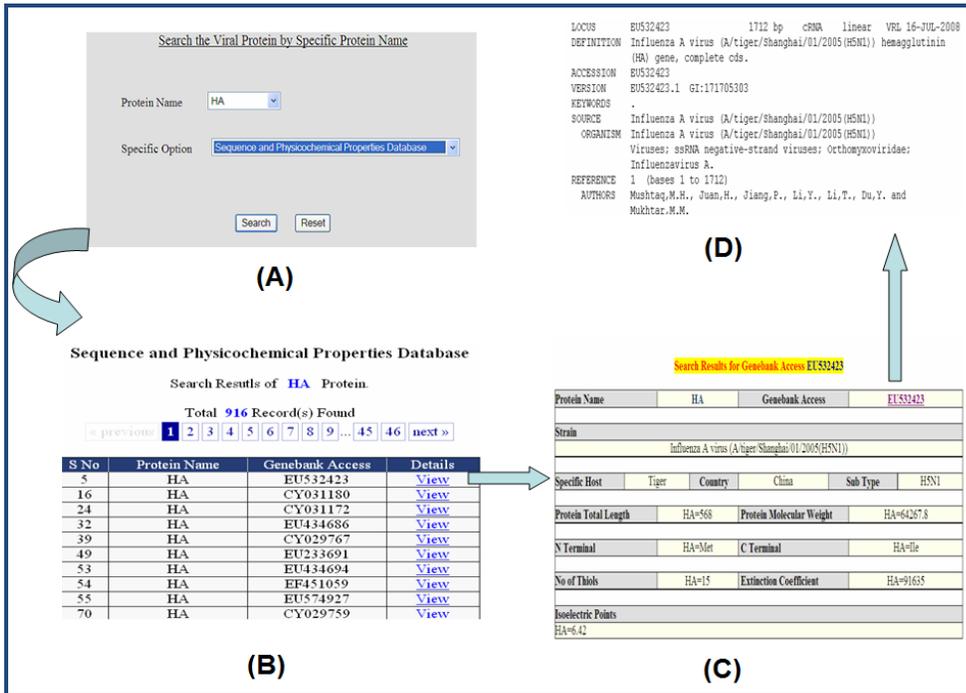


Figure S1: Search box is displayed in Figure 1(A) where viral protein was searched by selecting specific protein name like HA from drop down menu and in specific search options physico-chemical properties database was selected. A portion of search result containing data on S.No, Protein Name, Genebank Access and Details are displayed in Figure 1(B). Figure 1(C) shows the proteomic data of particular record and Figure 1(D) shows a portion of sequence data.

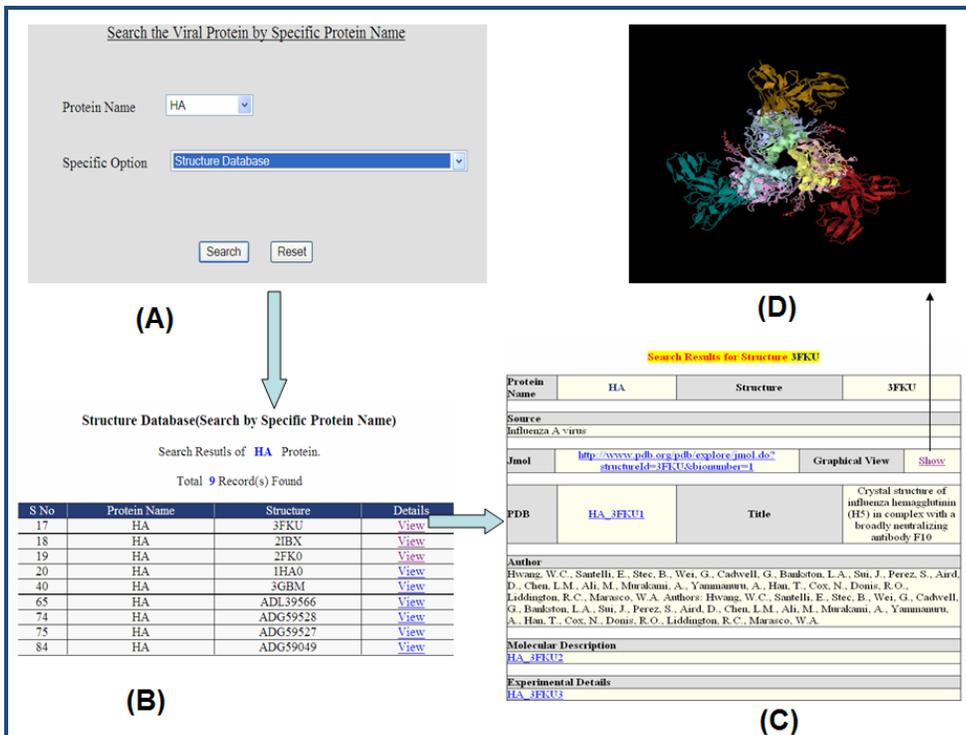


Figure S2: Search box is displayed in Figure 2(A) where viral protein was searched by selecting specific protein name like HA and in specific search options structure database was selected. Figure 2(B) displays the data on S.No, Protein Name, Structure and Details. Figure 2(C) shows the structure related informations and Figure 2(D) the structure of particular viral protein.

Search the Viral Protein by Specific Host and Country

Protein Name:

Specific Host:

Country:

(A)

↓

Search Viral Protein by Specific host and country

Search Protein : **HA**
 Specific Host : **Chicken**
 Country : **China**

Total **61** Record(s) Found

« previous **1** 2 3 4 next »

S No	Protein Name	Genebank Access	Specific Host	Details
337	HA	CY028980	Chicken	View
345	HA	CY028979	Chicken	View
670	HA	AF509026	Chicken	View
753	HA	CY028962	Chicken	View
1061	HA	CY028950	Chicken	View
1367	HA	CY028981	Chicken	View
1424	HA	CY028938	Chicken	View

(B)

Figure S3: The search box is displayed in Figure 3(A) where multiple options like protein name, specific host and country were selected from drop down menus. The results show (Figure 3(B)) data on S.No, Protein Name, Genebank Access, Specific Host and Details that contains proteomic data.