In-silico Molecular Analysis of Mutated Sequences of HFE1, HFE2, TFR2 and SLC40A1 causing Hemochromatosis Disease

Bilal Hussain[1*], Hassan Tariq[1], Tayyaba Sultana[2], Shahid Mahboob[2], Tariq Niaz[1]

[1]*Department of Bioinformatics GC University Faisalabad-38000, Pakistan*

[2]*Department of Zoology GC University Faisalabad-38000, Pakistan*
E-mail: profbilal@yahoo.com

[*]*Corresponding author*

Received: March 07, 2011 **Accepted: May 02, 2011**

Published: May 20, 2011

Abstract: *Hemochromatosis is a disorder in iron metabolism that is characterized by excess iron absorption. There are two forms of hemochromatosis: primary hemochromatosis is caused by a problem with your genes. Secondary or acquired hemochromatosis can be caused by other diseases. The main objective of this study is to analyze structure of different types of proteins involved in hemochromatosis. As this is the most threatening disease all over the world including Pakistan but unfortunately no data is available about it so it will be a great step to analyze the structure of its different proteins by using different online tools that gave different results according to their potential as each tool must show the same results for the same protein. Protein structure prediction is one of the most important goals persuaded by Bioinformatics for evolutionary studies and drug designing. Other objective of this project is to provide the prevalence of hemochromatosis in Faisalabad and structure prediction of its proteins to find the conserved regions. At the end the creation of a database is done that would contain all necessary information about diseases. Hereditary hemochromatosis is mainly caused by a defect in a gene called HFE. There are several types of genetic hemochromatosis. These include: Type I or Classic (HHC); Type II or Juvenile (JHC); Type III or Transferrin Receptor Mutation; and Type IV or Ferroportin Mutation. Most types of hereditary haemochromatosis have autosomal recessive inheritance, while type IV has autosomal dominant inheritance.*

Keywords*: Hemochromatosis, Types, Analysis.*

Introduction

Hemochromatosis is a disorder in iron metabolism in which there is an excess iron absorption and saturation of iron-binding proteins. The primary affected tissues are the liver, pancreas, and skin. The excess iron deposition leads to bronze pigmentation of the organs and skin [6].

Primary hemochromatosis is referred to as type 1 hemochromatosis. It has autosomal recessive inheritance. The locus causing type 1 hemochromatosis has been designated the HFE1 locus and is a major histocompatibility complex (MHC) class-1 gene located on chromosome 6p21.3. Normal HFE1 has been shown to form a complex with the transferrin receptor and in so doing is thought to regulate the rate of iron transfer into cells. A mutation in HFE1 will lead to increased iron uptake and storage. The majority of hereditary hemochromatosis patients have inherited a mutation in HFE1 that results in the substitution of Cys 282 for a Tyr. This mutation cause loss of conformation of one of the immunoglobulin

domains in HFE1.Another mutation found in certain forms of hereditary hemochromatosis also affects the HFE1 locus and causes a change of His 63 to Asp. This latter mutation is found along with the more common C282Y mutations resulting in a compound heterozygosity. As a result of the C282Y mutation the HFE1 protein remains trapped in the intracellular compartment. Because it cannot associate with the transferin receptor there is a reduced uptake of iron by intestinal cells. It is thought that this defect in intestinal iron-uptake results in an increase in the expression of the divalent metal transporter (DMT-1) on the brush border of the intestinal villus cells. Excess DMT-1 expression leads to an inappropriate increase in intestinal iron absorption [6].

Juvenile hemochromatosis is inherited in autosomal recessive manner. It is characterized by onset of severe iron overload typically in the first to third decades of life. Males and females are equally affected. The full-length normal hemojuvelin protein comprising 426 amino acids is predicted to be approximately 41 KD in size. In-silico analysis revealed that a 35-amino acid hydrophobic signal peptide is present at N-terminal. At the C-terminal there is a transmembrane domain and a glycophosphatidyl inositol (GPI) in addition to signal sequence [8].

Type III is caused by mutations in the transferrin receptor-2 gene (TFR2) on chromosome 7q22. This disorder is referred to as HFE3. The function of TFR2 in the liver is not to act in the uptake of transferrin-bound iron but to sense iron levels and to act as a regulator of hepcidin function. The clinical features of type III hemochromatosis are similar to those of classic type 1 disease. Type III hemochromatosis is inherited as an autosomal recessive disorder. Inheritance is autosomal recessive [6].

Hemochromatosis Type IV is also called ferroportin disease because it is caused by mutations in the ferroportin gene on chromosome 2q32. The ferroportin gene is also identified as IREG1 (iron-regulated gene 1), SCL40A1 (solute carrier family 40). Ferroportin is a multiple transmembrane-domain containing iron transport protein. Type IV hemochromatosis is inherited as an autosomal dominant disease. Individuals with mutations in the gene encoding ferroportin also have iron overload, but unlike juvenile hemochromatosis and HFE-associated hereditary hemochromatosis, they show macrophages that are iron laden. Affected individuals have high serum ferritin concentration despite normal/low transferrin-iron saturation at early stages of the disease. It is transmitted in an autosomal dominant manner [7].

Iron overload occurs in utero. This severe, often fatal iron overload syndrome usually presents at birth. Inheritance is unknown, but autosomal recessive and mitochondrial inheritance have been postulated. No locus has been identified. Neonatal hemochromatosis (NH) is characterized by hepatic failure in the newborn period and heavy iron staining in the liver [2].

Materials and methods

The work on Hemochromatosis was done in Bioinformatics laboratory of GC, University Faisalabad. Selected genes for different types of hemochromatosis were analyzed i.e. HFE1, HFE2, TFR2 and SLC40A1.

Gene that causes the Hemochromatosis was selected:

- HFE1 gene was selected because it causes hemochromatosis type I.
- HFE2, gene was selected because it causes hemochromatosis type II.
- TFR2, gene was selected because it causes hemochromatosis type III.
- SLC40A1 gene was selected because it causes hemochromatosis type IV.

Basic information of Hemochromatosis genes was taken from NCBI's (National Centre for Biotechnology Information) genome project data base [4].

- In primary structure prediction tools ProtParam tool, Compute pI/Mw tool, ProtScale, REP etc. are used [4].
- In secondary structure prediction tools SOPMA and HNN are used.
- In tertiary structure visualization RasMol is used [4].

ProtParam is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered sequence. Compute pI/Mw is a program for determining the theoretical pI (isoelectric point) and Mw (molecular weight) from a SWISS-PROT or TrEMBL entry, or for a user sequence. REP searches a protein sequence for a repeat [4].

ProtScale allows computing and representing the profile produced by any amino acid scale on a selected protein. An amino acid scale is defined by a numerical value assigned to each type of amino acid. The most frequently used scales are hydrophobic scales.

SOPMA (self-optimized predictions method with alignment) It does significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. This is an improvement of SOPM method. This improved SOPM method (SOPMA) correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins.

HNN (Hirerical Neural Network) prediction method can be seen as an improvement in the famous classifier developed by Qian and Sejnowski, and derived from the system NETtalk. It is made up of two networks: a sequence-to-structure network and structure-to-structure network [4].

Results

In the present study sequences for HFE1, HFE2, TFR2, and SLC40A1 causing Hemochromatosis (Iron overloading disease) were analyzed to compare different types of hemochromatosis. The primary, secondary and tertiary tools ProtParam, Compute pI/Mw, REP, RADAR, ProtScale, Sopma, and HNN etc. showed different results for above mentioned sequences.

Visits of many hospitals and Laboratories research institutes such as Allied hospital, Faisal hospital, Al_Noor hospital, Social security hospital, PINM, and DHQ, etc. were made to know the prevalence of hemochromatosis in Faisalabad. But unfortunately no data is available about the percent age of patients of this disease due to the lack of knowledge about this disease. Different expasy tools (ProtParam, Compute pI/Mw, REP, RADAR, ProtScale, Sopma, and HNN) were used for the primary, secondary and tertiary structure prediction of different types of proteins causing hemochromatosis. They gave different result according to their potential. Results for four different types of proteins are shown in the Tables 1-5.

Discussion

HFE1, HFE2, TFR2, and SLC40A1 sequences were analyszed by using a tool ProtParam and found interesting variations among individual aminoacids and molecular weights when compared to the results of other tools used in the project. Compute pI/Mw is a program for determining the theoretical pI (isoelectric point) and Mw (molecular weight) from a SWISS-

PROT or TrEMBL entry, or for a user sequence. HFE1 has the isoelectric point 4.65 while its weight is 636510.69. HFE2 has the isoelectric point 4.84 while its weight is 216239.20. TFR2 has the isoelectric point 4.40 while its weight is 1609900.42. SLC40A1 has the isoelectric point 4.46 while its weight is 1700429.26. It shows that if isoelectric point is high then the molecular weight is less Table 2. REP searches a protein sequence for a repeat. According to this tool HFE1, HFE2, TFR2, and SLC40A1 have no repeats.

Table 1. Result from ProtParam

Genes	**HFE1**	**HFE2**	**TFR2**	**SLC40A1**
Amino acids		**Repeats (Composition)**		
Ala (A)	300 (21.2%)	464 (24.3%)	2024 (24.6%)	3506 (30.3%)
Arg (R)	4 (0.3%)	3 (0.2%)	8 (0.1%)	11 (0.1%)
Asn (N)	5 (0.4%)	5 (0.3%)	5 (0.1%)	5 (0.0%)
Asp (D)	0 (0.0%)	2 (0.1%)	0 (0.0%)	1 (0.0%)
Cys (C)	342 (24.1%)	476 (24.9%)	2010 (24.4%)	1955 (16.9%)
Gln (Q)	0 (0.0%)	0 (0.0%)	0 (0.0%0	0 (0.0%)
Glu (E)	7 (0.5%)	6 (0.3%)	6 (0.1%)	10 (0.1%)
Gly (G)	353 (24.9%)	516 (27.0%)	2235 (27.1%)	2280 (19.7%)
His (H)	4 (0.3%)	4 (0.2%)	2 (0.0%)	2 (0.0%)
Ile (I)	4 (0.3%)	7 (0.4%)	2 (0.0%)	4 (0.0%)
Leu (L)	4 (0.3%)	1 (0.1%)	0 (0.0%)	4 ()0.0%
Lys (K)	0 (0.0%)	0 (0.0%)	3 (0.0%)	0 (0.0%0
Met (M)	5 (0.4%)	5 (0.3%)	3 (0.0%)	6 90.1%)
Phe (F)	3 (0.2%)	1 (0.1%)	2 (0.0%)	2 (0.0%)
Pro (P)	1 (0.1%)	5 (0.3%)	4 (0.0%)	2 (0.0%)
Ser (S)	5 (0.4%)	3 (0.2%)	1926 (23.4%)	6 (0.1%)
Thr (T)	370 (26.1%)	402 (21.1%)	0 (0.0%)	3754 (32.5%)
Trp (W)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Tyr (Y)	1 (0.1%)	0 (0.0%)	0 (0.0%)	1 (0.0%)
Val (V)	0 (0.0%)	0 (0.0%)	7 (0.1%)	0 (0.0%)
Pyl (O)	8 (0.6%)	7 (0.4%)	?	9 (0.1%)
Sec (U)	1 (0.1%)	0 (0.0%)	?	2 (0.0%)
Molecular weight	119509.4	157338.0	677902.8	967433.9
Theoretical pI	5.64	5.03	6.16	5.88

Table 2. Result from Compute pI/Mw

Genes	**HFE1**	**HFE2**	**TFR2**	**SLC40A1**
Theoretical pI	4.65	4.84	4.40	4.46
Molecular Weight	636510.69	216239.20	1609900.42	1700429.26

SOPMA (self-optimized predictions method with alignment) It does significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. This is an improvement of SOPM method. This improved SOPM method (SOPMA) correctly predicts 69.5% of amino acids for a three-state description of the secondary structure (alpha-helix, beta-sheet and coil) in a whole database containing 126 chains of non-homologous (less than 25% identity) proteins Table 4. HNN (hirerical Neural Network) prediction method can be seen as an improvement in the famous classifier developed by Qian and Sejnowski, and derived from the system NETtalk. It is made up of two networks: a sequence-to-structure network and structure-to- structure network Table 5.

Table 3. Result from ProtScale

Genes	HFE1	HFE2	TFR2	SLC40A1
Amino Acids	**Amino acid Composition**			
Ala	1.800	1.800	1.800	1.800
Arg	-4.500	-4.500	-4.500	-4.500
Asn	-3.500	-3.500	-3.500	-3.500
Asp	-3.500	-3.500	-3.500	-3.500
Cys	2.500	2.500	2.500	2.500
Gln	-3.500	-3.500	3.500	3.500
Glu	-3.500	-3.500	-3.500	-3.500
Gly	-0.400	-0.400	-0.400	-0.400
Thr	-0.700	-0.700	-0.700	-0.700
His	-3.200	-3.200	-3.200	-3.200
Ile	4.500	4.500	4.500	4.500
Leu	3.800	3.800	3.800	3.800
Lys	3.900	-3.900	3.900	3.900
Met	1.900	1.900	1.900	1.900
Phe	2.800	2.800	2.800	2.800
Pro	-1.600	-1.600	-1.600	-1.600
Ser	-0.800	-0.800	-0.800	-0.800
Val	4.200	4.200	4.200	4.200

Table 4. Result from SOPMA

Genes	HFE1	HFE2	TFR2	SLC40A1
Alpha helix	874 (28.67%)	235 (27.52%)	513 (25.10%)	818 (31.20%)
310 helix	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Pi helix	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Beta bridge	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Extended strand	633 (20.77%)	139(16.28%)	394 (19.28%)	710 (27.08%)
Beta turn	297 (9.74%)	51 (5.97%)	187 (9.15%)	244 (9.31%)
Bend region	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Random coil	1244(40.81%)	429(50.23%)	950 (46.48%)	850(32.42%)

The outcome of this study is the complete characterization and analysis of different types of hemochromatosis. By this analysis we can determine the effect of drugs on these proteins [4, 9]. Domain and motif identification provides detailed information about hemochromatosis. In this project bioinformatics study of hemochromatosis was done which have never done before. So it is important to work on hemochromatosis to aware the people with this life threatening disease [1]. Results were taken and further analyzed from different databases. Unfortunately no data is available about the percent age of patients of this disease due to the lack of knowledge about this disease. Major finding of the project is the variability of the results of the tools used for the molecular analysis of the same sequence as found for seven different types of the hemochromatosis.

Table 5. Result from HNN

Genes	HFE1	HFE2	TFR2	SLC40A1
Alpha helix	991 (32.51%)	212 (24.82%)	997 (24.68%)	1375 (41.53%)
310 helix	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Pi helix	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Beta bridge	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Extended strand	495 (16.24%)	105 (12.30%)	515 (12.75%)	600 (18.12%)
Beta turn	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Bend region	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Random coil	1562 (51.25%)	537 (62.88%)	2527 (62.56%)	1336 (40.35%)

References

1. Borgaonkar M. R. (2003). Hemochromatosis, More Common than You Think, Can Fam Physician, 49, 36-43.
2. Driscoll S. G., A. M. Hayes, H. L. Levy (1988). Neonatal Hemochromatosis: Evidence for Autosomal Recessive Transmission, Genet, 43, A232.
3. Domenico I., D. M. Ward, G. Musci, J. Kaplan (2006). Iron Overload due to Mutations in Ferroportin, Haematologica, 91, 92-95.
4. Gasteiger E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch (2005). Protein Identification and Analysis Tools on the ExPASy Server, John M. Walker (Ed): The Proteomics Protocols Handbook, Humana Press, 571-607.
5. Kontoghiorghes G. J., A. Spyrou, A. Kolnagou (2010). Iron Chelation Therapy in Hereditary Hemochromatosis and Thalassemia Intermedia Regulatory and Non Regulatory Mechanisms of Increased Iron Absorption, Hemoglobin, 34(3), 251-264.
6. Burke W., E. Thomson, M. J. Khoury, S. M. McDonnell, N. Press, P. C. Adams, J. C. Barton, E. Beutler, G. Brittenham, A. Buchanan, E. W. Clayton, M. E. Cogswell, E. M. Meslin, A. G. Motulsky, L. W. Powell, E. Sigal, B. S. Wilfond, F. S. Collins (1998). Hereditary Hemochromatosis: Gene Discovery and its Implications for Population-based Screening, JAMA, 280(2), 172-178.
7. Montosi G., A. Donovan, A. Totaro, C. Garuti, E. Pignatti, S. Cassanelli, C. C. Trenor, P. Gasparini, N. C. Andrews, A. Pietrangelo (2001). Autosomal-dominant Hemochromatosis is Associated with a Mutation in the Ferroportin (SLC11A3) Gene, J Clin Invest, 108, 619-623.
8. Zhang A. S., A. P. West Jr., A. E. Wyman, P. J. Bjorkman, C. A. Enns (2005). Interaction of Hemojuvelin with Neogenin Results in Iron Accumulation in Human Embryonic Kidney 293 Cells, J Biol Chem, 280, 33885-33894.
9. Zlocha J., L. Kovacs, S. Pozgayova, V. Kupcova, S. Durinova (2006). Molecular Genetic Diagnostics and Screening of Hereditary Hemochromatosis, Vnitr Lek, 52(6), 602-608.

Bilal Hussain, M.Phil., Ph.D. Scholar
E-mail: profbilal@yahoo.com

Bilal Hussain is a Ph.D. Scholar from the Government College University – Faisalabad with the distinction of Bronze Medal in the session. Since 2003 he is working as a Lecturer in the Department of Bioinformatics and Biotechnology. He is among the founders of Department of Bioinformatics and Biotechnology in GC University Faisalabad. He has specialization in the field of Molecular Biology and fisheries. He is performing many administrative duties in the department of Bioinformatics and Biotechnology and is an active member of research group of the department with many publications. He is working on molecular study of genetic disorders.

Hassan Tariq, M.Sc. in Bioimformatics
E-mail: hassantariq9@yahoo.com

Hassan Tariq graduated from the Government College University – Faisalabad in 2008 with the distinction of securing second highest percentage in the session. Since 2009 he is a working as a Research Officer in the Department of Bioinformatics and Biotechnology. He is the founder of Database and Software Engineering Research Group in the Department of Bioinformatics and Biotechnology. He is also an active member of teaching Faculty of the department, teaching some of the interesting subjects like Bioinformatics Software Development, also publishing textbooks. Scientific interests: Biological Databases, Data Mining, Software Engineering in Bioinformatics, Web Engineering.

Dr. Tayyaba Sultans, Ph.D. in Genetics
E-mail: arif143@yahoo.com

Dr. Tayyaba Sultana, Assistant professor, Department of Zoology, GC University, Faisalabad, Pakistan. She is HEC approved Ph.D. Supervisor too. During her academic career she got the distinction's scholarships like Quaid-i-Azam Merit Scholarship, INFAQ Foundation Merit Scholarship and The University Grants Commission (UGC) Merit Scholarship. She has a partial Research Project execution on my credit working as Co-link coordinator (2006-2009) under the British council-Higher Education Commission (JHELP-II) on 'Capacity building for molecular biology studies in Fish Production. She has more than ten research publications in the national and international impact factor research journals and one book by the VDM Publishers (Paperback Feb. 23, 2010) it is available on www.amazon.com.

Shahid Mahboob, Ph.D.
Email: shahidmahboob60@hotmail.com

Prof. Shahid Mahboob got his Ph.D. from University of Agriculture, Faisalabad, in 1992. He is working as Professor of Zoology in GC University, Faisalabad. He has awarded Prof. Mirza Azhar Beg Gold medal as young Researcher in Fisheries by Zoological Society of Pakistan. He has been awarded Best University teacher Award 2003 by Higher education Commission of Pakistan. He has served as Vice Chancellor and Dean Faculty of Science & Technology of GC University, Faisalabad. He has published resrach paper in diversified areas of biological sciences particularly using fish as model.

Tariq Niaz, M.Sc. in Entomology
E-mail: tniaz08@hotmail.com

Tariq Niaz graduated from the University of Agriculture – Faisalabad in 1979 with the distinction of securing third highest percentage in the session. He is serving Ayub Agriculture Research Institute, Faisalabad since 1980 as an entomologist. He has also supervised several thesis of M.Sc. and M.Phil. student. He has published 40 research papers and review articles in International and national peer review journals and proceedings of conferences/workshops. Scientific interests: Plant protection, Bioassay of pesticides, Biological databases, Bioinformatics, Software development, Web engineering.