

# The Gene Gateway Workbook

A collection of activities derived from the tutorials at Gene Gateway, a guide to online data sources for learning about genetic disorders, genes, and proteins.



**Human Genome Landmarks**  
Selected Genes, Traits, and Disorders  
[www.doe.gov/genomes/posters/landmarks](http://www.doe.gov/genomes/posters/landmarks)

Multiple myeloma oncogene  
Orofacial cleft  
Leukemia, acute nonlymphocytic  
Fanconi anemia, complementation group E  
Ankylosing spondylitis  
Stickler syndrome, type II  
OSMED syndrome  
Weissenbacher-Zweymuller syndrome  
Deafness, nonsyndromic sensorineural  
Dyslexia  
**Hemochromatosis** ←  
Porphyria variegata  
Pemphigoid, susceptibility to  
Immune suppression to streptococcal antigen  
Sialidosis, types I and II  
Panbronchiolitis, diffuse  
Psoriasis susceptibility

To view the chromosomes of the Human Genome Landmarks poster online, order your free copy of the poster, or download additional copies of this workbook, go to the Gene Gateway Web site:  
[www.DOEgenomes.org/genegateway](http://www.DOEgenomes.org/genegateway)

**Using hereditary hemochromatosis as a model, access a variety of Web sites and databases to:**

- Learn about a genetic disorder and its associated gene.
- Identify the most common disease-causing mutation.
- Find the gene locus on a chromosome map.
- Examine the sequence and structure of a gene.
- Access the amino acid sequence of a gene's protein product.
- Explore the 3-D structure of the gene's protein product.



## Table of Contents

Introduction .....	5
Why use hereditary hemochromatosis as a model? .....	6
Some basic concepts to understand before starting .....	6
<u>Activity #1</u> .....	7
Online Resources: OMIM and GeneReviews	
- Learn about the genetic disorder and its associated gene.	
- Identify the most common disease-causing mutation.	
<u>Activity #2</u> .....	13
Online Resource: NCBI Map Viewer	
- Find the hereditary hemochromatosis gene locus on a chromosome map.	
<u>Activity #3</u> .....	19
Online Resources: Entrez Gene, GenBank, and GenAtlas	
- Examine the sequence and structure of a gene.	
<u>Activity #4</u> .....	29
Online Resource: Swiss-Prot	
- Access the amino acid sequence of a gene's protein product.	
<u>Activity #5</u> .....	35
Online Resources: Protein Data Bank and Protein Explorer	
- Explore the 3-D structure of the gene's protein product.	
Table of Standard Genetic Code .....	47
Hereditary Hemochromatosis Worksheet .....	49
Contact Information .....	52



## Introduction

The Gene Gateway Workbook is a collection of activities with screenshots and step-by-step instructions designed to introduce new users to genetic disorder and bioinformatics resources freely available on the Web. It should take about 3 hours to complete all five activities.

The workbook activities were derived from more detailed guides and tutorials available at the Gene Gateway Web site ([www.DOEgenomes.org/genegateway](http://www.DOEgenomes.org/genegateway)). The Gene Gateway Web site was created as a resource for learning more about the genes, traits, and disorders listed on the Human Genome Landmarks poster, but it can be used to investigate any gene or genetic disorder of interest.

Many guides to genome Web resources are designed for bioscience researchers and are too technical for non-experts. This workbook and other Gene Gateway resources target a more general audience: teachers, high school and college students, patients with disorders and their families, and anyone else who wants to learn more about how life works at a molecular level.

This workbook shows you how to get started using bioinformatics resources that often intimidate and overwhelm new users. It also demonstrates how information from one resource, such as annotated protein sequence data from Swiss-Prot, can be used to reinforce and clarify information available from another resource, such as 3-D structures from Protein Data Bank. Gene Gateway provides users with a systematic approach to using multiple bioinformatics databases together to gain a better understanding of how genes and proteins can contribute to the development of a particular genetic condition.

Using the genetic disorder hereditary hemochromatosis as a model, this workbook shows you how to access:

- Online Mendelian Inheritance in Man (OMIM) and GeneReviews to learn about a genetic disorder, its associated gene or genes, and common disease-causing mutations
- NCBI Map Viewer to find a gene locus on a chromosome map
- Entrez Gene, GenBank, and GenAtlas to examine the sequence and structure of a gene
- Swiss-Prot to find the annotated amino acid sequence of a gene's protein product
- Protein Data Bank and Protein Explorer to view and modify the 3-D structure of the gene's protein product

Skills gained by working through the activities in this workbook can be applied to learning about other genetic disorders, genes, and proteins.

This workbook and other genome science resources available from the DOE Genomes Web site ([www.DOEgenomes.org](http://www.DOEgenomes.org)) are products of the Office of Biological and Environmental Research of the U.S. Department of Energy Office of Science.

## Why use hereditary hemochromatosis as a model?

- It is the most common autosomal recessive disorder affecting individuals of Northern European descent (about 1 in 200 Caucasians develop hereditary hemochromatosis).
- Effective methods for prevention and treatment are available with early diagnosis.
- It is caused by changes in the DNA sequence of a single gene, so the genetic basis of this disorder is easier to understand than for more complex disorders caused by alterations in multiple genes.
- The gene and its protein product are relatively well studied. Three-dimensional structures of the protein product are available in Protein Data Bank (PDB), the international repository for macromolecular structure data.

## Some basic concepts to understand before starting:

- Genes are the basic physical and functional units of heredity. Each gene is located on a particular region of a chromosome and has a specific ordered sequence of nucleotides (the building blocks of DNA).
- Central dogma of molecular biology: DNA → RNA → Protein
  - Genetic information is stored in DNA.
  - Segments of DNA that encode proteins or other functional products are called genes.
  - Gene sequences are transcribed into messenger RNA intermediates (mRNA).
  - mRNA intermediates are translated into proteins that perform most life functions.
- Eukaryotic genes have introns and exons. Exons contain nucleotides that are translated into amino acids of proteins. Exons are separated from one another by intervening segments of junk DNA called introns. Introns do not code for protein. They are removed when eukaryotic mRNA is processed. Exons make up those segments of mRNA that are spliced back together after the introns are removed; the intron-free mRNA is used as a template to make proteins.
- Special cellular components called ribosomes use the triplet genetic code to translate the nucleotides of a mRNA sequence into the amino acid sequence of a protein. The Table of Standard Genetic Code is provided in the back of this workbook.
- There are 20 different amino acids. Proteins are created by linking amino acids together in a linear fashion to form polypeptide chains.
- Protein polypeptide chains fold into three-dimensional structures that can associate with other protein structures to perform specific functions.

## Activity #1

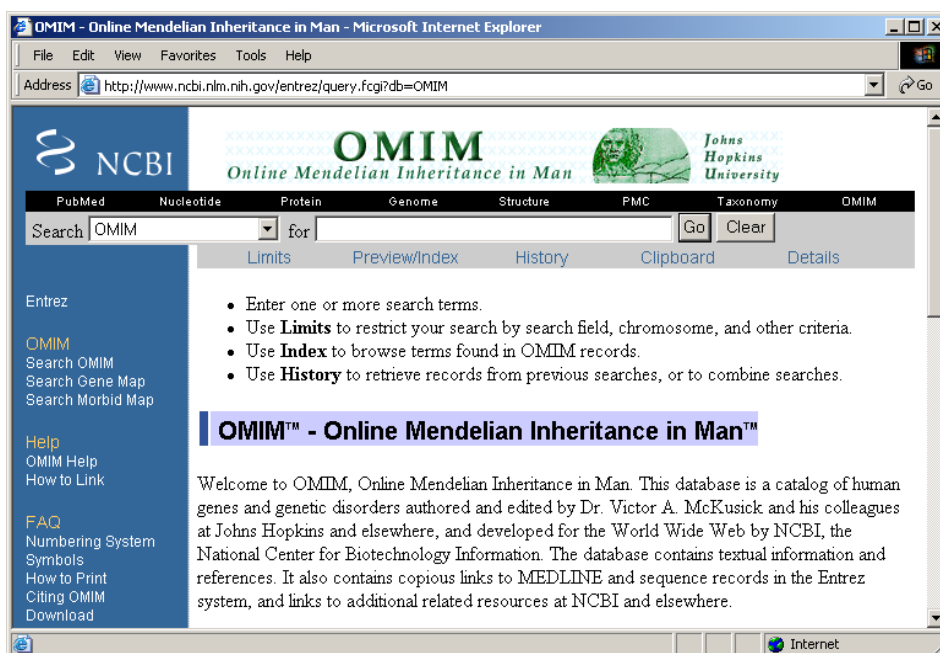
### Online Resources: OMIM and GeneReviews

- Learn about the genetic disorder and its associated gene.
- Identify the most common disease-causing mutation.

### Online Mendelian Inheritance in Man (OMIM)

OMIM is a large, searchable, up-to-date database of human genes, genetic traits, and disorders created and edited by Victor A. McKusick, M.D. and his colleagues at Johns Hopkins School of Medicine. The OMIM database is accessible through the National Center for Biotechnology Information (NCBI) suite of online resources. Each record in OMIM summarizes research defining what is currently known about a particular gene, trait, or disorder.

To access OMIM, let's go to the NCBI Web site (<http://www.ncbi.nlm.nih.gov/>), and then click on OMIM in the search box at the top. A screenshot of the OMIM home page is shown below.



The URL for the OMIM home page is: <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

Although the easiest way to search OMIM is to simply type a disorder name in the search box at the top, another option for searching OMIM is to use search field qualifiers. By adding search field qualifiers in square brackets to each search term and combining terms using Boolean operators (OR, AND, or NOT), you can execute a much more specific search in a single step.

This activity only demonstrates how to use a couple of OMIM's field qualifiers. More information about field qualifiers and other search options is available from OMIM Help (<http://www.ncbi.nlm.nih.gov/Omim/omimhelp.html>). In addition to OMIM, field qualifiers can be used to search other NCBI databases such as PubMed, Gene, and the nucleotide and protein sequence databases. Check the Help documentation of these resources to find out more.

Most of the genes, disorders and traits listed on the Human Genome Landmarks (HGL) poster were taken from the title fields of OMIM records. The field qualifier for the title field is [TI] or [TITL]. Since we selected our disorder from the HGL poster, we also know that hemochromatosis is found on chromosome 6. The field qualifier for specifying a particular chromosome is [CH] or [CHR].

1. To use a field qualifier in your search, simply add the qualifier to the end of your search term. For example, to search for hemochromatosis on chromosome 6 enter **hemochromatosis[TI] AND 6[CHR]** as shown in the search box below. Be sure to capitalize any Boolean operator (AND, OR, and NOT) you use in your search statements. Click **Go** to submit your search.

PubMed Nucleotide Protein Genome Structure PMC  
 Search OMIM for hemochromatosis[TI] AND 6[CHR] Go Clear

**NOTE:** Limiting a search to a particular chromosome may not work for disorders caused by alterations in multiple genes, such as breast cancer or colon cancer. These disorders are linked to genes on several different chromosomes; therefore, limiting your search to just one chromosome may not yield any results.

2. The search should return one result. Clicking on the MIM number [+235200](#) opens the full OMIM record for hemochromatosis shown below.

OMIM - HEMOCHROMATOSIS; HFE - Microsoft Internet Explorer  
 Address: http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=235200

NCBI  
 OMIM  
 Online Mendelian Inheritance in Man  
 Johns Hopkins University

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM  
 Search OMIM for [ ] Go Clear  
 Limits Preview/Index History Clipboard Details

Display Detailed Show: 20 Send to Text

[+235200](#) GeneTests, Links  
**HEMOCHROMATOSIS; HFE**

*Alternative titles; symbols*

HLAH  
 HEMOCHROMATOSIS, HEREDITARY; HH  
 HEMOCHROMATOSIS GENE, INCLUDED; HFE, INCLUDED

Gene map locus [6p21.3](#)

**TEXT**

**DESCRIPTION**

The clinical features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatocellular carcinoma (HCC; [114550](#)),



### 3. Let's examine some of the features of this record:

- Each record includes a blue navigation menu on the left with quick links to different sections within the record.
- Each OMIM record is assigned a unique six-digit MIM number located at the top of each entry. For hereditary hemochromatosis, the MIM number is 235200. As a unique identifier, the MIM number can be used to search other databases for information about a particular disorder. Clicking on the MIM number link will open the record in a simpler, frame-free format that is more suitable for printing. The plus sign (+) in front of the MIM number means that this entry refers to a gene with a known sequence and phenotype.
- A number sign (#) in front of the six-digit MIM number means that there is evidence that mutations in multiple loci may contribute to the development of this condition. For additional information about the symbols preceding MIM numbers, see OMIM Frequently Asked Questions ([http://www.ncbi.nlm.nih.gov/Omim/omimfaq.html#mim\\_number\\_symbols](http://www.ncbi.nlm.nih.gov/Omim/omimfaq.html#mim_number_symbols)).
- Below the MIM number, you will find the disorder name and the official gene symbol. The official gene symbol, which is **HFE** for hemochromatosis, serves as a unique identifier for a gene. To be "official," a gene symbol must have been approved by the HUGO Gene Nomenclature Committee (<http://www.gene.ucl.ac.uk/nomenclature/>). **The gene symbol is especially useful when searching other databases (such as sequence, genome-mapping, and structure databases) for gene-specific information.**



**NOTE:** For single-gene disorders like hemochromatosis, the official gene symbol will be included in the record title. For complex disorders like breast cancer, official symbols for associated genes will be described in the first paragraph of text.

- The gene map locus describes where a gene can be found on a chromosome. For the gene locus **6p21.3**, 6 is the chromosome number, p indicates the short arm of the chromosome, and 21.3 is a number assigned to a particular region of the chromosome. The gene map locus links to OMIM's Gene Map, a table of genes organized by cytogenetic location.
- The amount of text within an OMIM record varies depending upon what is known about a particular gene, disorder, or trait. Since hemochromatosis is well studied, there is a lot of information about this disorder and its gene. Some of the different types of information that may be included in an OMIM record are disorder description, nomenclature, clinical features, mapping, genotype and phenotype correlations, diagnosis, population genetics, gene structure, gene function, and animal models.
- If available, selecting the **Gene Structure** link (see the blue navigation column on left) provides information about the size and number of exons in the gene.
- Although it is not a part of every OMIM record, another useful section is **Allelic Variants** (see the blue navigation column on left). This section typically describes some of the most notable gene mutations associated with the development of disorders. Select the **View List** link under **Allelic Variants** to see a listing of important mutations that have been identified for this gene.

In addition to OMIM, another good resource for learning about genetic disorders and associated genes is GeneReviews, which is described in the next part of this activity.

## GeneReviews from GeneTests.org

GeneReviews is another excellent source for finding up-to-date, authoritative information on genetic disorders. All entries are written and reviewed by physicians, so the language is similar to that of medical text. While the amount and kind of content can vary greatly from record to record in OMIM, all reports in GeneReviews will provide similar kinds of information.

In addition to disease characteristics, diagnosis and testing, therapies, and counseling, each entry contains a **Molecular Genetics** section that describes the science behind disease development. This section identifies the mode of inheritance, the gene(s) involved, normal and abnormal gene products, and links to genomic databases.

Let's go to the GeneTests Web site to find a GeneReview for hereditary hemochromatosis:

<http://www.genetests.org/>

GeneTests Home Page - Microsoft Internet Explorer

Address: <http://www.genetests.org/>

Home Page | About GeneTests | **GeneReviews** | Laboratory Directory | Clinic Directory | Educational Materials

Funded by the National Institutes of Health

**10/05/04**

- 265 *GeneReviews*
- 1,098 Clinics
- 588 Laboratories testing for
- 1,064 Diseases
  - 734 Clinical
  - 330 Research only
- 32,330 Entries viewed per day

[More usage statistics](#)

Welcome to the **GeneTests** Web site, a publicly funded medical genetics information resource developed for physicians, other healthcare providers, and researchers, available at no cost to all interested persons. Use of this Web site assumes acceptance of the [terms of use](#).

Direct questions to [genetests@genetests.org](mailto:genetests@genetests.org).

**At This Site**

- ▶ **GeneReviews**  
Online publication of expert-authored disease reviews
- ▶ **Laboratory Directory**  
International directory of genetic testing

**What's New**

**New Features**

- ▶ Updated PowerPoint® Presentation
- ▶ New Service Listings in the Clinic Directory
- ▶ PowerPoint® Slide Show: *GeneReviews*
- ▶ Links to Laboratory Web Sites

**New in GeneReviews**

**New Lab Listings**

- ▶ 7 new listings

1. Click on **GeneReviews** in the navigation bar at the top.
2. Once you get to the **Search by Disease** screen at **GeneReviews**, enter **hemochromatosis** into the search box.
3. Beside the search result "HFE-Associated Hereditary Hemochromatosis," select the **Reviews** link to access the hereditary hemochromatosis review shown on the next page.

The screenshot shows a Microsoft Internet Explorer browser window displaying the GeneReviews website. The address bar shows the URL: <http://www.genetests.org/servlet/access?db=geneclinics&site=gt&id=888891&key=G4cheu3P5Re5x&gry=&fcn=y&fw=7P1P&filename=/profiles/hemochromatosis/>. The page features a navigation menu with links for Home Page, About GeneTests, GeneReviews, Laboratory Directory, Clinic Directory, and Educational Materials. The main content area is titled "HFE-Associated Hereditary Hemochromatosis" and includes a "Printable Copy" link. The page is funded by the NIH and developed at the University of Washington, Seattle. The authors listed are Kris V Kowdley, MD; Jonathan F Tait, MD, PhD; Robin L Bennett, MS; and Arno G Motulsky, MD. The page also provides information on the initial posting date (3 April 2000) and the last revision date (13 September 2004). A sidebar on the left contains a list of links for Summary, Diagnosis, Clinical Description, Differential Diagnosis, Management, Genetic Counseling, Molecular Genetics, Resources, References, Author Information, and Top of Page. At the bottom of the page, there is a section for "Disable Glossary" and "Title Index".

4. Access and read the **Summary, Diagnosis, Management, and Molecular Genetics** sections of this report. Use the information in GeneReviews and OMIM to answer the Questions for Activity #1 on the Hereditary Hemochromatosis Worksheet included in the back of this workbook.



## Activity #2

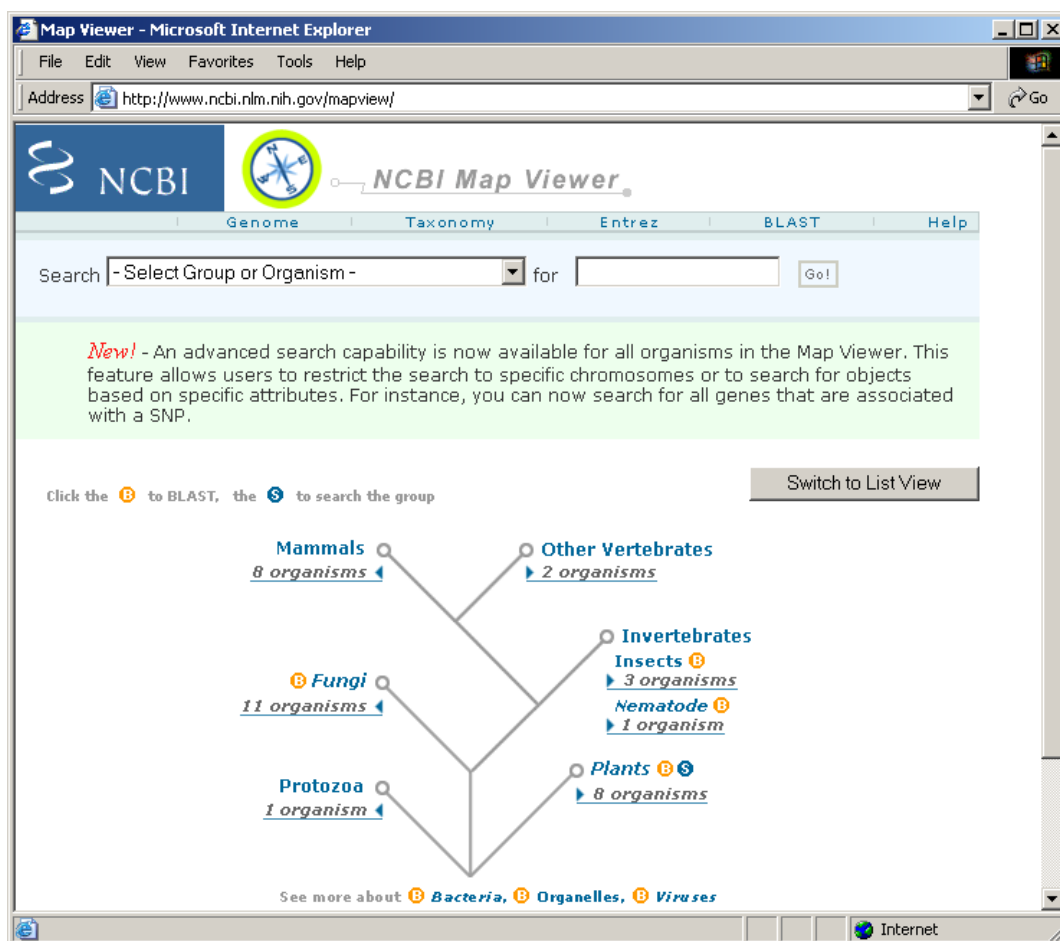
### Online Resource: NCBI Map Viewer

- Find a gene locus on a chromosome map.

NCBI Map Viewer is a Web-based tool for viewing and searching an organism's complete genome. Users can also view maps of individual chromosomes and zoom into specific regions within chromosomes to explore the genome at the sequence level.

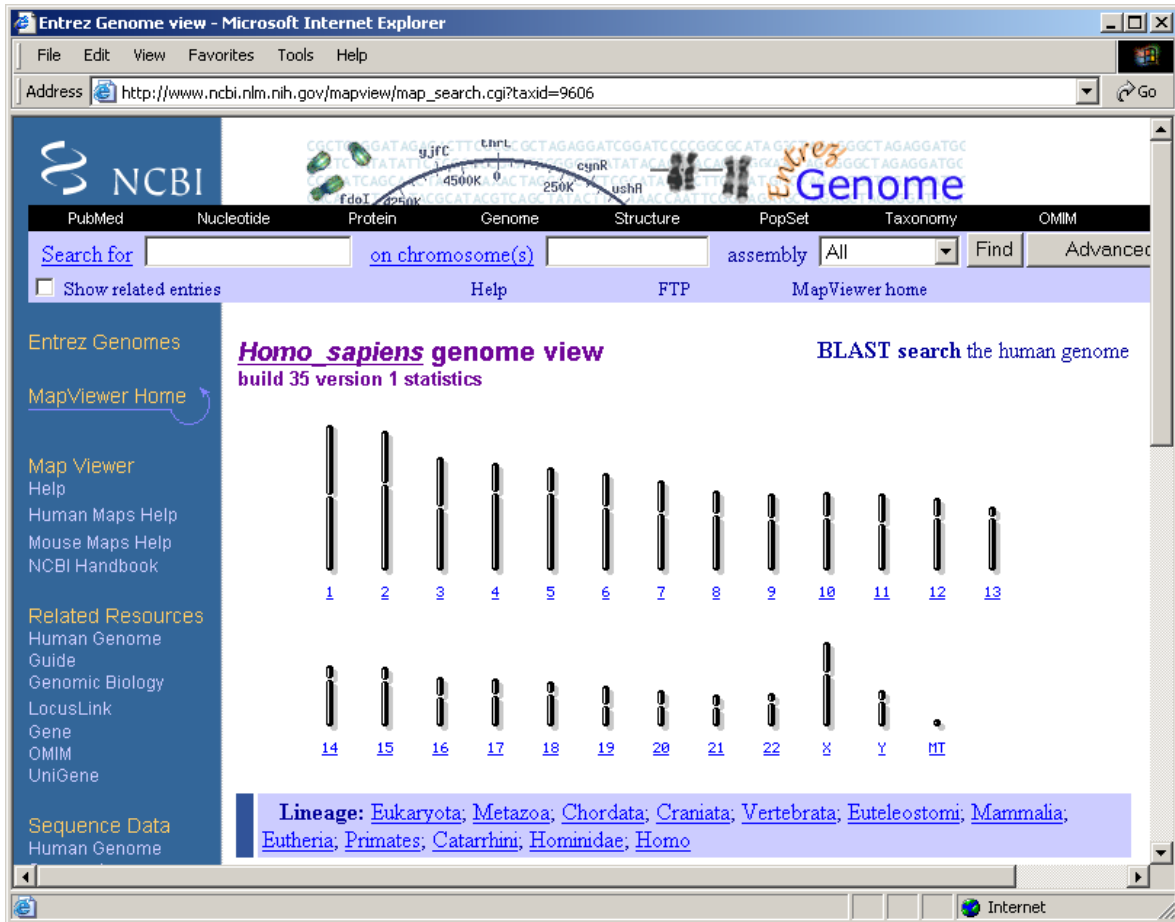
Map Viewer provides access to several different types of maps for different organisms. Many of these maps are meaningful only to scientific researchers. A discussion of all the different types of maps and genomic data is beyond the scope of this activity, which will focus only on how to locate a specific gene locus on a chromosome map.

From the NCBI home page (<http://www.ncbi.nlm.nih.gov/>), select **Map Viewer** from the alphabetized list of "Hot Spots" on the right.



The URL for NCBI Map Viewer is: <http://www.ncbi.nlm.nih.gov/mapview/>

On the Map Viewer home page, click on **Mammals**, and select **Homo sapiens (human)** from the pop-up menu to view the entire human genome. This will launch the *Homo sapiens* genome view shown in the following screenshot.



The URL for the *Homo sapiens* genome view is: [http://www.ncbi.nlm.nih.gov/mapview/map\\_search.cgi](http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi)

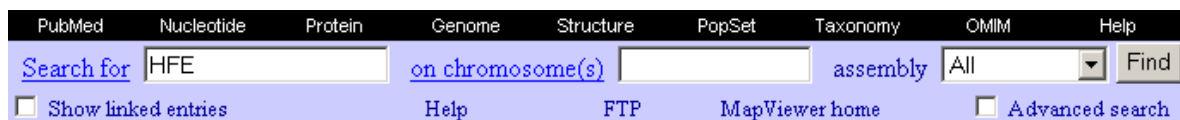
In Activity #1, we learned that the official symbol for the hereditary hemochromatosis gene is HFE, and its locus is 6p21.3. Let's find this locus on chromosome 6.

**What is a locus?**

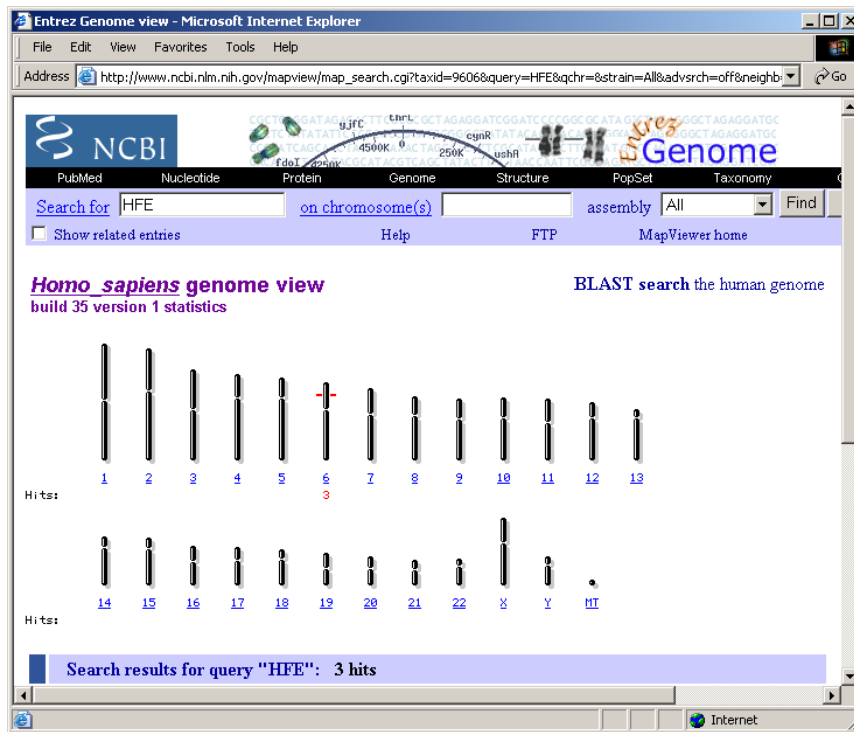
A locus describes the region of a chromosome where a gene is located. **6p21.3** is the locus for the HFE gene. **6** is the chromosome number, **p** indicates the short arm of the chromosome, and **21.3** is the number assigned to a particular region on a chromosome. When chromosomes are stained in the lab, light and dark bands appear, and each band is numbered. The higher the number, the farther away the band is from the centromere.

**Short and Long Arms of a Chromosome**

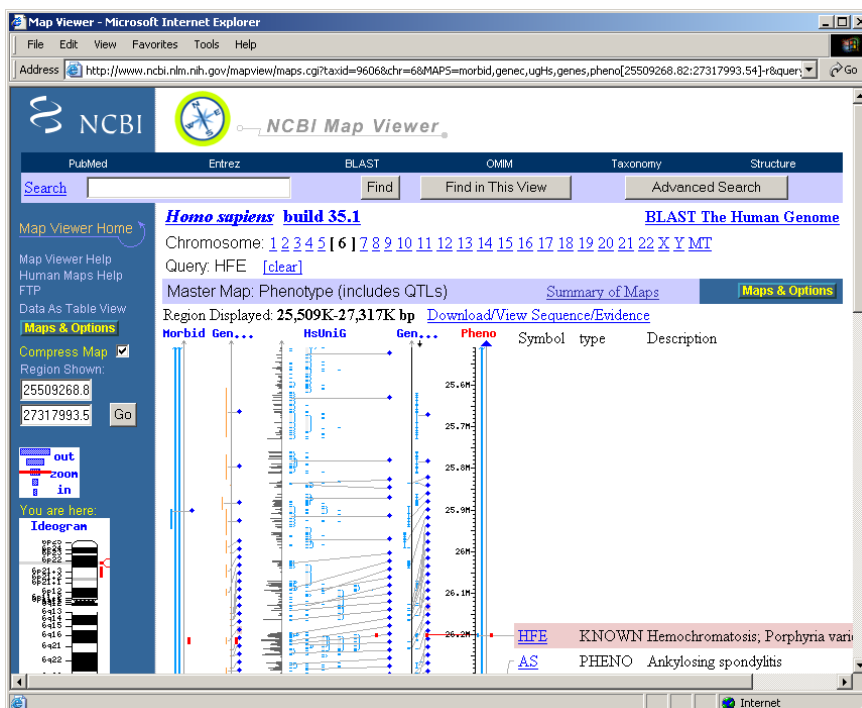
1. In the text box at the top of the page search for **HFE** as shown in the search box below.



2. Two red tick marks should be displayed on chromosome 6 in the genome view indicating the approximate location of HFE gene. The “3” below chromosome 6 indicates that there were three hits for our query. See the following screenshot.



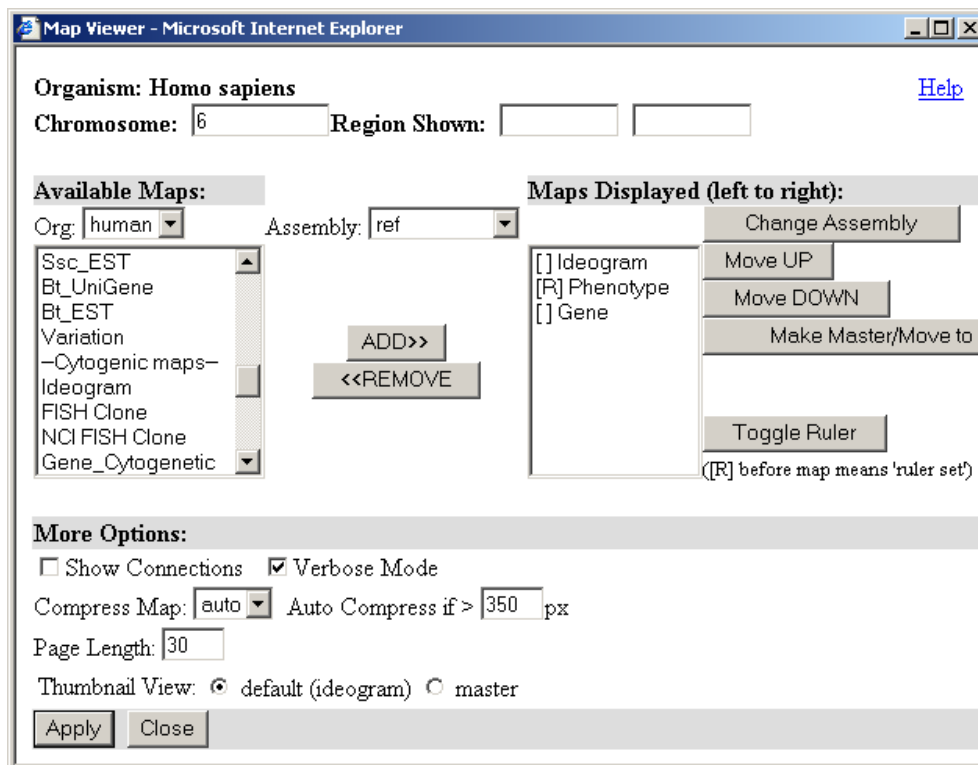
3. In the genome view click on the number [6](#) link below the chromosome. This will open a view of chromosome 6 that should look like the screenshot below. In the next step we will modify this view so that we can see an ideogram of chromosome 6 marked with the HFE gene locus.



3. Let's modify the display options by clicking on **Maps & Options**. This will open a window for customizing map options. Make the following adjustments. Before clicking the **Apply** button, your options window should resemble the screenshot below.

- Delete the numbers defining the **Region Shown** at the top of the options window.
- Remove all the maps listed under **Maps Displayed**, except the **Gene** and **Phenotype** maps.
- Under **Available Maps** select **Ideogram** and click the **Add** button. The **Gene** map should be designated as your Master Map. To make a map the master, simply highlight it and click the **Make Master/Move to Bottom** button. In the chromosome view, a master map is shown at the right edge of the display along with its details and descriptive text.

How the Maps & Options window should look:



- Click **Apply** at bottom and close screen.

#### About the maps:

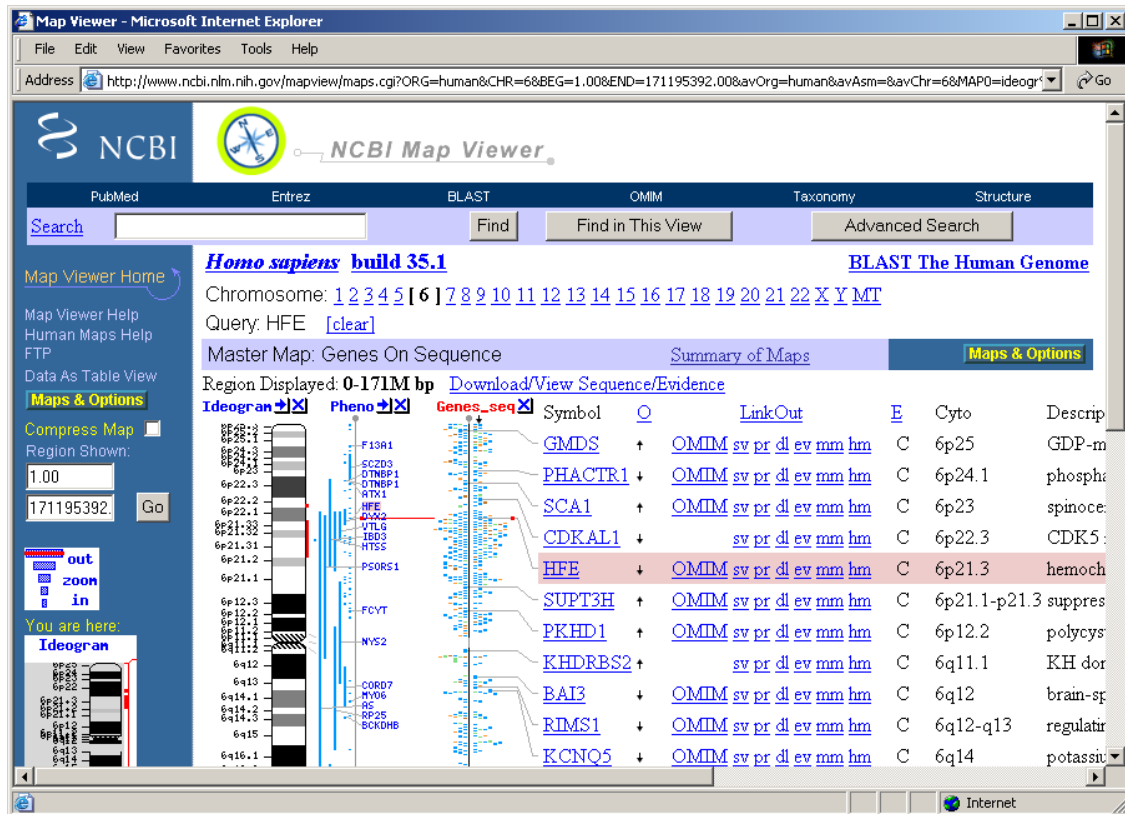
**Ideogram** - Shows the G-banding pattern of a chromosome at 850 band resolution.

**Phenotype** – Shows where loci associated with phenotypes can be found on the assembled human genome sequence. Gene symbols on this map link to OMIM.

**Gene** - Includes known and putative genes that have been annotated on genomic contigs (a group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome).

4. The new map of chromosome 6 should resemble the following screenshot. Make a note of where 6p21.3 is on the ideogram of chromosome 6.





Features of the **Genes\_seq** map (the master map in the screenshot above):

- Clicking on the Ideogram, Pheno, or **Genes\_seq** maps will open a pop-up window with options for zooming in on the displayed maps. You can also zoom in and out using the zoom option in the blue navigation column on the left.
- By default, Map Viewer displays 20 genes at a time on the **Genes\_seq** map. To see a more complete listing of genes on the chromosome, select the **Data As Table View** link above **Maps & Options** in the blue navigation column on the left.
- The Genes\_seq map provides a LinkOut feature that links you from Map Viewer to gene-specific entries in other NCBI resources.
  - **HFE** – Links to the HFE entry in NCBI's Entrez Gene database that brings together a variety of gene-specific information together in one interlinked system.
  - **OMIM** – Links to the HFE entry in the Online Mendelian Inheritance in Man (OMIM) database covered in Activity #1.
  - **sv** – The Sequence Viewer link lets you drill down to the sequence level of the genomic contig that contains a particular gene. A contig is a group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome. This link takes you to a graphic showing the position of the gene within the genomic sequence.
  - **pr** – Links to the reference sequence of the gene's protein product.

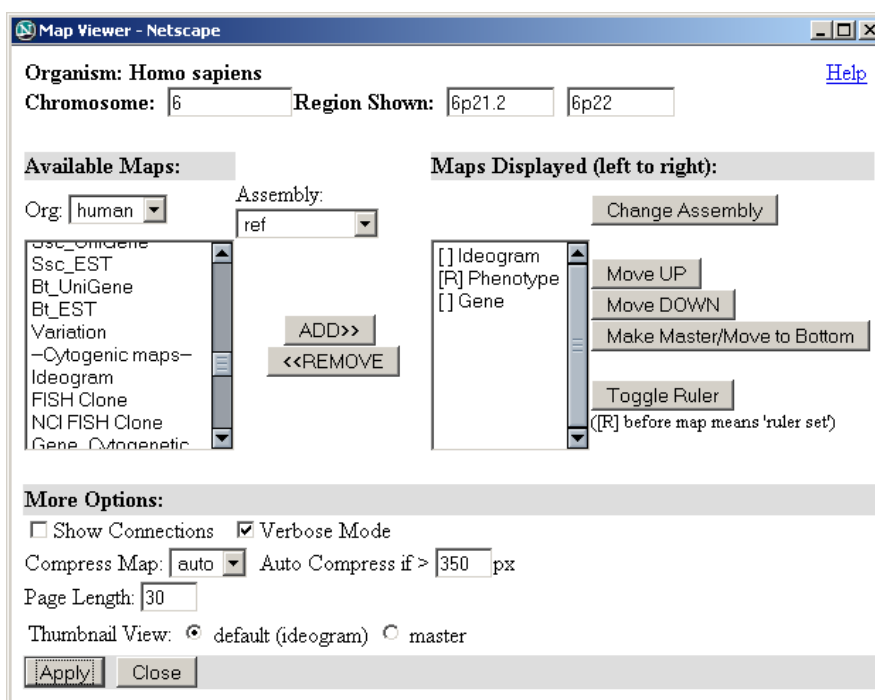
- [dl](#) – Links to a page for downloading the sequence data for a particular chromosome region.
- [ev](#) – Links to Evidence Viewer, which provides the biological evidence supporting a particular gene model. It displays all RefSeq models, GenBank mRNAs, known or potential transcripts, and ESTs (expressed sequence tags) that align to the area of interest.
- [mm](#) – Links to Model Maker, which allows you to view the evidence that was used to build a gene model based on assembled genomic sequence. You can also create your own version of a model by selecting exons of interest.
- [hm](#) – Links to Homologene, a resource for comparing genes in homologous segments of DNA from different organisms.

### Zooming in on region 6p21.3:

5. The locus 6p21.3 is flanked by the 6p21.2 and the 6p22 bands. Let's zoom in on the 6p21.3 locus.

- Click on **Maps & Options** again to open the options window.
- For **Region Shown:** enter 6p21.2 in the first box and 6p22 in the second box.

How Maps & Options window should look:



- Click the **Apply** button at the bottom, and close the box.

6. Scroll down to the bottom of the map to examine the **View Summary** information. Use this information and the new map of chromosome 6 to answer the questions for Activity #2 on the Hereditary Hemochromatosis Worksheet.

## Activity #3

### Online Resources: Entrez Gene, GenBank, and GenAtlas

- Examine the sequence and structure of a gene.

This activity covers how to use NCBI's Entrez Gene (formerly LocusLink) to access the genomic DNA sequence of the hereditary hemochromatosis gene. We will examine some of the features of a record from NCBI's GenBank and then use GenAtlas to learn about the structure (e.g., intron and exon composition, coding sequence) of a gene.

In sequence databases such as GenBank, genomic DNA sequences from eukaryotic organisms contain both exons and introns, while mRNA sequences are intron-free DNA sequences. We will look at the genomic sequence of a gene to get a better understanding of its structure.

Entrez Gene is a NCBI resource that serves as a single-query interface for accessing sequence and descriptive information for specific genes defined by available sequence data.

To begin, let's go to the Entrez Gene home page.

<http://www.ncbi.nih.gov/entrez/query.fcgi?db=gene>

Gene - Microsoft Internet Explorer

Address <http://www.ncbi.nih.gov/entrez/query.fcgi?db=gene>

NCBI Entrez Gene

Entrez PubMed Nucleotide Protein Genome Structure PMC Taxonomy Books OMIM

Search Gene for  Go Clear  current records only

Limits Preview/Index History Clipboard Details

Entrez  
SITE MAP  
Entrez Help

Gene  
Search  
Gene Help

FAQ

FTP site

Related sites  
Entrez Genome  
Genomic Biology  
HomoloGene  
LocusLink  
Map Viewer  
OMIM  
RefSeq  
UniGene

Feedback

- Enter one or more search terms.
- More information about available fields is available [here](#).
- Consider use of the limits and preview/index functions.
- Remember, boolean operators (AND, OR, NOT) must be in uppercase.

**Gene** Background

Gene provides a unified query environment for genes defined by sequence and/or in NCBI's Map Viewer. You can query on names, symbols, accessions, publications, GO terms, chromosome numbers, E.C. numbers, and many other attributes associated with genes and the products they encode.

Because Gene is now an Entrez database, all the familiar and useful functions are now available, including Preview/Index, History, and LinkOut.

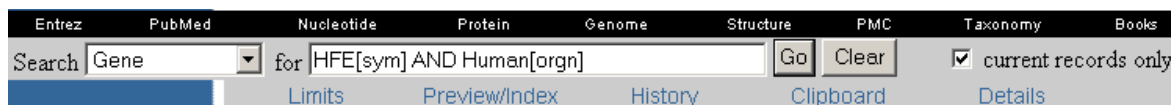
**Please note:** Entrez Gene is under active development. We welcome your [suggestions](#). We have also added a choice in our Feedback/Corrections [form](#) for suggestions; both paths reach the same staff.

[Getting started](#) Sample queries

Look for genes by name part and multiple species [transporter AND \("Drosophila melanogaster"\[orgn\] OR "Mus musculus"\[orgn\]\) more...](#)

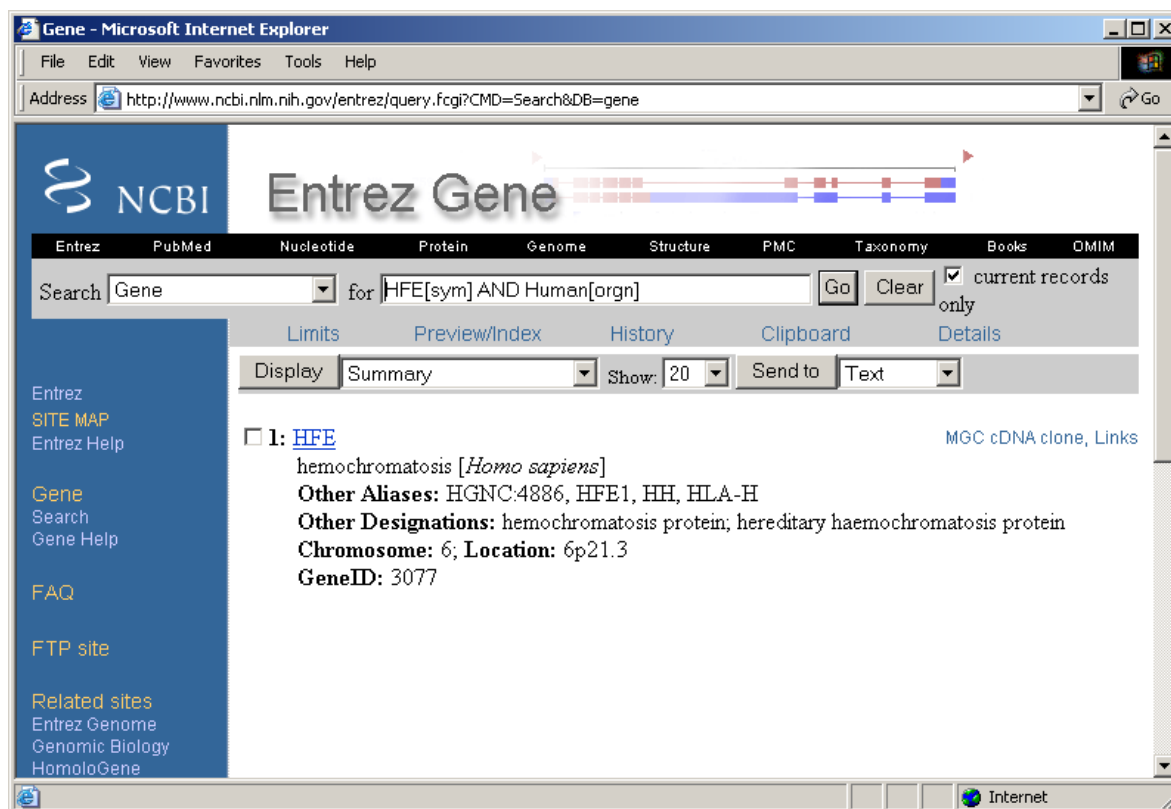
Look for genes by chromosome and symbol [\(II\[chr\] OR 2\[chr\]\) AND adh\\*\[sym\] more...](#)

1. In the search box at the top of the page, enter **HFE[sym] AND Human[orgn]**.

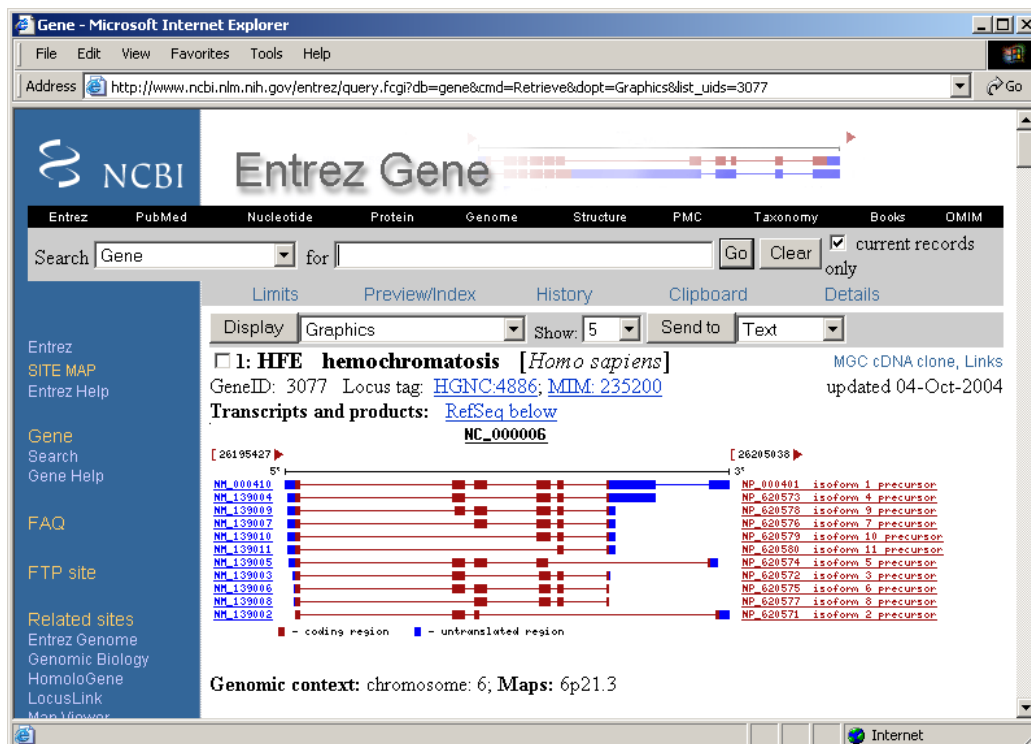


**Search Tip:** Adding [sym] to the end of your query term tells Entrez Gene that you are searching by gene symbol only. If you do not specify that you want to search the gene symbol field, the search will return multiple records that include the query term anywhere within its text. Adding [orgn] to a search term limits the search to genes from a specific organism. For more information on options for refining your search, see the Search Field Descriptions and Qualifiers section of Entrez Help: [http://www.ncbi.nlm.nih.gov/entrez/query/static/help/Summary\\_Matrices.html](http://www.ncbi.nlm.nih.gov/entrez/query/static/help/Summary_Matrices.html)

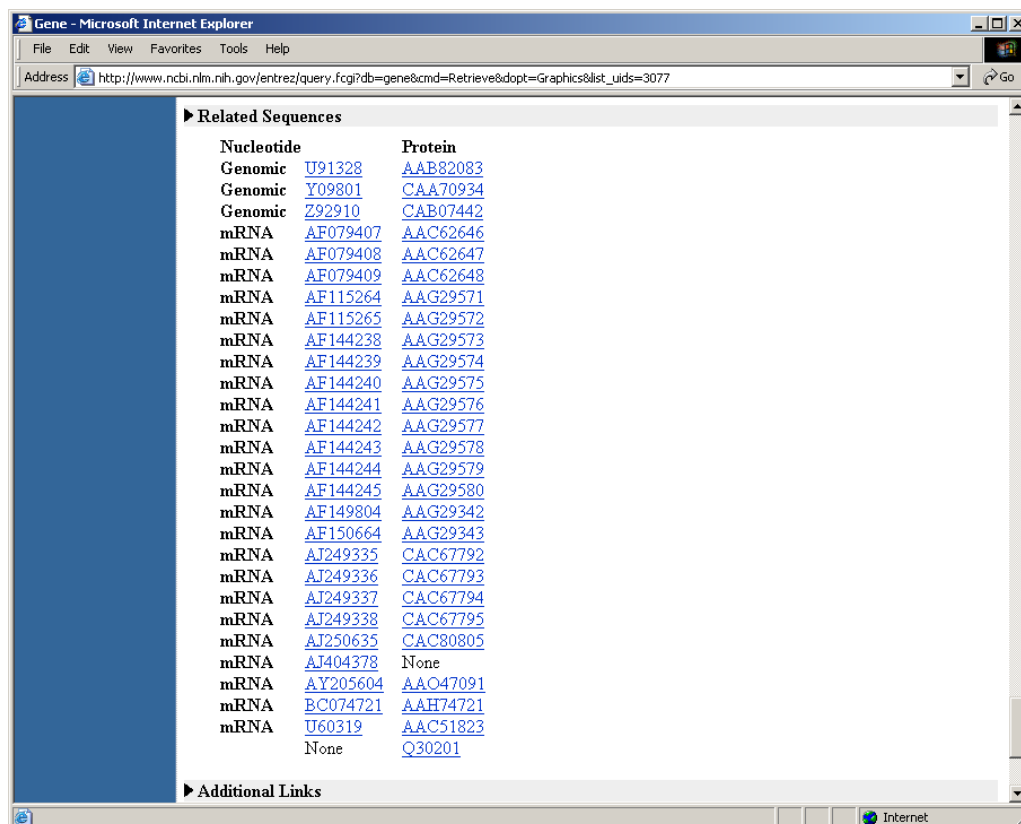
2. Submitting this search should retrieve a single result:



3. Click on the gene symbol [HFE](#) to pull up the Entrez Gene record. The record for HFE is shown on the following page. In the **Transcripts and products:** section at the top of this record (see the screenshot on the following page), notice that there are multiple transcripts represented by diagrams. This means that for the HFE gene, there can be more than one mRNA transcript. For example, an exon included in one transcript might be left out in another transcript. This is described in the **Summary:** section (you will need to scroll down to find this section). Use the information provided in the **Summary:** to answer Question 1 for Activity #3.



4. Scroll down and browse the contents of this record. Near the bottom of the page, you will see to the **Related Sequences** section shown below.



5. To find a genomic sequence record (including both introns and exons) for HFE, in the **Related Sequences** section, select a nucleotide sequence number labeled "Genomic." Click [Z92910](#) to open this GenBank sequence record. A screenshot of this record is shown below.

#### How did you know which genomic sequence to select?

The problem with archival sequence databases like NCBI's GenBank is that they usually have multiple sequence records for the same gene. You may need to open each record individually and browse through definition, sequence annotation, and comments to figure out exactly what nucleotide sequence is contained within each record.

For example, the [U91328](#) record contains the sequence of a genomic segment that not only includes the HFE gene sequence but also sequences for the histone 2A-like protein gene, RoRet gene, and sodium phosphate transporter (NPT3) gene. [Y09801](#) contains only sequence information for the HFE promoter and the HFE gene's first exon. Of the three genomic records listed for HFE, only [Z92910](#), which has seven exons and six introns, is the best candidate for examining HFE's genomic sequence.

GenBank Record [Z92910](#) - The genomic sequence of the human HFE gene.

The screenshot displays the NCBI Sequence Viewer interface for the human HFE gene (Z92910). The browser window title is "NCBI Sequence Viewer - Microsoft Internet Explorer". The address bar shows the URL: <http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?db=Nucleotide&dopt=GenBank&val=1890179>. The main content area shows the NCBI logo and a search bar with "Nucleotide" selected. Below the search bar are navigation tabs: Entrez, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and Books. The "Nucleotide" tab is active, and the search results show "1: Z92910. Homo sapiens HFE ...[gi:1890179]". The record details are as follows:

LOCUS HSHFE 12146 bp DNA linear PRI 23-JUL-1999  
 DEFINITION Homo sapiens HFE gene.  
 ACCESSION Z92910  
 VERSION Z92910.1 GI:1890179  
 KEYWORDS haemochromatosis; HFE gene.  
 SOURCE Homo sapiens (human)  
 ORGANISM [Homo sapiens](#)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 858)  
 AUTHORS Albig,W., Drabent,B., Burmester,N., Bode,C. and Doenecke,D.  
 TITLE The haemochromatosis candidate gene HFE (HLA-H) of man and mouse is  
 located in syntenic regions within the histone gene cluster  
 JOURNAL J. Cell. Biochem. 69 (2), 117-126 (1998)  
 MEDLINE [98208340](#)  
 PUBMED [9548560](#)  
 REFERENCE 2 (bases 1 to 12146)  
 AUTHORS Albig,W.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAR-1997) Albig W., Georg-August-Universitaet  
 Goettingen, Biochemie und Molekulare Zellbiologie, Humboldtallee  
 23, Goettingen, FRG, 37073  
 FEATURES Location/Qualifiers  
 source 1..12146

6. Scroll down the [Z92910](#) record to the **Features** section. The different features characterized for this gene are explained on the following page.

FEATURES	Location/Qualifiers
source	1..12146 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="6" /map="6p" /clone="ICRFy901D1223" /clone_lib="ICRF Y&C-library"
<a href="#">gene</a>	1028..10637 /gene="HFE"
<a href="#">exon</a>	1028..1324 /gene="HFE" /number=1
<a href="#">CDS</a>	join(1249..1324,4652..4915,5125..5400,6494..6769, 6928..7041,7995..8035) /gene="HFE" /function="iron metabolism" /note="haemochromatosis candidate gene" /codon_start=1 /protein_id="CAB07442.1" /db_xref="GI:1890180" /db_xref="GOA:Q30201" /db_xref="UniProt/Swiss-Prot:Q30201" /translation="MGPRARPALLLLMLLQTAVLQGRLLRSHSLHYLFMGASEQDLGL SLFEALGYVDDQLFVFDHESRRVEPRTPWVSSRISSQMWLQLSQSLKGDHMFVDF WTIMENHNHSKESHTLQVILGCEMQEDNSTEGYWKYGYDGDHLEFCPDTLDWRAAEP RAWPTKLEWERHKIRARQNRAYLERDCPAQLQQLLELGRGVLDQQVPLVKVTHVTS SVTTLRCRALNYYPQNIITMKWLKDKQPMDAKEFEPKDVLPNGDGTQGWITLAVPPGE EQRYTCQVEHPGLDQPLIVIEWEPSPSGLVIGVISGIAVFWVILF IGILF IILRRKQG SRGAMGHYVLAERE"
<a href="#">intron</a>	1325..4651 /gene="HFE" /number=1



Some of the features of the sequence in GenBank Record Z92910 include:

**source** - The source feature must be included in each sequence record. The source gives the length of the entire sequence included in the record and the scientific name of the source organism. Other types of information that the submitter may include in this field are chromosome number, map location, and clone or strain identification.

**gene** - Gives the numbers of the nucleotides where the gene stops and starts. **This link opens a new sequence record that shows only the gene sequence.**

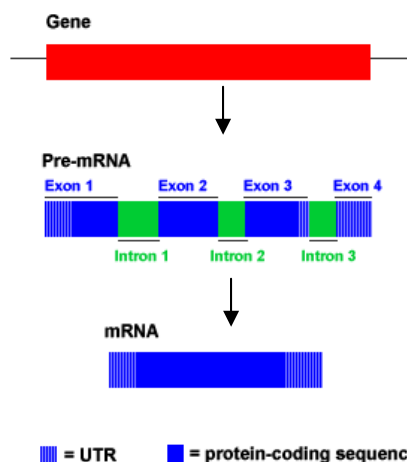
**exon** - Gives the numbers of the nucleotides where each exon begins and ends. You will see several of these entries as you scroll down. Each exon is a sequence segment that codes for a portion of processed (intron-free) mRNA. The name of the gene to which the exon belongs and the exon number are provided. **This link opens a new sequence record that shows only the exon sequence.**

**CDS** - The coding sequence consists of the nucleotides that actually code for the amino acids of the protein product. This feature includes the coding sequence's amino acid translation and may also contain gene name, gene product function, a link to protein sequence record, and cross-references to other database entries. **This link opens a new sequence record that shows only the coding sequence.**

**intron** - Gives the numbers of the nucleotides where each intron begins and ends. An intron is a segment of noncoding sequence that is transcribed but removed from the transcript by splicing together the exons (coding portions) on either side of it. **This link opens a new sequence record that shows only the intron sequence.**

### What's the difference between exons and coding sequence?

Most texts define exons as short segments of protein coding sequence. This is a bit of an oversimplification. Exons are those segments of sequence that are spliced together after the introns have been removed from the pre-mRNA. Yes, the coding sequence is contained in exons, but it is possible for some exons to contain no coding sequence. Portions of exons or even entire exons may contain sequence that is not translated into amino acids. These are the untranslated regions or UTRs. UTRs are found upstream and downstream of the protein-coding sequence. See diagram below.



7. Examine the reference section, the features section, and the sequence at the bottom of this record, and then answer questions 2–6 of the Questions for Activity #3. Questions 7–8 will be answered using GenAtlas, which is covered in the next section of this activity.

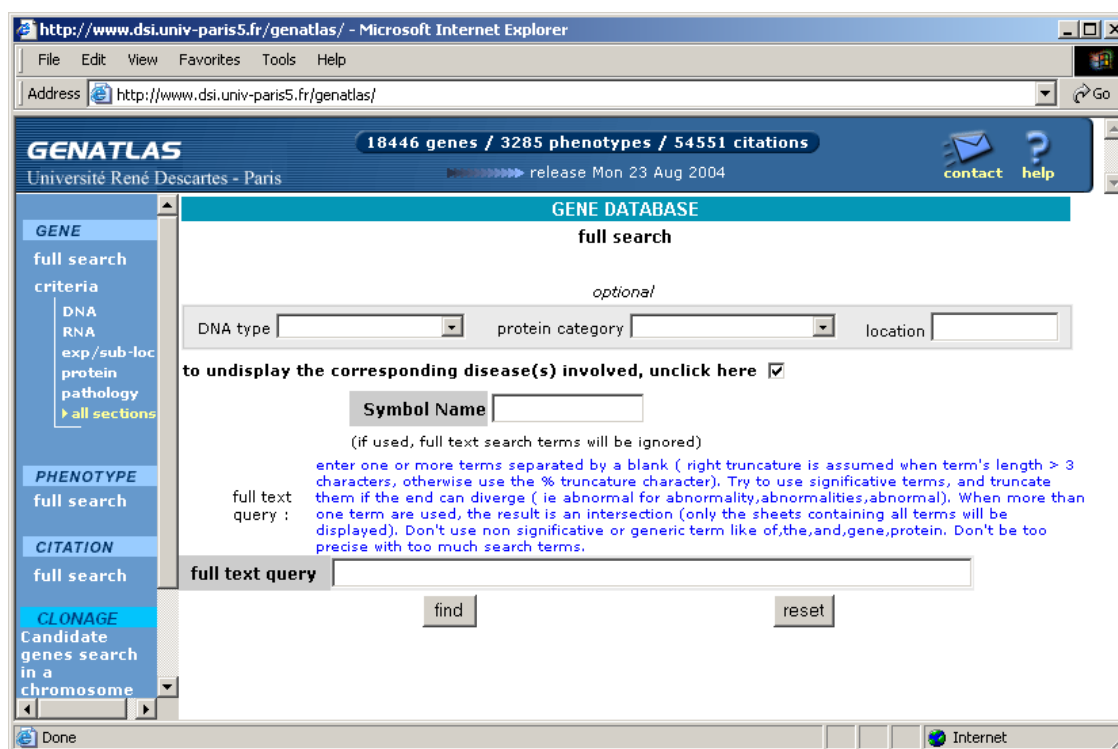


## Learning about Gene Structure with GenAtlas

GenAtlas is another useful tool for learning about the structure of genes. This resource is a compilation of Human Genome Project mapping information reported in the scientific literature. GenAtlas consists of three different types of databases: a gene database with more than 18,000 records, a phenotype database, and a citation database that contains references for the previous two databases. This activity will focus on using the gene database to learn about the coding and non-coding portions of genes.

Let's start by going to the GenAtlas home page:

<http://www.dsi.univ-paris5.fr/genatlas/>



1. To find the record for the hemochromatosis gene, enter the gene symbol **HFE** in the **Symbol Name** search box as shown below and click **Find** to submit your search.

**Symbol Name**

2. Select HFE from the search results to open the record. See a screenshot of the HFE record on the next page.

HFE record in GenAtlas:

**GENATLAS : GENE Database**

18446 genes / 3285 phenotypes / 54551 citations

Université René Descartes - Paris

release Mon 23 Aug 2004

**GENE**

full search

criteria

DNA  
RNA  
exp/sub-loc  
protein  
pathology  
all sections

**PHENOTYPE**

full search

**CITATION**

full search

**CLONAGE**

Candidate genes search in a chromosome region

Awstats user statistics

**FLASH GENE**

Symbol: HFE *last update : 9/10/2003*

Genew name: hemochromatosis

Corresponding disease: [HFE](#), [VGP](#)

Location: 6p21.3

Synonym symbol(s): HLA-H

	DNA	RNA	EXP/sub-loc	PROTEIN	PATHOLOGY
TYPE	functioning gene				
SPECIAL FEATURE	gene in gene, antisens HFE antisense RNA partially covers HFE mRNA				
STRUCTURE	9,6 kb 7 Exons				

present in the contig : [NT\\_007592](#) of Genbank

DNA size 9.61 Kb mRNA size 2717 bp 7 exons [NM\\_000410](#) [see the exons](#)

16945699 1 2 3 4 5 6 7 16955308

DNA size 7.94 Kb mRNA size 1922 bp 5 exons [NM\\_139004](#) [see the exons](#)

3. Examine the HFE record. As you scroll down the page you will notice that there are at least 11 variants for this gene. We saw this before when we were using Entrez Gene. Multiple mRNA variants for a gene indicate alternative splicing, which involves splicing the same mRNA transcript different ways to produce different proteins. For example, an exon included in one variant may be spliced out with the introns in another variant. Multiple protein products can be generated from alternative splicing of the same mRNA. **We will examine only the first variant (NM\_000410), which is the largest (7 exons), most inclusive splice variant for HFE.**

4. Click on [see the exons](#) for the first variant. This will take you to a page providing a color-coded breakdown of the DNA sequence of each exon (see the screenshot on the following page). The blue bases are adjacent intronic DNA. **Note that only portions of the intron sequence adjacent to the exons are shown.** Much of the intronic DNA sequence has been left out. Start and stop codons are in red. The bases of the exon sequence are capitalized in black. Bold portions of the exon sequence make up the coding sequence, and the unbolded portions do not code for protein.

Sequence of the first HFE Variant (**NM\_000410**) from GenAtlas

GENATLAS  
Université René Descartes - Paris  
18446 genes / 3285 phenotypes / 54551 citations  
release Mon 23 Aug 2004

Gene	HFE	Variant 1
DNA :	9.61 Kb	<a href="#">NT_007592</a>
mRNA :	2717 bp	<a href="#">NM_000410</a>
CDS :	1932 bp	<a href="#">NP_000401</a>

NON CODANT MRNA, **CDS**, [initiator and stop codon](#), [genomic and intronic adjacent sequences](#), [allelic variation](#).

**EXON 1 297 bp**  
 aaaaagctttatatttctaaagtcaaataagacataagttggtctaaagttgagataaaat  
 ttttaaatgtatgattgaattttgaaaatcataaatatttaaatatctaaagttcagatc  
 agaacattgccaagctactttccccaatcaacaacacccttcaggatttaaaaaccaag  
 GGGGACACTGGATCACCTAGTGTTCACAAGCAGGTACCTTCTGCTGTAGGAGAGAGAGA  
 ACTAAAGTCTGAAAGACCTGTTGCTTTTACCAGGAAGTTTTACTGGGCATCTCCTGAG  
 CCTAGGCAATAGCTGTAGGGTGACTTCTGGAGCCATCCCGTTTTCCCGCCCCCAAAG  
 AAGCGGAGATTTAACGGGGACGTGCGGCCAGAGCTGGGAAAT**GGGCCCGAGCCAGGC**  
 CGGGCTTCTCCTCCTGATGCTTTTGCAGACCGGGTCTGCAGGGGCGCTTGCTGC  
 gtgagtccgaggggctgcggcggaactagggcgcggggggtggaaaaatcgaaactag  
 ctttttcttgcgcttgggagtttgctaaacttggaggacctgctcaacctatccgcaa  
 gcccctctccctactttctgcgtccagacccctgagggagtgctaccactgaactgca

**EXON 2 264 bp**  
 aagcacacaaggaaagagcaccaggactgtcatatggaagaaagacaggactgcaactc  
 acccttcacaaaatgaggaccagacacagctgatggtatgagttgatgcaggtgtgtgga  
 gcctcaacatcctgctccctcctactacacatggttaaggcctgtgctctgtctccag  
 GTTCACACTCTCTGCACTACCTCTTCATGGGTGCCTCAGAGCAGGACCTTGGTCTTTCCT  
 TGTTTGAAGCTTTGGGCTACGTGGATGACCAGCTGTTGTTTCTATGATCATGAGAGTC  
 GCCGTGTGGAGCCCCGAACTCCATGGGTTTCCAGTAGAATTTCAAGCCAGATGTGGCTGC  
 AGCTGAGTCAGAGTCTGAAAGGGTGGGATCACATGTTCACTGTTGACTTCTGGACTATTA  
 TGGAAAATCACAACCACAGCAAGG  
 gtatgtggagagggggcctcaccttctgaggttgcagagcttttcatctttcatgca

5. Use the DNA information about first HFE Variant (**NM\_000410**) to answer questions 7–8 of the Questions for Activity #3 in the Hereditary Hemochromatosis Worksheet.

**NOTE:** While GENATLAS provides a nice color-coded break down of gene structure, it is always a good idea to compare the sequence information from this source with sequence data accessible from NCBI. It is possible for some resources to have errors in the sequence data they present, so crosschecking among multiple sources is recommended.



## **Activity #4**

### **Online Resource: Swiss-Prot**

- Access the amino acid sequence of a gene's protein product.

This activity covers how to use the Swiss-Prot protein sequence database to learn about the amino acid sequence and other features of the hereditary hemochromatosis protein.

The protein sequence database Swiss-Prot was developed by groups at the Swiss Institute of Bioinformatics (SIB) and the European Bioinformatics Institute (EBI). Swiss-Prot is noted for its detailed annotation (descriptions of protein function and labeling of domains and other key features within proteins) of protein sequence data. TrEMBL is a computer-annotated database companion to Swiss-Prot that holds sequence data until it can be properly annotated, reviewed, and added to Swiss-Prot.

Let's start by going to the Swiss-Prot home page:

<http://us.expasy.org/sprot/>



1. Scroll down to the **Access to Swiss-Prot and TrEMBL** section and select [Advanced search in Swiss-Prot and TrEMBL](#). A screen shot of the advanced search page is shown on the next page.

Swiss-Prot/TrEMBL Advanced Search - Microsoft Internet Explorer

Address <http://us.expasy.org/sprot/sprot-search.html>

**swissprot** **Swiss-Prot/TrEMBL Advanced Search**

This search program uses [SRS](#) to perform queries. Simpler forms are available to [search by description](#) or [by full text](#). Available connectors within a field are "&" (and), "|" (or) and "!" (but not). You can prefix your search terms by ! to specify "not" (this is not possible in SRS). Example queries:

- To retrieve all AP1 complex proteins from mouse (AP1S1, AP1G1, etc. but not MIAP1, IQGAP1, ...), specify *Gene Name: ap1\**, *Organism: Mus*, and deselect "Append and prefix \* to query terms".
- To retrieve the three human beta-adrenergic receptor proteins in Swiss-Prot, but not the beta-adrenergic receptor kinases, specify *Description: beta&adrenergic&receptor&kinase*, *Organism: Homo sapiens*, and select "Append and prefix \* to query terms".

Search  Swiss-Prot  TrEMBL

Description

AND  Gene name

AND  Organism  or choose from the list

Append and prefix \* to query terms

view of  results

The URL for Swiss-Prot/TrEMBL Advanced Search is: <http://us.expasy.org/sprot/sprot-search.html>

2. Scroll down to the search boxes. Remove the check in the box next to TrEMBL. We want only sequences from Swiss-Prot. In the **Gene name** search box enter **HFE**. In the **Organism** box enter **human**. Submit your query.

3. You should retrieve two results.

Swiss-Prot gene name: hfe - Microsoft Internet Explorer

Address <http://us.expasy.org/cgi-bin/get-entries?db=sp&DE=&GNC=AND&GN=hfe&OC=human&wild=1&view=full&num=100>

ExPASy Home page Site Map Search ExPASy Contact us Swiss-Prot

Hosted by NCSC US Mirror sites: Australia Bolivia Canada China Korea Switzerland Taiwan

Search  Swiss-Prot/TrEMBL for  Go Clear

**Swiss-Prot gene name: hfe**

**Gene name and organism**

The following is a list of the 2 Swiss-Prot entries with the gene name *hfe* and from taxon *human*.

**Entries in Swiss-Prot:**

Send selected sequences to  Retrieve sequences (FASTA format)

	Entry name	AC	Gene names	Description	Organisms	Length
<input type="checkbox"/>	HFE_HUMAN	<a href="#">Q30201</a>	HFE, HLAH	Hereditary hemochromatosis protein precursor (HLA-H)	Homo sapiens (Human)	348
<input type="checkbox"/>	RGMC_HUMAN	<a href="#">Q6ZVN8</a>	HFE2, HJV, RGMC	Hemojuvelin precursor (Hemochromatosis type 2 protein) (RGM domain family member C)	Homo sapiens (Human)	426

<http://us.expasy.org/sitemap.html>

4. Select the AC number [Q30201](#) for the HFE-HUMAN entry. This record is shown below.

**NiceProt View of Swiss-Prot: Q30201**

Search:  for

**NiceProt View of Swiss-Prot: [Q30201](#)**

[Printer-friendly view](#) [Submit update](#) [Quick BlastP search](#)

[\[Entry info\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#) [\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

*Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.*

**Entry information**

Entry name	HFE_HUMAN
Primary accession number	Q30201
Secondary accession numbers	O75929 O75930 O75931 Q96KU5 Q96KU7 Q96KU8 Q9HC64 Q9HC68 Q9HC70 Q9HC83
Entered in Swiss-Prot in	Release 35, November 1997
Sequence was last modified in	Release 35, November 1997
Annotations were last modified in	Release 45, October 2004

**Name and origin of the protein**

Protein name	Hereditary hemochromatosis protein [Precursor]
Synonym	HLA-H
Gene name	Name: HFE Synonyms: HLAH

5. Look at the **Protein Name** field. Notice that this protein is designated as a precursor protein. This means that part of the protein chain needs to be cutoff by a proteolytic enzyme to form the “mature” functional protein.

6. Using the navigation links at the top of the record, go to the **Features** section.

**Features**

[Feature table viewer](#) [Feature aligner](#)

Key	From	To	Length	Description	FTId
SIGNAL	<a href="#">1</a>	<a href="#">22</a>	22		
CHAIN	<a href="#">23</a>	<a href="#">348</a>	326	HEREDITARY HEMOCHROMATOSIS PROTEIN.	
DOMAIN	<a href="#">23</a>	<a href="#">114</a>	92	EXTRACELLULAR ALPHA-1.	
DOMAIN	<a href="#">115</a>	<a href="#">205</a>	91	EXTRACELLULAR ALPHA-2.	
DOMAIN	<a href="#">206</a>	<a href="#">297</a>	92	EXTRACELLULAR ALPHA-3.	
DOMAIN	<a href="#">298</a>	<a href="#">306</a>	9	CONNECTING PEPTIDE.	
TRANSMEM	<a href="#">307</a>	<a href="#">330</a>	24	POTENTIAL.	
DOMAIN	<a href="#">331</a>	<a href="#">348</a>	18	CYTOPLASMIC TAIL.	
DISULFID	<a href="#">124</a>	<a href="#">187</a>			
DISULFID	<a href="#">225</a>	<a href="#">282</a>			
CARBOHYD	<a href="#">110</a>	<a href="#">110</a>		N-LINKED (GLCNAC...) (POTENTIAL).	
CARBOHYD	<a href="#">130</a>	<a href="#">130</a>		N-LINKED (GLCNAC...) (POTENTIAL).	
CARBOHYD	<a href="#">234</a>	<a href="#">234</a>		N-LINKED (GLCNAC...) (POTENTIAL).	
VARSPLIC	<a href="#">26</a>	<a href="#">114</a>		RSHSLHYLFMGASEQDLGLSLFEALGYVDDQLFVFDHES RRVEPTPVVSSRISSQNWULQLSQSLKGDHMFVDFWTIMENH NHSKE -> Q (in <a href="#">isoform 2</a> and <a href="#">isoform 4</a> ).	VSP_003218

7. Select the [Feature aligner](#) link. This will open a new screen with a list of selected features within the HFE protein. See the screenshot below.

Feature aligner - Microsoft Internet Explorer  
 Address: http://us.expasy.org/cgi-bin/aligner?Q30201

Search: Swiss-Prot/TrEMBL for [ ] Go Clear

### Feature aligner

Selected features of [Q30201](#) (HFE\_HUMAN) Hereditary hemochromatosis protein precursor (HLA-H) [Homo sapiens (Human)].

Key	Position	Length	Description
<input type="checkbox"/> CHAIN	<a href="#">23-348</a>	326	HEREDITARY HEMOCHROMATOSIS PROTEIN RLLRSHSLHY LFMGASEQDL GLSLFEALGY VDDQLFVFD HESRRVEPRT PUVSSRISSQ MWLQSQSLK GWDHMTVDF WTIMENHNHS KESHTLQVIL GCEMQEDNST EGYWKYGYDG QDHLEFCPDT LDWRAAEPRRA WPTKLEWERH KIRARONRAY LERDCPAQLQ QLELGRGVL DQQVPPPLVKV THHVTSSVTT LRCRALNYYP QNITMKWLKD KQPMDAKEFE PKDVLNPGDG TYQGWITLAV PPGEQRYTC QVEHPGLDQP LIVINEPSPS GTLVIGVISG IAVFVVLFI GILFIILRRK QGSRGAMGHY VLAERE
<input type="checkbox"/> DOMAIN	<a href="#">23-114</a>	92	EXTRACELLULAR ALPHA-1 RLLRSHSLHY LFMGASEQDL GLSLFEALGY VDDQLFVFD HESRRVEPRT PUVSSRISSQ MWLQSQSLK GWDHMTVDF WTIMENHNHS KE
<input type="checkbox"/> DOMAIN	<a href="#">115-205</a>	91	EXTRACELLULAR ALPHA-2 SHTLQVILGC EMQEDNSTEG YWKYGYDGQD HLEFCPDTLD WRAAEPRAMP TKLEWERHKI RARONRAYLE RDCPAQLQQL LELGRGVLDQ Q
<input type="checkbox"/> DOMAIN	<a href="#">206-297</a>	92	EXTRACELLULAR ALPHA-3 VPPLVKVTHH VTSSVTTLRC RALNYYPQNI TMKWLKDKQP MDAKEFEPKD VLPNGDGTQ GWITLAVPPG EEQRYTCQVE HPGLDQPLIV IW
<input type="checkbox"/> DOMAIN	<a href="#">298-306</a>	9	CONNECTING PEPTIDE EPSPSGTLV
<input type="checkbox"/> TRANSMEM	<a href="#">307-330</a>	24	(POTENTIAL) IGVISGIAVF VVILFIGILF IILR
<input type="checkbox"/> DOMAIN	<a href="#">331-348</a>	18	CYTOPLASMIC TAIL KRQSGRGAMG HYVLAERE

8. Notice that the protein chain includes only amino acids 23–348. The first 22 amino acids are not associated with any domains (domains are functional units within a protein). This portion of the protein sequence is cleaved from the larger precursor sequence to make the mature, functional HFE protein.

9. Swiss-Prot records are known for their detailed sequence annotation. Notice how each domain is broken down into segments of corresponding amino acids within the protein chain. Select the [23–348](#) position link to access a new page that shows this portion within the context of the entire protein sequence (see the screen shot on the next page).



Swiss-Prot: HFE\_HUMAN - Microsoft Internet Explorer

Address: <http://us.expasy.org/cgi-bin/sprot-ft-details.pl?Q30201@CHAIN@23@348>

ExpASY Home page Site Map Search ExpASY Contact us Swiss-Prot

Search:  for

## Swiss-Prot: HFE\_HUMAN

The section of the sequence [HFE\\_HUMAN](#) (Q30201) you have selected corresponds to:

CHAIN 23 348 HEREDITARY HEMOCHROMATOSIS PROTEIN.

In one-letter code:

```

1      11      21      31      41      51
1  MGPRARPALL LLMLLQTAVL QGRLLRSHSL HYLFGASEQ DLGLSLFEAL GYVDDQLFVF 60
61  YDHESRRVPE RTPWVSSRIS SQMWLQLSQS LKGWDHMFV DFWTIMENHN HSKESHTLQV 120
121 ILGCEMQEDN STEGYWRYGY DGDHLEFCP DTLDWRAAEP RAWPTKLEWE RHKIRARQNR 180
181 AYLERDCPAQ LQQLLELGRG VLDQQVPLV KVTHHTSSV TTLRCRALNY YPQNITMKWL 240
241 KDKQPMDAKE FEPRKDVLPNG DGTYQGWITL AVPPGEEQRY TCQVEHPGLD QPLIVWEPS 300
301 PSGLTVIGVI SGIAVFVVIL FIGILFIILR KRQSGRGAMG HYVLAERE

```

In three-letter code:

```

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
1 Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu 15
16 Gln Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu 30
31 His Tyr Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser 45
46 Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe 60

```

10. The selected section of the protein sequence is highlighted in red. Another nice feature is that the protein sequence is represented using both one-letter and three-letter amino acid abbreviations.

11. Select the [HFE\\_HUMAN](#) link at the top of the page to return to the main Swiss-Prot record for the HFE protein. Return to the Features section of the record.

12. Scroll down to the part of the features section that describes the amino acid position of the protein's secondary structures (e.g., STRAND, TURN, HELIX). You can use this information to figure out which parts of the amino acid sequence form beta-strands, alpha helices, or the turns between these units of secondary structure.

13. The sequence and feature information presented in this record will help you gain a better understanding of the protein structure examined in Activity #5. Continue with Activity #5 before answering the questions for activities #4 and #5 in the worksheet.



## **Activity #5**

### **Online Resources: Protein Data Bank and Protein Explorer**

- Explore the sequence and structure of the gene's protein product.

This activity demonstrates how to find molecular structure data in Protein Data Bank (PDB) and how to use Protein Explorer (PE) to view a three-dimensional representation of a protein. PDB is an international archive of 3D structural information for biological macromolecules. Protein Explorer is free software that uses a program called "Chime" to display three-dimensional molecular structures in your browser.

#### **Before you begin:**

**Download and install the Chime plug-in from the Chime Web site:**

<http://www.mdl.com/downloads/downloadable/index.jsp> (Free registration is required.)

The latest version of Protein Explorer works in Internet Explorer 6 (Windows only) as well as Netscape 4.7x (Windows or Macintosh PPC). **Check your browser to see if it is compatible.** If you are not sure what version your browser is, select **Help** from the menu bar at the top of your browser, and then select the **About Netscape** or **About Internet Explorer** option. If you do not have a compatible browser version, you can download the version you need from one of the following links.

Microsoft Internet Explorer home page

<http://www.microsoft.com/windows/ie/downloads/default.msp>

Netscape Archived Products

<http://wp.netscape.com/download/archive.html>

#### **Why use Protein Explorer to view 3-D structures?**

Protein Explorer (PE) shares a command language and many other powerful features with RasMol, however it has menu-based options for manipulating structures making it easier to use and more appropriate for occasional or more general users.

Protein Explorer was created by Eric Martz, a professor with the Department of Microbiology at the University of Massachusetts. For more information about Protein Explorer, see <http://proteinexplorer.org>.

#### **Some Protein Structure Basics**

- Proteins are created by linking amino acids together in a linear fashion to form polypeptide chains. There are 20 different amino acids. The amino acid sequence of a polypeptide chain is the **primary structure** of a protein.
- Amino acids have different chemical properties. For example, some amino acid residues are strictly hydrophobic ("water fearing") and must be protected from aqueous environments, while other amino acids are hydrophilic ("water loving"). The substitution of

just one amino acid for another with very different chemical properties can have serious consequences for a protein's structure and function.

- The folding of regions within the polypeptide chain into alpha helices and beta sheets is a protein's **secondary structure**.
- The packing of the entire polypeptide chain into a three-dimensional globular unit is a protein's **tertiary structure**.
- If a protein molecule is a complex of more than one polypeptide chain, then the complete structure of this molecule is called a protein's **quaternary structure**.
- A domain is a discrete portion of a protein with its own function and specific three-dimensional structure. The combination of domains in a single protein determines its overall function.
- Different parts of a polypeptide chain can be linked by disulfide bridges that form between two cysteine residues. Disulfide bridges (or disulfide bonds) stabilize a protein's three-dimensional structure. The loss of a disulfide bridge would be detrimental to a protein's overall structure.

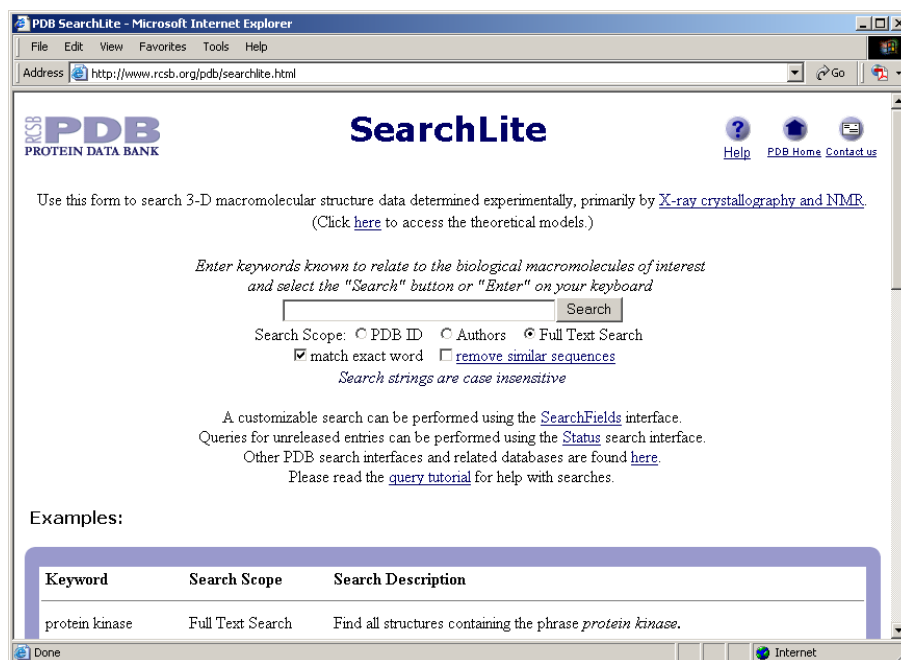
## Finding a Structure Record in PDB

To begin, we need to find a record for the hereditary hemochromatosis protein in PDB.

The URL for Protein Data Bank (PDB): <http://www.rcsb.org/pdb/>

The screenshot shows the RCSB Protein Data Bank website. The browser window title is "The RCSB Protein Data Bank - Microsoft Internet Explorer". The address bar contains "http://www.rcsb.org/pdb/". The main content area includes the PDB logo, a welcome message, and a search section titled "Search the Archive". The search section has a search box and options for "PDB ID", "Authors", and "Full Text Search". There are also links for "QuickSearch!", "SearchLite", "SearchFields", and "Status Search". A banner on the left says "We are building a new home for your molecules." and another on the right lists "PDB Mirrors" with links to various institutions like San Diego Supercomputer Center, Rutgers University, etc.

- Under the "Search the Archive" section of the home page, click the [SearchLite](#) link. **SearchLite** is used to search the text of PDB files, and it provides examples of search strings that can be used to query the PDB database. See a screenshot of the SearchLite page below.



- In the search box, enter **hemochromatosis** and submit your search.
- The search should return ten structures. Read the **Title** and **Compound** fields for each record to determine which structure is the best candidate for examination. The **Compound** field is very helpful for figuring out exactly how many different molecules are included in the structure. Notice that two molecules (Hemochromatosis Protein and  $\beta$ -2-Microglobulin) are included in the first structure (1A6Z) and three (the previous two plus transferrin receptor) in the second structure (1DE4). The first two results from the search are shown below.

<input type="checkbox"/>	<b>1A6Z</b>	Deposited: 04-Mar-1998 Exp. Method: X-ray Diffraction Resolution: 2.60 Å { <a href="#">EXPLORE</a> }
<i>Title</i> Hfe (Human) Hemochromatosis Protein		
<i>Classification</i> Mhc Class I Complex		
<i>Compound</i> Mol_Id: 1; Molecule: Hfe; Chain: A, C; Engineered: Yes Mol_Id: 2; Molecule: $\beta$ -2-Microglobulin; Chain: B, D		
<input type="checkbox"/>	<b>1DE4</b>	Deposited: 12-Nov-1999 Exp. Method: X-ray Diffraction Resolution: 2.80 Å { <a href="#">EXPLORE</a> }
<i>Title</i> Hemochromatosis Protein Hfe Complexed With Transferrin Receptor		
<i>Classification</i> Metal Transport Inhibitor/Receptor		
<i>Compound</i> Mol_Id: 1; Molecule: Hemochromatosis Protein; Chain: A, D, G; Fragment: Ectodomain; Synonym: Hfe; Engineered: Yes Mol_Id: 2; Molecule: $\beta$ -2-Microglobulin; Chain: B, E, H; Engineered: Yes Mol_Id: 3; Molecule: Transferrin Receptor; Chain: C, F, I; Fragment: Ectodomain; Engineered: Yes		

- The best record is 1A6Z. Click on the [EXPLORE](#) link on the right side of the entry to open this PDB record. The record for 1A6Z is shown on the following page.

**Structure Explorer - 1A6Z**

Try the Structure Explorer page for [1A6Z](#) from the new, reengineered RCSB PDB Web site!

**Summary Information**

**Title:** Hfe (Human) Hemochromatosis Protein

**Compound:** Mol\_Id: 1; Molecule: Hfe; Chain: A, C; Engineered: Yes  
Mol\_Id: 2; Molecule:  $\beta$ -2-Microglobulin; Chain: B, D

**Authors:** J. A. Lebron, M. J. Bennett, D. E. Vaughn, A. J. Chirino, P. M. Snow, G. A. Mintier, J. N. Feder, P. J. Bjorkman

**Exp. Method:** X-ray Diffraction

**Classification:** Mhc Class I Complex

**Source:** Homo sapiens

**Primary Citation:** [Lebron, J. A., Bennett, M. J., Vaughn, D. E., Chirino, A. J., Snow, P. M., Mintier, G. A., Feder, J. N., Bjorkman, P. J.](#): Crystal structure of the hemochromatosis protein HFE and characterization of its interaction with transferrin receptor. *Cell* 93 pp. 111 (1998)

**Deposition Date:** 04-Mar-1998      **Release Date:** 23-Mar-1999

### The Importance of the Primary Citation

To understand what you are really looking at in this structure, you would need to access the article listed as the primary citation. Although the full text for some articles may be freely available online, most articles are accessible only through a university library that provides public access to its collections. Sometimes you can purchase a copy of the article online from the publisher.

The article for this structure has been accessed, and here are some of the main points to understand about this structure:

- Only the soluble portion of the HFE polypeptide chain is included in the 1A6Z structure. The transmembrane domain of the HFE protein is missing. This structure has only 275 of the 348 amino acids in the complete HFE protein sequence.
- The first 22 amino acids of the HFE polypeptide sequence have been left out because they are not part of the mature, functional protein. Therefore, the first amino acid in this structure is really the 23<sup>rd</sup>, and cysteine 260 is the cysteine residue involved in the C282Y mutation that we learned about in Activity #1.
- Each HFE polypeptide chain is complexed with another polypeptide chain called beta-2 microglobulin.
- The 1A6Z structure consists of two separate molecules of the HFE/beta-2 microglobulin complex.

5. Make note of the *Compound*: information below the title on this page. It indicates that this structure is a complex of HFE protein and beta-2-microglobulin. Chains A and C are HFE polypeptide chains, and chains B and D are beta-2-microglobulin.

6. Let's look at the sequence data for this structure by clicking on the [Sequence Details](#) link in the purple navigation column on the left. At the Sequence Details page, scroll down to examine the amino acid sequences of the chains in this structure.

**Chain 1A6Z:A**

Compound	Hfe		
Type	Protein		
Molecular Weight	32285		
Number of Residues	275		
Number of Alpha	6	Content of Alpha	21.45
Number of Beta	19	Content of Beta	37.82

**Sequence and secondary structure**

```

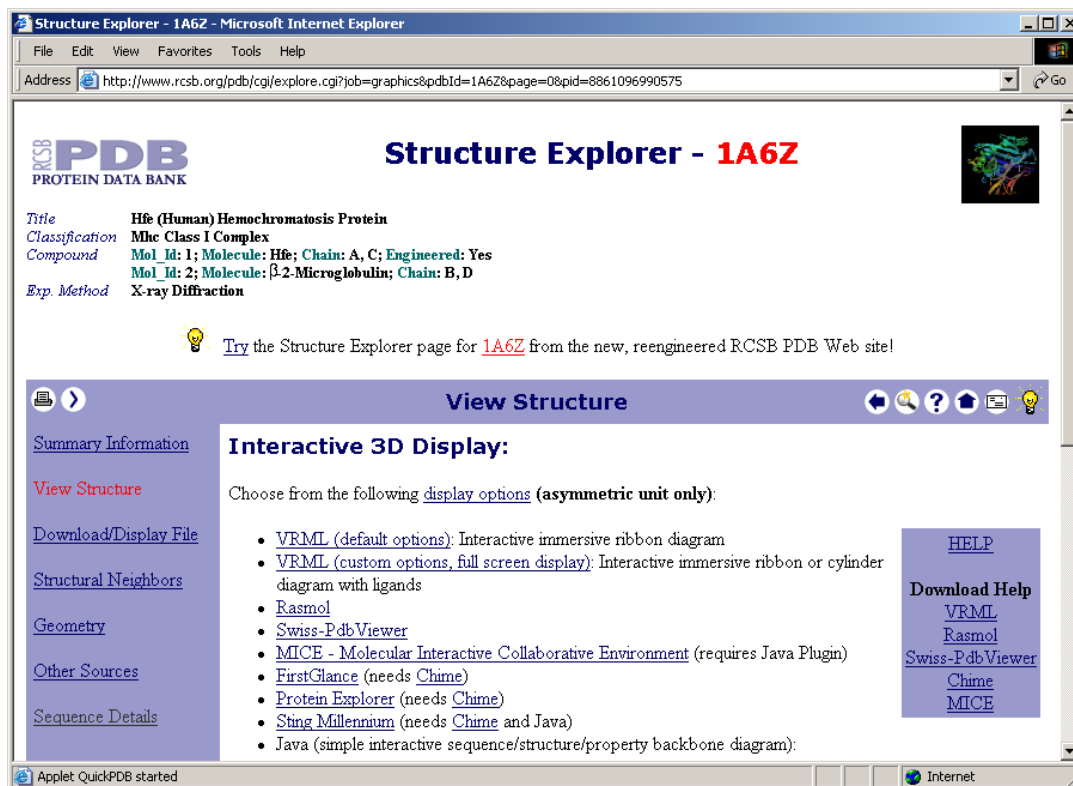
1  RLLRSHSLHY LFMGASEQDL GLSLFEALGY VDDQLFVFD HESRRVEPRT
   EEEE EEEEE SS SSB EEEE ETTEE EEEE SSS EEE S
51  PWSRRISSQ MWLQSQSLK GWDHMTVDV WTIMENHNHS KESHTLQVIL
   SSS SSTTT HHHHTHHHH HHHHHHHHH HHHHTTT S SS EEEEE
101 GCMEQEDNST EGYNKYGYDG QDHLEFCPDT LDWRAAEPR WPTKLEWERH
   EEEE TTS E EEEEEEEET EE EEEETT TEEESSHH HTHHHHTT
151 KIRARONRAY LERDCPAQLQ QLLELGRGVL DQVPPLVKV THHVTSSVTT
   HHHHHHHHH HTHHHHHHH HHHHHHTTS S B EEEE EEEE SS EE
201 LRCRALNYPY QNITMKULKD KQPHDAKEFE PKDVLPNGDG TYQGUITLAV
   EEEEESEES S EEEEEET TEE GGGG EEEE TTS EEEEEEE
251 PPGEEQRYTC QVEHPGLDQP LIVIW
   TTGGGEEE EEE TTSS EEEE
  
```

7. There is a **Sequence and secondary structure** section for each polypeptide chain within the structure. Beneath each line of sequence, there is a line of letters (E's, S's, T's, H's, etc.) that describe the polypeptide's secondary structure [though beyond the scope of this activity the letters represent: H=helix; B=residue in isolated beta bridge; E=extended beta strand; G=310 helix; I=pi helix; T=hydrogen bonded turn; and S=bend].

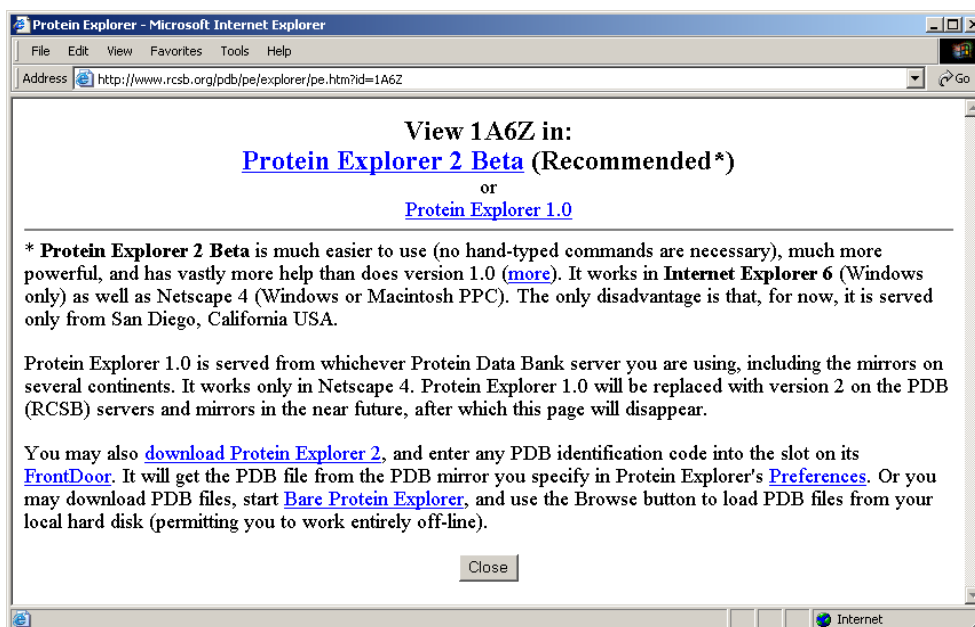
Note that the sequence for the HFE chain is only 275 amino acids long. **Use the amino acid sequence of one of the HFE chains to answer question 1 of the Questions for Activities #4 and #5.**

### Viewing the Structure

8. Select the [View Structure](#) link in the navigation column on the left. This will take you to the part of the record that gives you options for interacting with the 3-D structure of the protein. Still images of the structure are also available for download. See the screenshot on the following page.



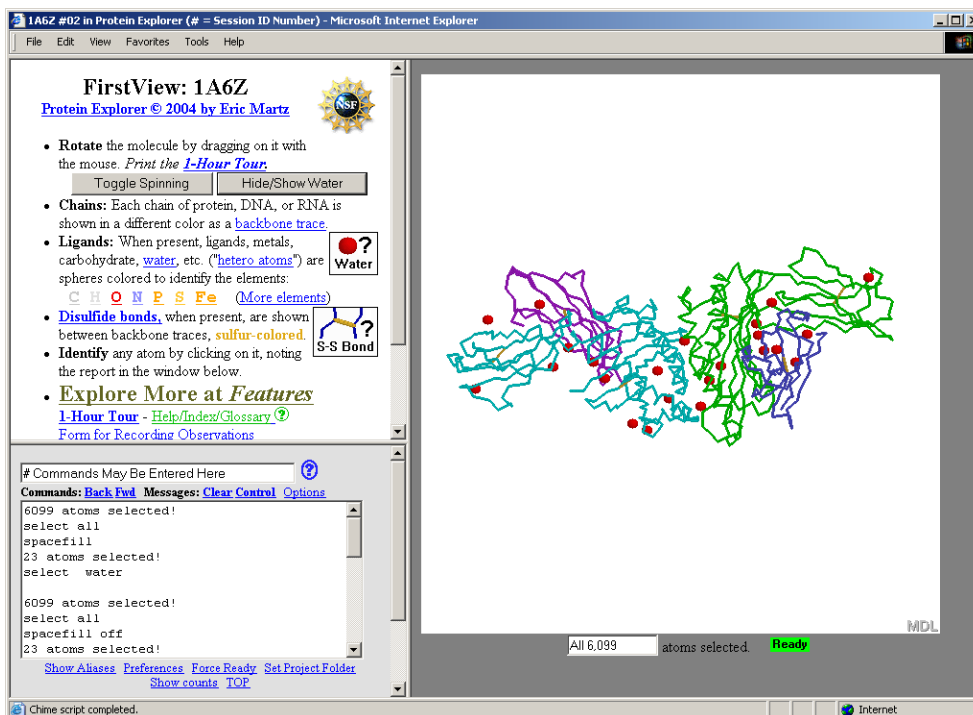
9. From the **Interactive 3D Display** options, select the [Protein Explorer](#) link. If you have downloaded and installed Chime and are using Netscape 4.7X (Mac or Windows) or Internet Explorer 6.0, the window below should open. If you have problems with this step, see PDB's Molecular Graphics HELP (<http://www.rcsb.org/pdb/help-graphics.html>) page to access Troubleshooting FAQs.



10. Select the recommended [Protein Explorer 2 Beta](#) link to launch Protein Explorer.



11. If a message window pops up, just click OK to close the window. The Protein Explorer browser window should look like the screenshot below. Once the structure is loaded, click the **Toggle Spinning** button to stop the molecule's spinning. Click on this button again if you want to restart the spinning. To move the molecule while spinning is turned off, simply click on the structure and drag your mouse to rotate it to a particular position.



**The browser window is divided into three different frames:**

- The first frame (**FirstView: 1A6Z**) in the upper left corner contains general information about the structure, display control buttons, and links to help files and tours.
- The second frame contains a text box for entering commands and a message box that displays information about the selected structure or particular part of the structure.
- The third frame, containing the actual structure, is the largest. Right clicking on the structure (Windows users) or clicking and holding the mouse button still (Mac users) will open a display options menu.

11. The red spheres around the structure are the oxygen atoms of water molecules. Whenever you want to identify a part of the structure, click on it with your mouse and a description will appear in the message box of the second frame. To remove the spheres, click the **Hide/Show Water** button in the first frame.

12. Protein chains are represented by different colors. By clicking on each chain and reading its report in the message box, you will find that the turquoise chain is chain A; the purple chain is chain B; the green chain is chain C; and the blue chain is chain D. When we looked at the sequence details earlier in this activity, we found that chains A and C are HFE polypeptide chains, and chains B and D are beta-2-microglobulin.

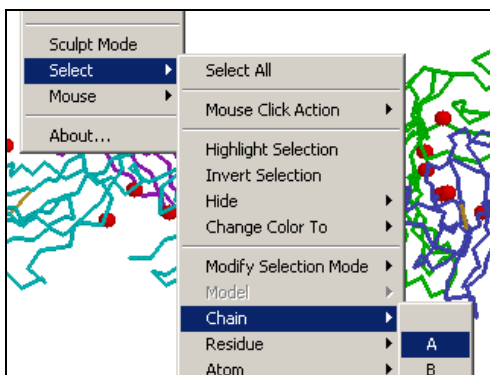
**The rest of this activity will focus on how you can customize display options:**

- By accessing options menus using mouse controls (right-clicking for Windows users or holding down the mouse for Mac users)
- By entering commands into the text box of frame 2 where it says “# Commands May Be Entered Here”

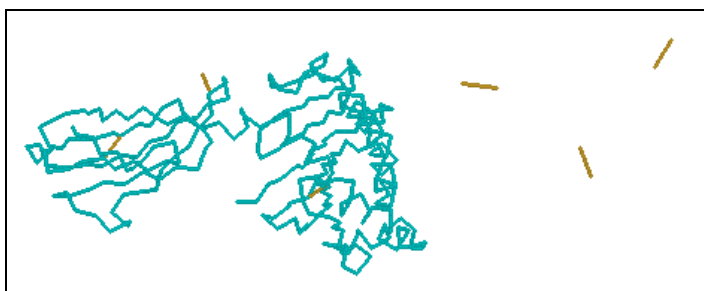
In this activity we will mainly use the options menus to modify the structure. To learn more about using commands in Protein Explorer see *Select Commands in Chime and RasMol* (<http://www.umass.edu/microbio/rasmol/seleccmd.htm>) by Eric Martz.

**NOTE:** If you ever need to start over, click on the [PE Site Map](#) link near the bottom of the **FirstView: 1A6Z** frame and then select [Reset Session](#).

13. Any time you want to alter the color or display option for any portion of a structure, you must first select it. Select chain A by accessing the options menu with your mouse and choosing **Select > Chain > A**.



14. Let's alter the structure so that only HFE chain A (the turquoise chain) is displayed. To do this we will need to hide chains B, C, and D. Since we have already selected chain A, we can hide all nonselected portions of the structure by choosing **Select > Hide > Hide Not Selected**. The structure should look something like the screenshot below.



The tan or brownish lines that remain after hiding chains B, C, and D represent disulfide bridges, which are special bonds that connect different parts of protein chains together. Protein Explorer displays disulfide bridges by default. If you look closely you should also be able to see two disulfide bridges in chain A. Disulfide bridges are important for stabilizing the three-dimensional structure of a protein, therefore making the protein less susceptible to degradation.

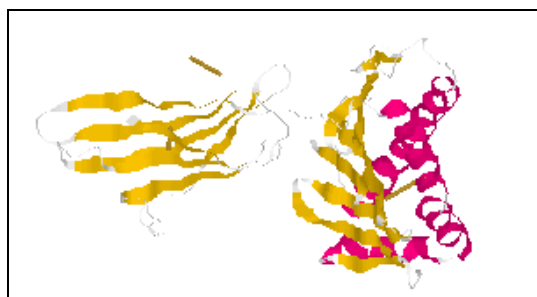
**A Note about Displaying Disulfide Bridges:**

There is an option for choosing not to display disulfide bridges, but a glitch with this option keeps it from working properly all the time. For this activity, let's leave the disulfide bridges displayed. But if you want to experiment with this option after completing this activity, to hide the disulfide bridges in all chains but chain A, you would:

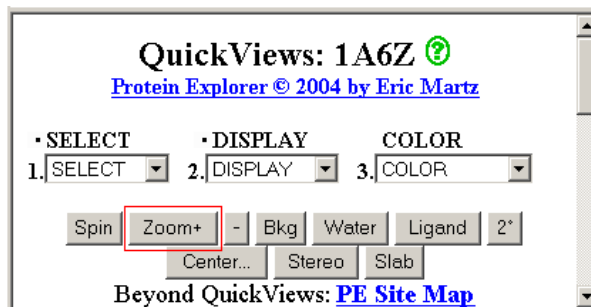
- Make sure chain A is selected by choosing **Select > Chain > A**.
- Select everything but A by choosing **Select > Invert Selection**.
- To hide all bridges in chains B, C, and D, choose **Options > Display Disulfide Bridges**.
- A new browser window will appear. Close this new window, and you should see that the disulfide bridges in chains B, C, and D have been removed.

15. Chain A should still be selected. To change the display of chain A from backbone (the default display type) to ribbons, choose **Display > Ribbons**. The ribbons display option shows where the alpha helices and beta-sheets are in the secondary structure of the protein.

16. In addition to selecting the ribbons display option; we can see more clearly where the alpha helices, turns, and beta sheets are by choosing **Color > Structure** from the menu options with the mouse. With this color option, the beta strands that make up the beta-sheets are gold, alpha helices are magenta, and turns are white. The structure should now resemble the image below.



17. To zoom in, select the [Explore More at Features](#) link in the **FirstView: 1A6Z** frame (the upper left frame). Then select [Explore More with Quick Views](#). Click the **Zoom +** button (see screen shot below) a couple of times.



18. To center chain A in the middle of the viewing frame click the **Center** button. When the "Click OK to center one atom by clicking on it..." window pops up, click OK, and then click on

the chain somewhere near its middle. The structure will center itself around the area that you clicked.

19. In Activity #4, we accessed the **Features** table of the Swiss-Prot sequence record for the human HFE protein. This table showed which amino acids were associated with the functional domains within the protein. This table indicated that cysteine 282 is part of the extracellular alpha-3 domain, which is comprised of amino acids 206–297. Let's highlight this domain in Protein Explorer by changing the color of these amino acids.

**REMEMBER:** The amino acids in the HFE chain from PDB are numbered differently from the amino acids in the Swiss-Prot record. The first 22 amino acids are missing from the HFE chain in the PDB structure; therefore, the amino acids in which we are interested are 184–275.

To select amino acids 184–275 type **select 184-275:a** into the command box of frame 2 and hit Enter or Return.



20. To change the color of these amino acids, choose **Select > Change Color To > Blue** from the menu options.

21. Now, highlight the cysteine residue that is altered in the C282Y mutation that causes hereditary hemochromatosis. This mutation changes a highly conserved amino acid at position 282 from cysteine (C) to tyrosine (Y).

**NOTE:** Since the HFE chain in the structure is 22 amino acids shorter than the complete sequence in Swiss-Prot, the cysteine altered in the C282Y mutation is cysteine 260 ( $282-22=260$ ).

Let's select this cysteine by typing **select 260:a** into the command box of frame 2 and hitting Enter or Return. This command selects residue 260 of chain A.

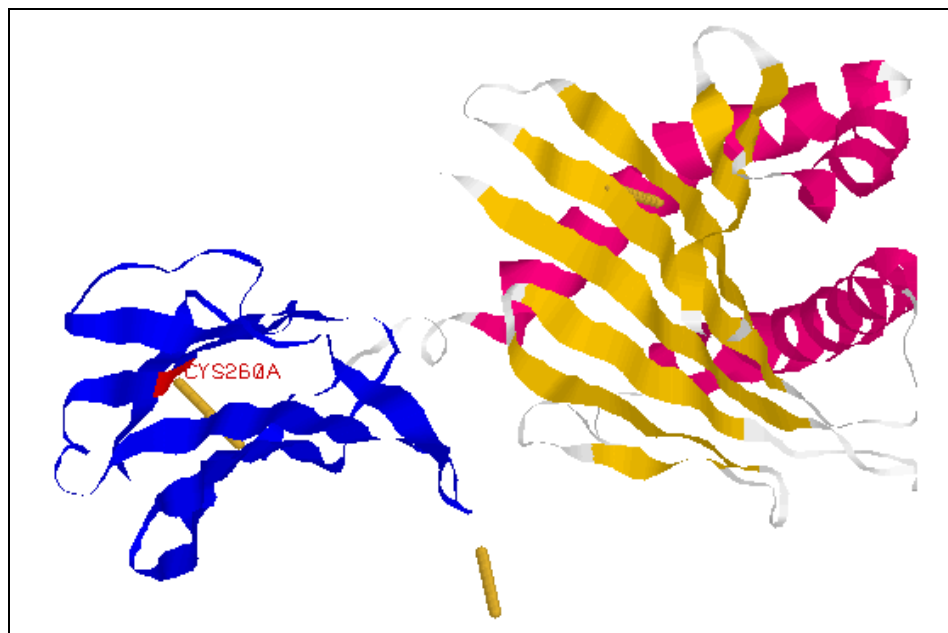


22. To label this residue, using the options menu, choose **Options > Labels**.

23. To change the color of this residue to red choose **Select > Change Color To > Red**. You may want to click on the structure and drag your mouse to move and rotate the structure into another position.

24. The end product of this activity should look something like the structure on the following page.

### Chain A of PDB Structure 1A6Z (Hfe (Human) Hemochromatosis Protein)



### Importance of Protein Structure and Its Role in Disease Development

By examining the structure of HFE, we can see that disulfide bonds are essential for establishing the proper three-dimensional formation of a protein. In the C282Y mutation, a cysteine residue is replaced by another amino acid, tyrosine, and the disulfide bond between cysteines 225 and 282 is lost. This is detrimental to the protein's structure. As a result, the HFE protein can no longer perform its normal function regulating iron uptake, and cells get overloaded with iron. This buildup of iron in cells leads to organ damage and other hemochromatosis symptoms and complications.



## Table of Standard Genetic Code

	<b>T</b>	<b>C</b>	<b>A</b>	<b>G</b>
<b>T</b>	TTT Phe (F) TTC Phe (F) TTA Leu (L) TTG Leu (L)	TCT Ser (S) TCC Ser (S) TCA Ser (S) TCG Ser (S)	TAT Tyr (Y) TAC <b>TAA STOP</b> <b>TAG STOP</b>	TGT Cys (C) TGC <b>TGA STOP</b> TGG Trp (W)
<b>C</b>	CTT Leu (L) CTC Leu (L) CTA Leu (L) CTG Leu (L)	CCT Pro (P) CCC Pro (P) CCA Pro (P) CCG Pro (P)	CAT His (H) CAC His (H) CAA Gln (Q) CAG Gln (Q)	CGT Arg (R) CGC Arg (R) CGA Arg (R) CGG Arg (R)
<b>A</b>	ATT Ile (I) ATC Ile (I) ATA Ile (I) <b>ATG Met (M) START</b>	ACT Thr (T) ACC Thr (T) ACA Thr (T) ACG Thr (T)	AAT Asn (N) AAC Asn (N) AAA Lys (K) AAG Lys (K)	AGT Ser (S) AGC Ser (S) AGA Arg (R) AGG Arg (R)
<b>G</b>	GTT Val (V) GTC Val (V) GTA Val (V) GTG Val (V)	GCT Ala (A) GCC Ala (A) GCA Ala (A) GCG Ala (A)	GAT Asp (D) GAC Asp (D) GAA Glu (E) GAG Glu (E)	GGT Gly (G) GGC Gly (G) GGA Gly (G) GGG Gly (G)

### Key to the Table of Standard Genetic Code

Alanine	ALA	A	Arginine	ARG	R
Asparagine	ASN	N	Aspartic acid	ASP	D
Cysteine	CYS	C	Glutamic acid	GLU	E
Glutamine	GLN	Q	Glycine	GLY	G
Histidine	HIS	H	Isoleucine	ILE	I
Leucine	LEU	L	Lysine	LYS	K
Methionine	MET	M	Phenylalanine	PHE	F
Proline	PRO	P	Serine	SER	S
Threonine	THR	T	Tryptophan	TRP	W
Tyrosine	TYR	Y	Valine	VAL	V

STOP = Termination Signal - signifies the end of a polypeptide chain





## Hereditary Hemochromatosis Worksheet

This worksheet provides questions to be answered as you complete the activities in the Gene Gateway Workbook.

### Questions for Activity #1

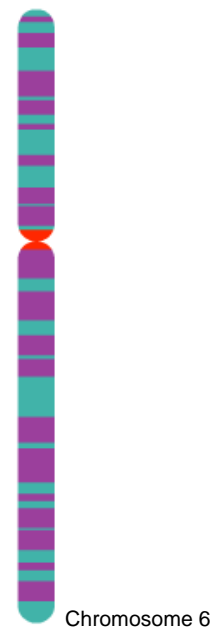
- 1) What are some of the characteristics of hereditary hemochromatosis? How is it treated?
- 2) What is the official gene symbol of the hereditary hemochromatosis gene, and what is its gene locus?
- 3) Briefly describe the CYS282TYR mutation included in the Allelic Variants section of OMIM and the Molecular Genetics section of the hereditary hemochromatosis GeneReview. Most individuals with hereditary hemochromatosis have two copies of this disease-causing mutation.

### Questions for Activity #2

- 1) On the diagram to the right, mark the region where the HFE gene can be found on chromosome 6.

At the end of Activity 2, we zoomed in on the 6p21.3 region of chromosome 6. Use the information in this chromosome view to answer the following questions:

- 2) What is the estimated total number of genes on chromosome 6?
- 3) How many genes are in the 6p21.3 region of the chromosome?
- 4) How many phenotypes are in the 6p21.3 region?



### Questions for Activity #3

1) Using the summary provided in Entrez Gene for HFE, briefly describe the function of the gene's protein product

#### Use the GenBank sequence record Z92910 to answer questions 2–6.

2) In the Features section of record Z92910, select the [gene](#) link. How many base pairs (bp) are in the genomic sequence of the HFE gene?

3) Return to the main record Z92910, and scroll through the Features section. How many exons have been identified in this sequence?

4) Select the [exon](#) link for exon 1. How many base pairs are in exon 1?

5) Return to the main record Z92910. Select the [CDS](#) link. How many base pairs are in the coding sequence?

6) Look at where exon 1 begins and where the CDS begins. How many nucleotides in exon 1 do not code for protein?

#### Use the information in the box about the first HFE variant ([NM\\_000410](#)) in GenAtlas to answer the following questions.

7) How large is the mRNA transcript?

8) Compare the size of the mRNA transcript with the size of the coding sequence you found for Question 5. How much of the mRNA is not coding sequence?

## Questions for Activities #4 and #5

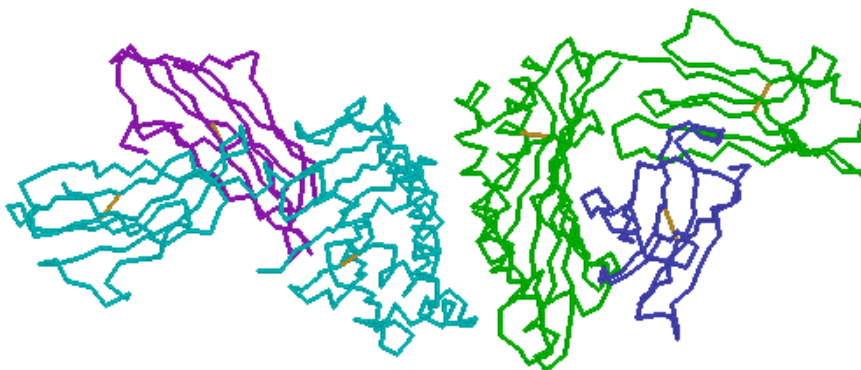
1) Examine the sequence from Swiss-Prot and the sequence of one of the HFE chains provided in the Sequence Details section of the PDB structure 1A6Z. In the Swiss-Prot sequence shown below, underline the portion of the amino acid sequence that is included in the PDB structure.

```

      10           20           30           40           50           60
      |           |           |           |           |           |
MGPRARPALL  LLMLLQTAVL  QGRLLRSHSL  HYLFMGASEQ  DLGLSLFEAL  GYVDDQLFVF
      |           |           |           |           |           |
      70           80           90           100          110          120
      |           |           |           |           |           |
YDHESRRVEP  RTPWVSSRIS  SQMWLQLSQS  LKGWDHMFTV  DFWTIMENHN  HSKESHTLQV
      |           |           |           |           |           |
     130          140          150          160          170          180
      |           |           |           |           |           |
ILGCEMQEDN  STEGYWKYGY  DGQDHLEFCP  DTLDWRAAEP  RAWPTKLEWE  RHKIRARQNR
      |           |           |           |           |           |
     190          200          210          220          230          240
      |           |           |           |           |           |
AYLERDCPAQ  LQQLLELGRG  VLDQQVPLV  KVTHHVTSSV  TTLRCRALNY  YPQNITMKWL
      |           |           |           |           |           |
     250          260          270          280          290          300
      |           |           |           |           |           |
KDKQPMDAKE  FEPKDVLPNG  DGTYQGWITL  AVPPGEEQRY  TCQVEHPGLD  QPLIVIWEPS
      |           |           |           |           |           |
     310          320          330          340
      |           |           |           |
PSGTLVIGVI  SGIAVFVIL  FIGILFIILR  KRQGSRGAMG  HYVLAERE

```

2) PDB structure 1A6Z is shown below. Label the chains as being either HFE or beta-2-microglobulin (b2m).



3) Why is the cysteine residue affected in the C282Y mutation important?

## Contact Information

This document was produced by the Genome Management Information System at Oak Ridge National Laboratory, Oak Ridge, Tennessee, July 2003. The content was last updated October 5, 2004.

If you have any questions or comments concerning this document, contact Jennifer Bownas, [bownasjl@ornl.gov](mailto:bownasjl@ornl.gov), 865/574-7582.

## For more information

Gene Gateway: <http://www.DOEgenomes.org/genegateway>

Human Genome Project Information: <http://www.ornl.gov/hgmis/home.html>

DOE Genome Research Programs: <http://www.DOEgenomes.org/>

U.S. Department of Energy (DOE)  
Office of Science  
Office of Biological and Environmental Research  
Genome Research Programs

