CHAPTER II

MATERIALS AND METHODS

2.1 Materials

2.1.1 Chemical Catalogue

Biochemicals, Organic Compounds for Research and Diagnostic Reagents (Sigma Chemical Company, USA)

ICN A World of Biomedical Products (ICN Pharmaceuticals, Inc., USA)

2.1.2 Protein Database

SWISS-PROT, URL = http://expasy.hcuge.ch/sprot/sprot-top.html

Protein Data Bank (PDB), URL = http://pdb.pdb.bnl.gov/

2.1.3 Computer Hardware

Power Macintosh 7100/66 AV (Apple Computer, Inc., USA), which includes the following hardware and software:

System Version: 7.5.3 Thai

System Size: 2,581 KB

Rom Version: 1917

AppleTalk Version: 58

32-Bit QuickDraw Version: 1.3

Hard Disk: 1,200,000 KB

Total Memory: 32,440 KB

2.1.4 Computer Software

2.1.4.1 Database Management Software

FileMaker Pro 3.0 (Claris Corporation, USA) was used for managing the CU Peptide Database. The DBMS is capable of searching the database rapidly and requires relatively low memory because of its *disk-caching* scheme, loading only the data that it requires at the moment. Furthermore, FileMaker Pro 3.0 has a CGI application support for Internet accessibility.

2.1.4.2 HTTP Server

WEBSTAR (Starnine corporation, USA) was used as a HTTP server in this research. It is one of the most popular server management programs on the Apple Macintosh.

2.1.4.3 CGI Application

A CGI application is a special application for communicating between the database management program and the HTTP server. *WEB FM* (Web Broadcasting Corporation, USA), the CGI application used in this research, includes a very fast interface with FileMaker Pro.

2.1.4.4 Internet Browser

Netscape Navigator (Netscape Communications Corporation) is one of the most popular browser on the market. It was used in this research for browsing the CU Peptide Database pages as well as for interpreting calculation routines written in JavaScript.

2.1.4.5 Molecular visualization

Rasmol (the UC Regents/Modular CHEM Consortium) was used to visualize molecular information obtained primarily from PDB (Protein Data

Bank, Brookhaven National Laboratory).

Table 2-1 Specifications of computer software used in this research

Software Name	Version	Size (on disk)	Memory Requirements
FileMaker Pro	3.0 v.1	2.2 MB	5,002 KB
Webstar WWW Server	v.1.2.2T	429 KB	1,682 KB
WEB FM	2.0 v.3	527 KB	1,024 KB
Netscape Navigator	3.0	4.3 MB	9,000 KB
Rasmol	2.6 ucb	280 KB	1,024 KB

2.2 Methods

2.2.1 System Analysis

System analysis was performed in the course of preliminary planning the contents and the look and feel of the database. First, biochemical requirements were considered and then, consistently with the biochemical requirements, biochemical data fields in the database were created for peptide data. Second, computer scripts were designed for calculating some properties of peptides such as molecular weight, net charge, isoelectric point and hydropathic index. Finally, in order to make the CU Peptide Database accessible to the global biochemical community via the Internet, a home page of the CU Peptide Database was designed.

2.2.2 CU Peptide Database Construction

The overall structure of the CU Peptide Database was determined. To simplify the calculation script in calculating some properties of peptides such as molecular weight and net charge, CU Peptide Database was designed to be composed of two relational database, Database1 and Database2. Database1 is composed of 971 peptides with their data. Database2 was created to store each residue's data. They linked together by a common "Peptide Name" field. Database1 retrieved each datum of residue of database2 and calculate the overall value. The overall structure of the system is shown in figure 2.1.

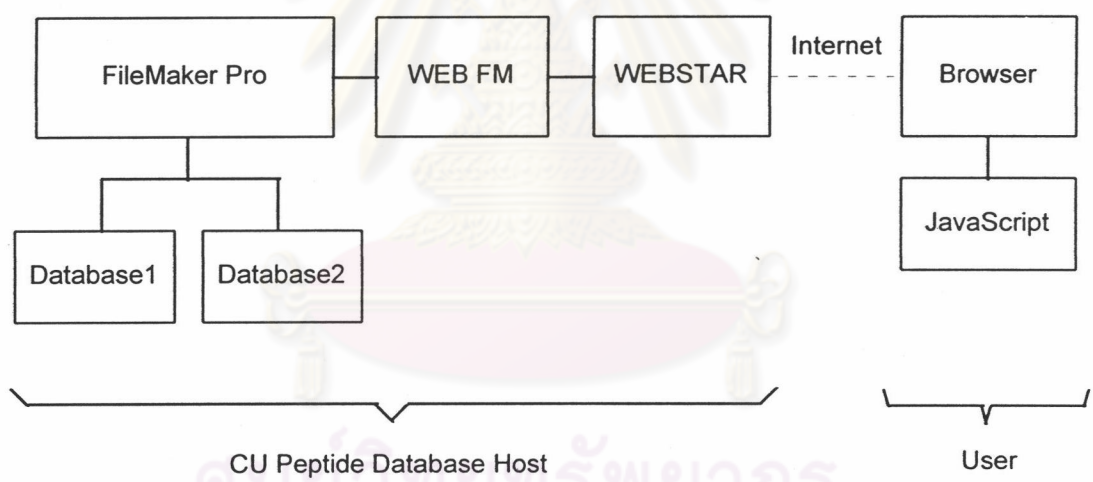

Figure 2-1 Overall structure of the CU Peptide Database

In FileMaker Pro, FileMaker scripts were used to calculate molecular weight, net charge, isoelectric point and hydropathic index. In CU Peptide Database web page, Scripts were created in JavaScript language and inserted into HTML for calculating molecular weight, net charge and isoelctric point.

2.2.2.1 Peptide Data Collection

Peptide data were collected from several chemical catalogues. The data include peptide name, category of the peptide, peptide sequence, source, disulfide bond and product code number.

2.2.2.2 Creation of CU Peptide Database

FileMaker Pro 3.0 was used to manage the CU Peptide Database.

A. Fields for an individual kind of peptide data in section 2.2.2.1, i.e. name, etc. were created and then were filled with data. Fields for molecular weight, net charge at any pH, isoelectric point, and hydropathic index were also created in order to contain calculated data described in the following section. The created fields are shown in Figure 2-2.

B. Scripts were written for calculating the molecular weight (MW), the net charge at any pH, the isoelectric point (pI) and the hydropathic index of each peptide. Biochemical foundation in the script are as follow: (See the script in appendix A.)

1. Molecular Weight

The MWs for all residues were calculated separately in the Database2, then the Database1 retrieved each residue's MW and summed to give total MW. The MWs of amino acids are shown in Table A-1 in Appendix A.

Peptide Name	ADRENOCORTICOTROPIC HORMONE (ACTH 1-39 ; Corticotropin A)
Sequence	Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Trp-Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Asn-Gly-Ala-Glu-Asp-Glu-Ser-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe
Category	Adrenocorticotropic Hormone and Fragments
Source	Human
Disufide Bridge	-
Product Code Number	A 0423
Molecular weight	
Net Charge	**At pH**
Isoelectric Point	
Hydrophatic Index	

Figure 2-2 Fields for storing peptide data created in FileMaker Pro 3.0

Table 2-2 Field attribute

No.	Field Name	Kind	Length (character)	Locked	Text Wrap
1.	Peptide Name	Text	64,000	✓	✓
2.	Sequence	Text	64,000	✓	✓
3.	Category	Text	64,000	✓	✓
4.	Source	Text	64,000	✓	✓
5.	Disulfide Bridge	Text	64,000	✓	✓
6.	Product Code Number	Text	64,000	✓	✓
7.	Molecular Weight	Calculation	255	✓	✓
8.	Net Charge	Calculation	255	✓	✓
9.	At pH	Value List	255	✓	✓
10.	Isoelectric Point	Calculation	255	✓	✓
11.	Hydropathic Index	Calculation	255	✓	✓

2. Net Charge

The net charge for all residues were calculated separately in the Database2, then the Database1 retrieved each residue's net charge and summed to give total net charge. The charge of carboxyl group and amino group of each residue were ignore since they all formed peptide bonds to their neighboring residues. Thus, only side chain group was considered. The net charge residue with unionizable group was set to zero. The net charge at a given pH of seven residue with ionizable group, which are Arginine, Lysine, Histidine, Aspartic acid, Glutamic acid, Cysteine and Tyrosine, were calculated as follow:

From the Henderson - Hasselbalch equation

$$pH = pK + \log \frac{[A^-]}{[HA]}$$

which pK = -log dissociation constant, [HA] is the concentration of acid and $[A^-]$ is the concentration of conjugate base

$$\text{So, } -\log \frac{[A^-]}{[HA]} = pK\text{-}pH$$

$$\log \frac{[HA]}{[A^-]} = pK\text{-}pH$$

$$\frac{[HA]}{[A^-]} = \frac{10^{(pK-pH)}}{1}$$

$[HA]$ and $[A^-]$ of an R group of a peptide are mixed in the solution, so they can be presented in ratio as follow:

$$[HA] = \frac{10^{(pK-pH)}}{(10^{(pK-pH)}+1)}$$

$$[A^-] = \frac{1}{(10^{(pK-pH)}+1)}$$

The net charge of any residue is the sum of charges in acid form and charges in conjugate base form. So:

net charge = ratio of [HA] x charge in acid form + ratio of $[A^-]$ x charge in base form

$$\text{net charge} = \frac{10^{(pK-pH)}}{(10^{(pK-pH)}+1)} \text{ x charge in acid form} + \frac{1}{(10^{(pK-pH)}+1)} \text{ x charge in base form}$$

Then, the net charge of all residues, including net charge of amino group at N-terminal and carboxylic group at C-terminal, were summed to give the total net charge. (See the charges in acid form and base form of ionizable groups of standard amino acids in Table A-1 in Appendix A.)

3. Isoelectric Point (pI)

The isoelectric point of a peptide is the surrounding pH at zero net charge on the peptide molecule. When the "Calculate pI" button was depressed, the pI script started calculating the net charge of the peptide at a very low pH value. If the net charge was still above zero, the script would run again at 0.0001 pH unit higher than last round, until the net charge approached zero (within -0.0001 to 0.0001 unit). The script would return the pH value, i.e. the calculated isoelectric point.

4. Hydropathic Index

Hydropathic index is a good predictor of which portions of a peptide are inside a protein and which portions are outside (Voet, 1990). In this database, The script assigned an appropriate hydropathy value to each amino acid residue of peptide and then summed the hydropathies of nine consecutive residue, starting at the amino terminal, within overlapping segments displaced from each other by one residue. A given sum was then plotted above the middle residue of the segment (Kyte and Doolittle, 1982). Thus, the first value corresponding to the sum of the hydropathies of residues 1 to 9 was plotted at location 5, the second value corresponding to the sum of the hydropathies of residues 2 to 10 was plotted at location 6, and so on. The hydropathies of 20 standard amino acids are shown in Table A-1 in Appendix A.

2.2.3 CU Peptide Database Web Page Construction

As the CU Peptide Database was designed to function on the WWW, pages of the database were created in HTML and JavaScript. (See the scripts in Appendix B.) The pages were composed of CU Peptide Database home page, Search Form page, Results page and Details page. An attractive logo and a background graphic were added to make CU Peptide Database interesting apart from giving biochemical data.

2.2.3.1 CU Peptide Database Introductory Page

The home page includes the definition and summary of the CU Peptide Database as well as links to the Search Form and to other biochemical databases.

2.2.3.2 Search Form Page

Form, which consist of Category, Peptide Name, Sequence, Source, Total residue, MW and pI field, were created for users in searching the desired peptide using several search criteria.

2.2.3.3 Results Page

Clicking *Search Now! button* in the Search Form page leads to the Results page. This web page returns lists of peptides that match the user's search criteria.

2.2.3.4 Details Page

Information of the peptide that the user has selected in the Result page were shown in table. The information includes Category, Sequence, Total Residue, Source, Disulfide bridge, Hydropathic Index, Product Code Number and Cross References to other biochemical database. The 213 major peptides were

searched for cross references in SWISS-PROT database and Protein Data Bank (PDB). the resulting URLs were stored in the "Cross Reference" field.

Calculation Area was also added if all residues in such peptide are standard amino acids. The Calculation Area allows users to calculate molecular weight, pI and net charge of the peptide at any pH. Furthermore, it lets users construct their own peptide and calculate the above properties of the peptide.

2.2.4 Primary Test of CU Peptide Database

Primary test of CU Peptide Database was performed for examining the CU Peptide Database. Accessing and searching the database on a remote computer were tested to ensure that CU Peptide Database can be accessed via the Internet. Several search criteria were performed including testing JavaScript responsible for auto-filling text in the form. Besides, interactive calculations of peptide properties in calculation area were tested.

2.2.5 Comparison of the Result from the CU Peptide Database with Data Derived from Experimental Studies

In order to demonstrate the potential use of the CU Peptide Database, the hydropathic index calculated from this database was compared with the sidedness (inside or outside the molecule) of the amino acid residues as observed on Rasmol, a molecular visualization program. The peptides used in this part of the study were peptides on the CU Peptide Database that could be cross-referenced to PDB in order to obtain three dimensional structure data.