

LBL-37809  
UC-408

**An Electronic Laboratory Notebook  
Based on HTML Forms**

Johanna E. Marsteller and Manfred D. Zorn

Information and Computing Sciences Division  
Lawrence Berkeley Laboratory  
University of California  
Berkeley, California 94720

October 1995

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

**DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED**

**DISCLAIMER**

**Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.**

---

# An Electronic Laboratory Notebook based on HTML Forms

---

Johanna E. Marstaller and Manfred D. Zorn

Software Documentation and Developers Manual\*  
LBLN Technical Report LBL-37809

---

## 1.0 Introduction

---

The Resource for Molecular Cytogenetics at Lawrence Berkeley National Laboratory and at the University of California, San Francisco, funded by the Department of Energy, has been created to facilitate the application of molecular cytogenetics in clinical and biological studies.

One of the goals of the Resource is the development of an evenly spaced set of P1 clones along each chromosome. This probe mapping effort starts with the selection of an appropriate target locus and the definition of a set of primers to select a corresponding clone out of a human P1 library. The P1 clone is subsequently mapped onto metaphase chromosomes to determine the position along the axis in fractional length from the p terminus of the chromosome.

To facilitate the free data exchange between researchers at UCSF and LBL we developed a laboratory notebook to support this probe mapping effort. Several constraints had to be taken into consideration: physically distributed users at LBL and UCSF, different hardware platforms (Mac-

---

\* This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, UC-408, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

intosh, Unix workstations), integration of available public information like GDB, ability to do rapid prototypes and frequent modifications of customized user interfaces. Recognizing the potential of the emerging tools based on the World Wide Web, we decided to build the laboratory notebook with a web browser as the user interface that connects to a simple Unix database system. The Mosaic/Netscape client allows to formulate retrieval and edit operations that are sent to the database. Results are filtered through Perl scripts which generate HTML documents with Hypertext links that are sent back to the Mosaic/Netscape client.

## **2.0 World Wide Web Server - Laboratory Notebook**

---

The World wide Web browsers with HTML based forms provide a fast and easy mechanism to create forms-based user interfaces. The computer scientist can sit down with the biologist and rapidly make changes in response to the user's comments. Furthermore the HTML forms work equally well on a number of different hardware platforms; thus the biologists may continue using their Macintosh computers and find a familiar interface if they have to work on a Unix workstation. The web browser can be run from any machine connected to the Internet: thus the users are free to enter or view information even away from their labs at home or while on travel. Access can be restricted by password and other means to secure the confidentiality of the data. A bonus that is hard to implement otherwise is the facile connection to outside resources. Linking local information to data in public databases is only a hypertext link away with little or no additional programming efforts.

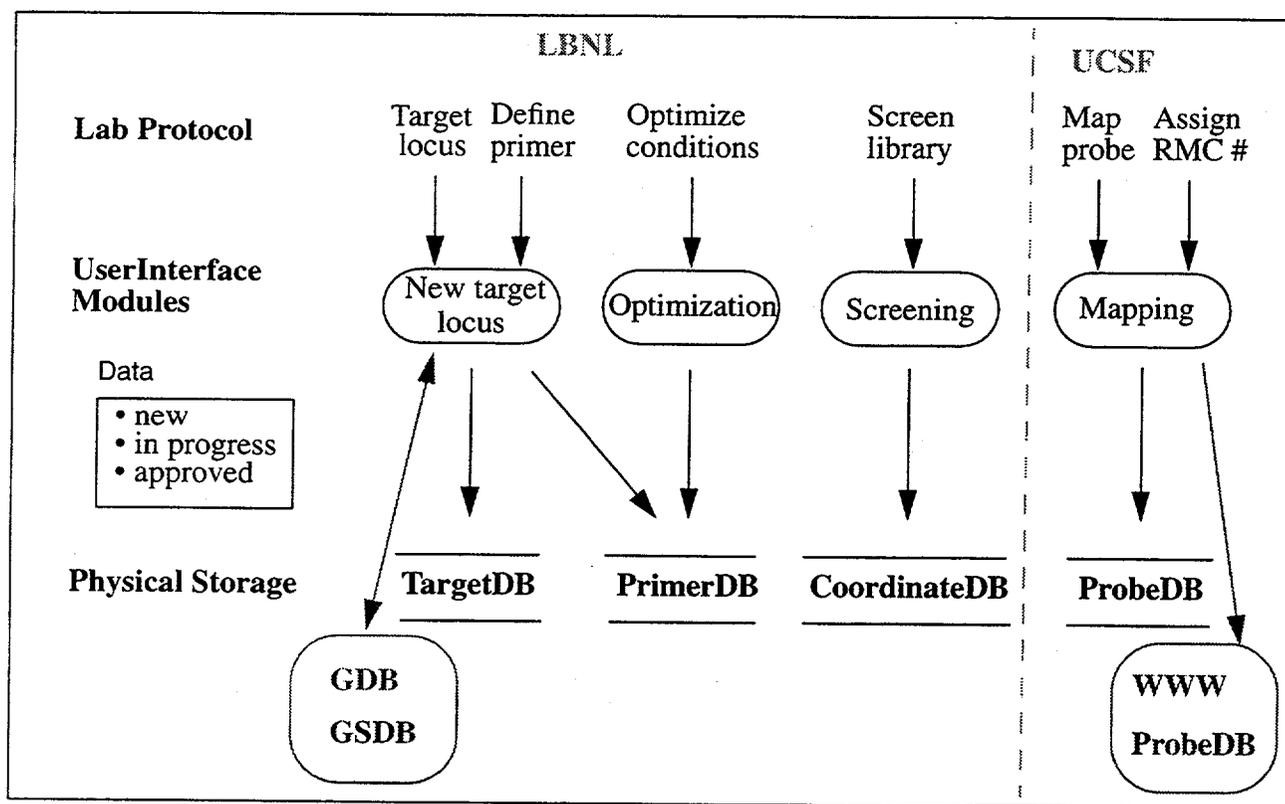
The electronic notebook records information that has traditionally been kept in handwritten laboratory notebooks. It keeps detailed information about the progress of the research, such as the optimization of primers, the screening of the primers and, finally, the mapping of the probes. The notebook provides two areas of services: Data entry, and reviewing of data in all stages.

### **2.1 Data Entry**

The data entry part of the laboratory notebook is organized in a number of sections that reflect the individual stages in the mapping process. Only scientists involved in the research of each stage have access to that stage via password.

Figure 1 describes the data flow for the probe mapping process. The dashed line shows the division of labor between LBNL and UCSF. The Lab Protocol identifies the individual steps involved in the mapping process. The mapping stages correspond to the different tasks as they are implemented in the electronic notebook. The determining of the target locus and the finding of a suitable primer is combined in the Target/Locus section of the data entry part. As data is accumulated in the different stages of the mapping process, it is assigned to different physical storage locations, i.e., Unix dbm files. When a new target locus is identified, information is retrieved from either GDB or GSDB automatically. Once the probes are made public, they are stored in a separate dbm file on a separate server to ensure protection of the data.

FIGURE 1. Data Flow



Each section builds on the preceding section and adds information specific to its particular stage. Entries are time-stamped and moved along the protocol steps automatically within a section. Releasing entries to the

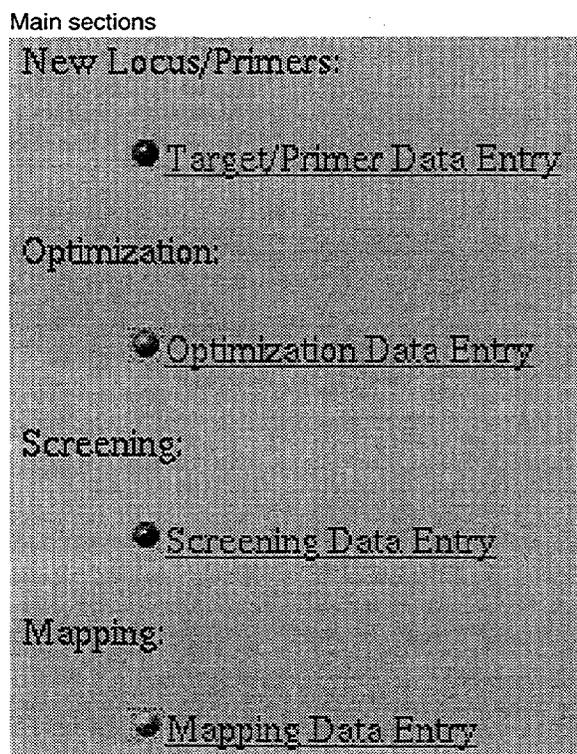
next stage requires explicit approval from the responsible scientist. Figure 3 exemplifies the protocol steps of all laboratory notebook data. Data is visible in various browsers depending on its status. Once data is approved in one stage, it migrates to the *New* browser of the next stage.

Each row in a browser refers to one locus target and the collected information from each stage of the mapping process. A search engine based on the criteria for each stage retrieves the data. Documents are selected for updating by clicking on the appropriate link. Once the documents are submitted the script returns the user to the respective browser.

The main sections of the data entry part of the laboratory notebook are shown in Figure 2.

---

FIGURE 2.



### 2.1.1 Target Primer Data Entry

The target locus is identified at the beginning of the probe mapping process. Dependent on the name of the target locus, the corresponding entry in either GDB or GSDB public genome database is retrieved, and a pointer into the public database is stored in the local notebook. In addi-

tion to local administrative information, a target may have a set of primer pairs defined. All scripts are stored in the directory

`/home/zorn/local/httpd/cgi-bin/primer.`

The README file in this directory gives an overview of the main scripts. (See Appendix A for a listing of the README files). The data entry choices of the first stage are:

---

FIGURE 3.

Data Entry Target Locus

DATA ENTRY TARGET/PRIMERS

[Home] [RmcGroup] [DataEntry]

---

New Locus/Primers:

- Create new Target
- In Progress
- Approved
- Discontinued

- Create new Target

The user indicates the target locus and its origin, i.e., GDB, GSDB or an internal target locus, using this form. (See Appendix B, Figure 17). Clicking *Information* retrieves data from the public databases (via Perl script *infoquery*). The retrieved data is then displayed in a new form where additional fields can be filled before submitting the data to the Target database file (Appendix B, Figure 19). The Perl script automatically stores a link to the GDB document in the new form. Once all data is entered in the form the user can submit the data to the Target database file. Clicking the *Submit* button invokes the Perl script *submit*. Essentially, this checks the status of the target data and stores the target related

data accordingly in the Target database file and the primerpair related data in the primer dbms file.

- In Progress
- Approved
- Discontinued

The three selections, *In Progress*, *Approved*, and *Discontinued* for Target data entry each retrieve a table from the database where a line in the table or browser describes one entry of the target entity in the three respective categories. Each line of the browser contains the creation or approval date, the corresponding primerpair and the name of the target locus, which is a link to the detailed document, and displays it in the form described in Appendix B, Figure 18. Documents retrieved from *In Progress* may be updated. However, once a target locus has been approved for optimization, the data as stored in the Target database file may not be altered. For that purpose, documents retrieved from the approved table are displayed in text format only (Appendix B, Figure 18).

Until a target locus with a set of primers has been approved, it is not visible, i.e., cannot be selected for updating in the optimization stage. Discontinued locus targets may be reselected for further research.

### 2.1.2 Optimization Data Entry

Newly released primer pairs are time-stamped in the notebook and appear on one of the primer pages. Each primer pair features a pair of sequences, names, and various stages of primer optimization to determine the optimal working conditions. Experimental results of the optimization are added to the primer features. While primer pairs are being optimized, several status pages provide overviews about the progress. Primer pairs that have been optimized and approved move on to the next stage. For the *Optimization* section all selections retrieve a group of documents matching the selection criteria. The following options are implemented:

- New
- In Progress (All)
- In Progress (Tested:Yes/Worked:No)
- In Progress (Tested:Yes/Worked:Weak)
- In Progress (Tested:Yes/Worked:Mispr.)

- In Progress (Tested:No/Worked:Yes)
- Approved
- Discontinued

*New* optimization data simply displays the release date, the locus and primerpair. Clicking on the locus/primerpair name retrieves a data entry form which contains information previously collected, such as locus/primerpair name and allows the experimental conditions of the optimization process to be entered. Once the optimization form is submitted by clicking on the *Submit* button, the Perl script pair stores the data according to its status. Only data which has been newly collected is submitted to the Primer database file. Data previously collected for the target locus cannot be modified at this time.

Since *In Progress* browsers can become rather large, the viewing criteria have been narrowed to specific stages in the optimization process. The columns of all *In Progress* browsers include: the starting date of the optimization process, the PCR conditions and the locus/primerpair name. Clicking on the locus/primerpair name retrieves the document.

Once primers have been approved for screening, their data can no longer be changed. Selecting approved data displays text view only.

Primers may be discontinued and set aside for later study.

See Appendix C for a detailed optimization document and lines of *In progress* browser.

### 2.1.3 Screening Data Entry

The screening of the libraries is currently done in four experiments testing a particular pool. Analogous to the primer optimization, screening information is entered during the process. Status pages allow the retrieval of results for each of the screening steps. See Appendix D for an example of a detailed screening document and a sample of an *In Progress* browser display.

The following browsers are currently implemented:

- New
- In Progress (All)
- In Progress(Plate Pool)

- In Progress (Matrix)
- In Progress (3rd Screening)
- In Progress (Confirmation Screening)
- In Progress (Failed)
- Approved

The columns of the *New* browser include the release date, the PCR conditions and the locus/primerpair name. Clicking on the locus/primerpair name retrieves a data entry form with input fields for the screening process.

All *In Progress* browsers reflect the different stages of the screening process. Each line displays the starting date of the screening process, the screening run results, the coordinates, if applicable, and the locus/primer pair name. Clicking on the name retrieves the document. If a screening run fails, data is not discarded but discontinued and can be retrieved from the *Failed* browser for further testing. See Appendix D for a detailed view of a screening browser and data entry document.

#### 2.1.4 Mapping Data Entry

Once a P1 clone has been identified, the coordinates are transferred to the mapping section. The FISH mapping information is again captured in a set of data entry pages and added to the notebook. Probes that have finished this process are assigned an RMC number, i.e., a unique number within the Resource for Molecular Cytogenetics, and moved on to the last section.

See Appendix E for an example of a detailed mapping document and a sample of an *In progress* and *done* browser. Data for this stage can be entered from the following selections:

- New
- Clones from other sources
- In Progress (All)
- In Progress (On Target)
- In Progress (Weak)
- In Progress (Wrong Chromosome)
- In Progress (No Signal)
- In Progress (On Ice)

- Done / With RMC Number
- Public / With RMC Number

Each line of the *New* mapping browser displays the date of release, the locus/primerpair name and the coordinate. Clicking on the coordinate displays a data entry form with fields to record the mapping of probes.

If the clones for the FISH mapping have not been optimized and screened at the Resource, all data is entered at this point. To be consistent with the previous data submission, all target locus information is stored in the targetdb, primerpair information in the primerdb, and the coordinate and library information, if available, is stored in the coorddb. See Appendix E, Figure 22. Currently development of chimeric probes is under testing but has not been released.

Each line of an *In Progress* browser displays starting date of the mapping process, the experiment number, the chromosome map, the locus/primer-name and the coordinate. Clicking on the coordinate retrieves the document for updating.

### 2.1.5 Finished Probes

The last section of the notebook holds the probe information until the probes are to be made public. The *Done* browser selects all completed mapped coordinates which have been assigned an RMC number. Upon initial assignment of a number the probe is restricted, i.e. only RMC group members may view data about this probe.

The *Public* browser reflects all probes which have been made accessible to the public. The data corresponds with the data on the public web server.

Currently, the process of moving the data to the public web server is not automated. The laboratory notebook has not been tested sufficiently to move accumulated data from the development server to the public server.

## 2.2 Browsers

The second section of the electronic notebook, the browsers, provides views of all approved data from all stages. All members have access to the browser section. This section does not allow the user to update data, however, members have the ability to add comments to retrieved documents. Essentially, the purpose of this section is to monitor the progress of the mapping process. The underlying data for the browser should be

the data accumulated in the data entry section. However, as mentioned before, not enough data has been entered electronically to display in the browsers. Currently, the data displayed in the browser section of the development web server corresponds with the data from UCSF of all finished probes, public and non public. This data originates in an excel spreadsheet and has been manipulated by Perl scripts to be displayed. Data may be viewed based on the following selections:

#### **2.2.1 Range by Chromosome**

This selection provides a view of all probes developed at the Resource. All completed mapped coordinates which have been assigned an rmc number are displayed. For each chromosome a browser displays all mapped probes using the following information in each line: RMC number, chromosome map, gene/locus name, FLpter and coordinate.

RMC number, gene/locus name, and chromosome map are links in each line of the browser. Clicking on the RMC number retrieves the detailed document. The gene/locus name, if it is given, links to GDB, based on the GDB ID number, to retrieve a detailed GDB document. The chromosome map, if available, is also a link to GDB and retrieves all probes for that region.

See Appendix F for examples of browsers and detailed documents.

#### **2.2.2 Search for a specific probe**

A search engine allows the user to retrieve a specific probe or a range of probes. The search may be based on one of the following criteria: RMC number, chromosome map, FLpter, and gene/locus name. An on-line help is available to the user to ensure the correct use of the search criteria.

#### **2.2.3 Display Map**

The physical map of each chromosome is displayed with the available probes ordered according to their FLpter.

#### **2.2.4 Probes in Progress**

Probes may be viewed at each stage in their approved status. Users may select probes from the following stages:

- Approved Locus/PrimerPair

Each line of the browser displays the locus and primer name of all approved Locus/PrimerPair and the date of approval, i.e., when the

primer has been released for optimization. Clicking on the locus/primer-pair name shows a document with the detailed information.

- Approved optimized primers

This selection shows a browser with all optimized primer pairs that have been approved for screening. The date of approval, PCR conditions and the locus primerpair name is shown in each line of the browser. Clicking on the locus/primerpair name brings up the document with the detailed information for viewing.

- Approved screened primers

This selection shows a browser with all screened primer pairs approved for mapping. Each line of the browser contains the date of approval, the locus/primerpair name and the coordinate. Clicking on the coordinate brings up a detailed document for viewing.

### 2.3 Directory Structure - Perl Scripts

The data entry part of the laboratory notebook is stored on the machine jenny.lbl.gov. The home directory for the data entry Web server is located in

```
/home/zorn/local/httpd
```

(from now on referred to as ~/). All HTML concerning Resource documents are stored in subdirectories of

```
~/htdocs/rmc
```

The directory structure reflects the stages in the mapping process (Figure 1), e.g.,

```
~/htdocs/rmc/public
```

holds all HTML documents for the RMC group section. The directory for the data-entry documents,

```
~/htdocs/rmc/data-entry
```

is divided into primer-resources and probe-resources. The Probe Mapping section of the laboratory notebook server is accessible

only with password. Each member of the Resource must have their login name with a corresponding password entered to be able to access *RMC Group* from the Home Page on jenny.lbl.gov.

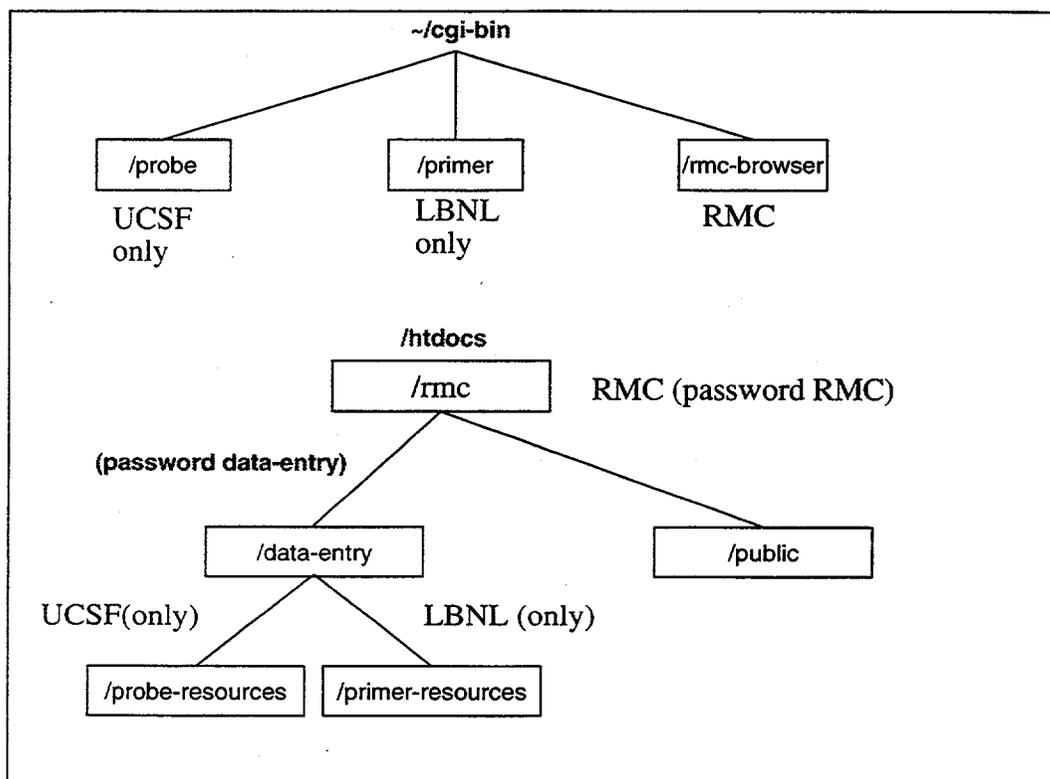
The *RMC group* and its members file is stored in

~/htdocs/.htgroup

This file must include all login names to the RMC directory. The data-entry section of the server has its own additional password protection.

FIGURE 4.

Directory Structure



### 2.3.1 Data Entry Structure

Perl scripts respond to the WWW client requests by displaying new forms and/or browsers with the requested data. All scripts manipulating data-entry for LBNL scientists are located in a subdirectory of

~/cgi-bin

To ensure the integrity of the data developed at the different sites of the Resource (LBNL and UCSF), each site can only update data in its respective sections of the notebook.

Scripts manipulating the Target, Primer and Screening sections are stored in

```
~/cgi-bin/primer
```

and scripts for the mapping section are stored in

```
~/cgi-bin/probe.
```

The scripts for the RMC group browsers are located in

```
~/cgi-bin/rmc-browser
```

(see Figure 4).

Because of the division of labor, LBNL has access to the new target/locus, the primer optimization, and the screening section of the data entry. The directory `~/cgi-bin/primer` contains different password entries than the directory `~/cgi-bin/probe`. The access to these directories is determined at the `~/htdocs/rmc` directory level. Anybody selecting *Data Entry* is requested to supply a password at that time. The password file for the data entry section is stored in

```
~/htdocs/rmc/.htpasswd and  
~/htdocs/rmc/.htgroup
```

To add and/or delete a user and their password to the data entry section the `httpd passwd` script has to be invoked. First add the login name of the new user to `~/htdocs/.htgroup` at the end of the `rmc` group. Type:

```
htpasswd /home/zorn/local/httpd/htdocs/.htpasswd  
<userlogin>
```

The script will prompt you for the password to be added.

Currently, the separation of the different passwords for different users from UCSF and LBNL has been disabled because of testing the user interface and underlying dbms files. To invoke the access restriction, either start a new password file, storing it according to WWW restric-

tions, or move the previously enacted files to the old location. (In the directory `~/htdocs/data-entry/primer-resources` move the files `htaccess` and `htpasswd` to `.htaccess` and `.htpasswd`, respectively, and in the directory `~/htdocs/data-entry/probe-resources` move the `htaccess` and `htpasswd` to `.htaccess` and `.htpasswd`, respectively. This will ensure that only the users in the described sections have access to the separate data entry forms.

All Unix dbms data files are stored in subdirectories of

`~/htdocs/supports.`

All data entries from the notebook are stored in

`~/supports/primerdata`

All data from the target locus is stored in the dbms files `~/supports/primerdata/targetdb.dir` and `targetdb.pag`. A Perl script `targetdb` can display the data in those dbms files. Data from the primer optimization is stored in the files `primerdb.dir` and `primerdb.pag`; again, a script `primerdb` allows the data to be displayed. `coorddb.dir` and `coorddb.pag` hold all screening data; `probedb.dir` and `probedb.pag` contain all information from the mapping process; the script `probedb` displays the data. Once data is made public, a copy of the relevant information for display in the RMC browsers is made to the `rmcdb.dir` and `rmcdb.pag`. A README file in the directory

`~/htdocs/supports/primerdata`

contains information on the various fields in the dbms files.

Once probes from the data entry section are made publicly available the data is stored in the directory

`~/supports/public`

using the files `rmcdb.dir` and `rmcdb.pag`. Currently, there is no automated procedure to take data which has become public into the underlying dbms file for the browsers. The data displayed in these browsers comes from a UCSF MS EXCEL spreadsheet which has been read into the dbms files after collecting data from GDB and storing these links in the dbms file.

Scripts that collect references and links from the public databases and fill the RMC database file can be found in

```
~/supports/primerdata/fillldb-scripts
```

A README file in this directory explains the sequence of steps.

### 2.3.2 Browser Structure

To describe the design of this application we use a newly proposed data model, the Relationship Management Data Model (RMM)\*. This data model is a set of logical objects used to provide an abstraction of a portion of the real world. Hypermedia applications present a challenge to design modeling. These application involve a complex domain and to capture it in a model is not easily done with existing methodologies. The Relationship Management Data Model seemed to be a methodology easily adapted for the design of hypermedia applications.

The focus of the RMM design methodology calls for a sequence of design step which we have included in this document. Figure 5 shows all the steps for the RMM methodology. Shaded are the ones included in this document. The requirement documents such as the ER diagram and the various navigation designs are described below. The corresponding appendices include the user-interface and screen design.

---

\*Tomas Isakowitz, Edward A. Stohr, and P. Balasubramanian, *RMM: A Methodology for Structured Hypermedia Design*, Communications of the ACM 38 (8) 34-44 (1995).

FIGURE 5. The RMM methodology stages

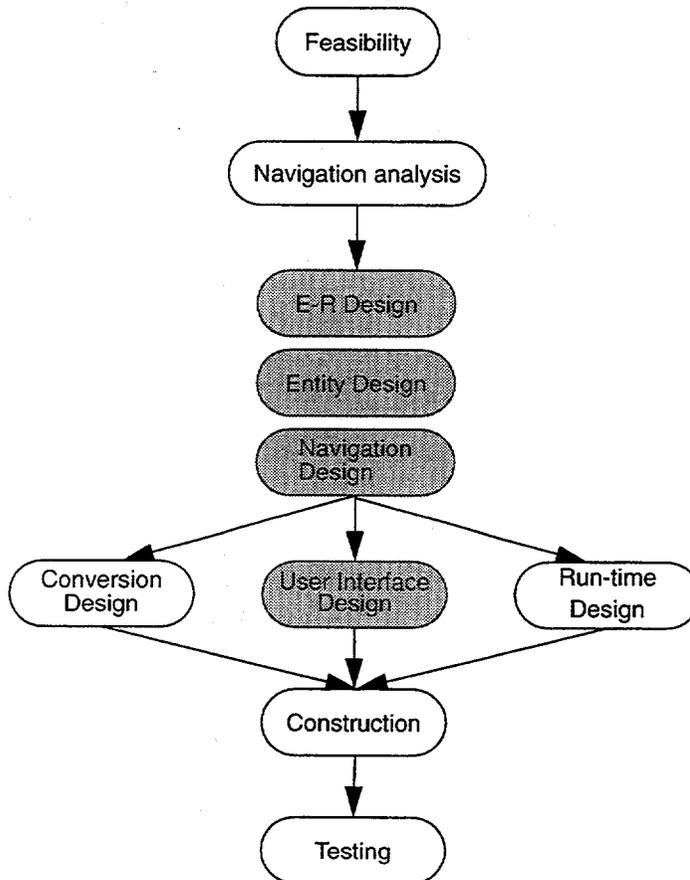
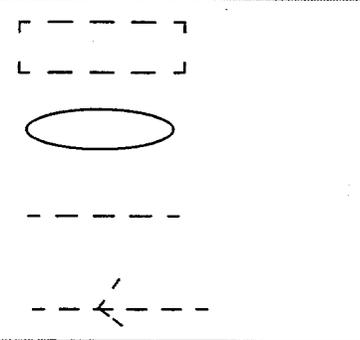
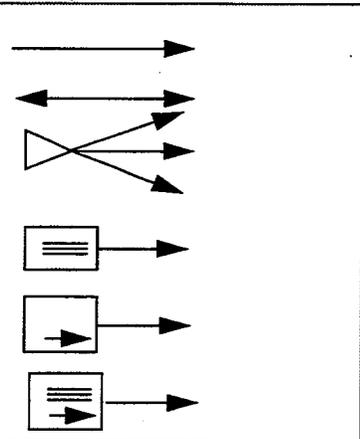


Figure 6 gives an overview of the terminology used to describe the Relationship Management Data Model of the Server. RMM lends itself to describe the regular structure of the Resource browsers. There are classes of objects and definable relationships between these classes and multiple instances of objects within each class (see the ER Diagram Figure 7).

The ER domain primitives are well known. The RMM domain primitive, *slice*, is used to group attributes of one entity in order to present the entity under different aspect or to different user groups. What is really interesting in this RMM model are the access primitives. They are specific to hypermedia application and are used to describe the paths that will enable hypertext navigation.

FIGURE 6.

The Relationship Management Data Model Primitives

ER Domain Primitives	<p>Entities</p> <p>Attributes</p> <p>One-One Associative Relationship</p> <p>One-Many Associative Relationship</p>	
RMD Domain Primitives	Slices	
Access Primitives	<p>Unidirectional Link</p> <p>Bidirectional Link</p> <p>Grouping</p> <p>Conditional Index</p> <p>Conditional Guided Tour</p> <p>Conditional Indexed Guided Tour</p>	

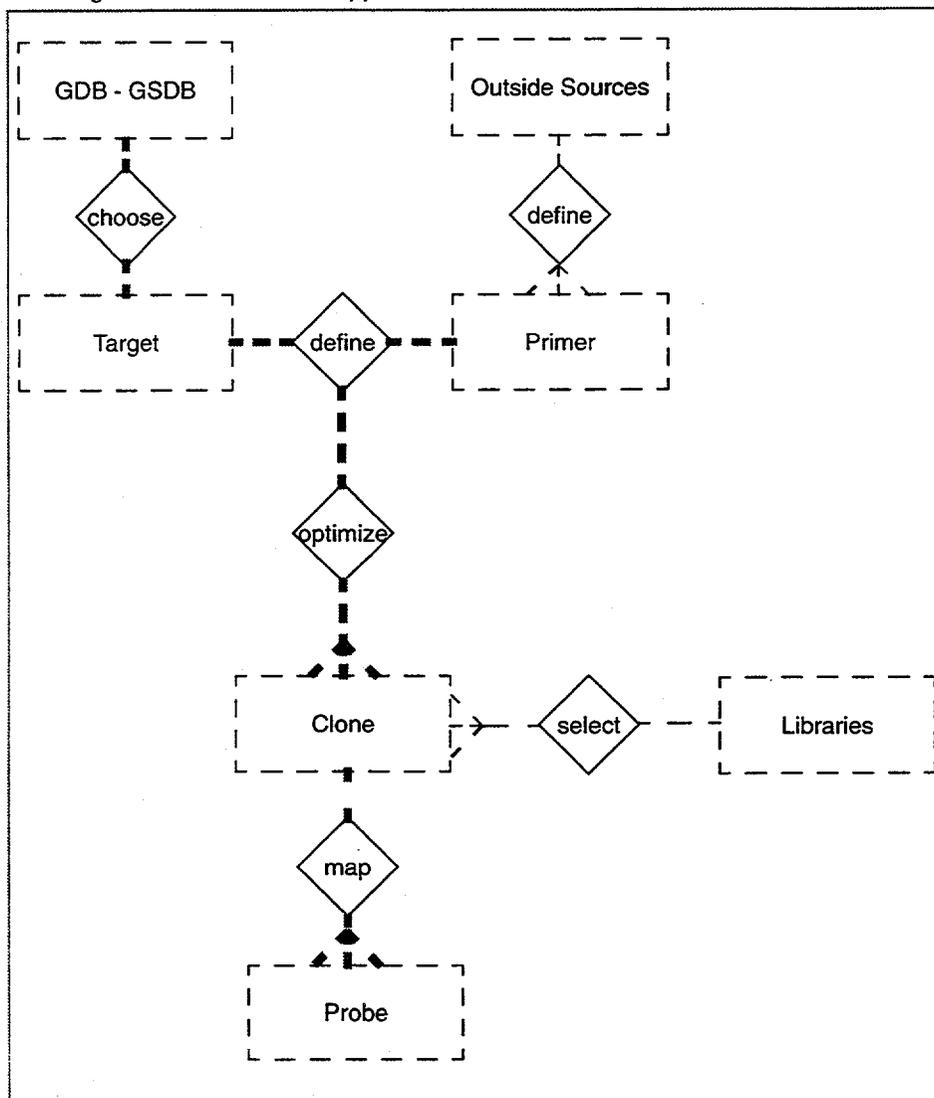
The uni and bidirectional links are used to navigate only within the boundaries of the instances of one entity, i.e. between slices. The grouping construct is a menu like mechanism that shows access to other parts of the application. The index and the guided tour show navigation across different entities.

### 2.3.3 Entity-Relationship Diagram

The entity-relationship diagram is based on the perception of the real world which consists of objects and relationships between these objects. An object, i.e., an entity, is represented as a rectangle. The relational model presented in Figure 7 captures the user requirements and their analysis. This model describes how the user, i.e., the biologist perceives the database.

FIGURE 7.

E-R Diagram for the Resource application



The model is primarily used to design database systems. The RMM methodology uses the ER diagram to analyze the domain and identify relationships across which navigation is supported. The relevant entities and their relations are selected as the basis for the hypermedia application. Figure 7 depicts the ER diagram for the Resource database and indicates in thick lines which associate relationships are supported by navigation. For example, given a collection of Primer instances, links are provided to instances of the corresponding targets and to the respective detailed information in the public databases such as GDB and GSDB.

#### 2.3.4 Slice Design

An entity may consist of a large number of attributes of different nature. For example the entity probe has attributes pertaining to hybridization quality, laboratory notes, and FISH mapping. Not all attributes are of interest to all users. Attributes are thus grouped into slices, according to a particular view. The entity probe is presented in the application in three slices. Depending on the permissions of the users, each slice may be selected for viewing. Scientists involved in the actual research of the probes may view all attributes contained in the slice *Research Area*. Members of the RMC group are presented with a subset of these attributes in the slice *RMC group*. Finally, the general public has access to yet a smaller number of attributes of the probe.

Attributes for the data-entry slice:

Laboratory specific

- experiment number
- date frozen
- internal ID number

Hybridization specific

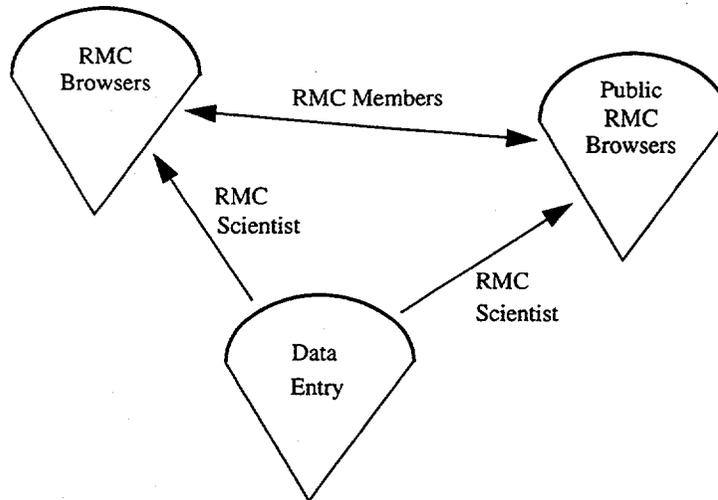
- interphase
- metaphase
- noise level

FISH mapping specific

- chromosome map
- FLpter
- standard error
- number of experiments
- donor
- references
- comments
- restriction

The other entities, such as target, primer, and coordinates are represented in a similar fashion. Each of the different user group views a different slice of the entity. See Appendix B for more details.

FIGURE 8. Slice Design for entity probe



### 2.3.5 Navigation Design

To describe the hypertext navigation paths each associative relationship appearing in the ER diagram is analyzed. Requirement analysis determines if an entity should be made accessible for navigation. For ease of view and explanation the navigation design is broken down for each of the menus resolving from displaying the entities of the ER Diagram. Looking at Figure 7, the ER Diagram, all entities except the Outside Sources and the libraries have been made accessible via navigation. Detailed documents retrieve information from all entities to be displayed.

FIGURE 9. Navigational Design Main

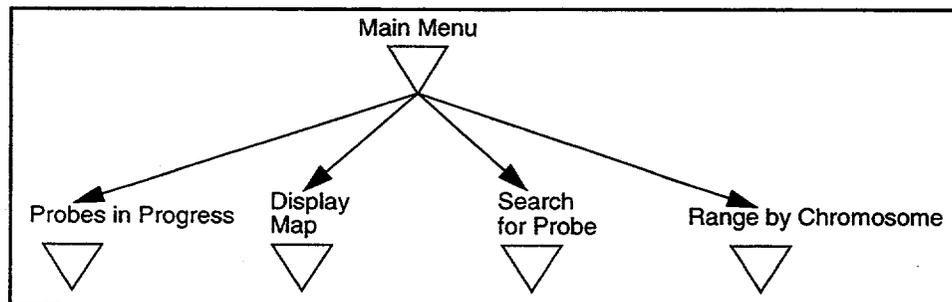
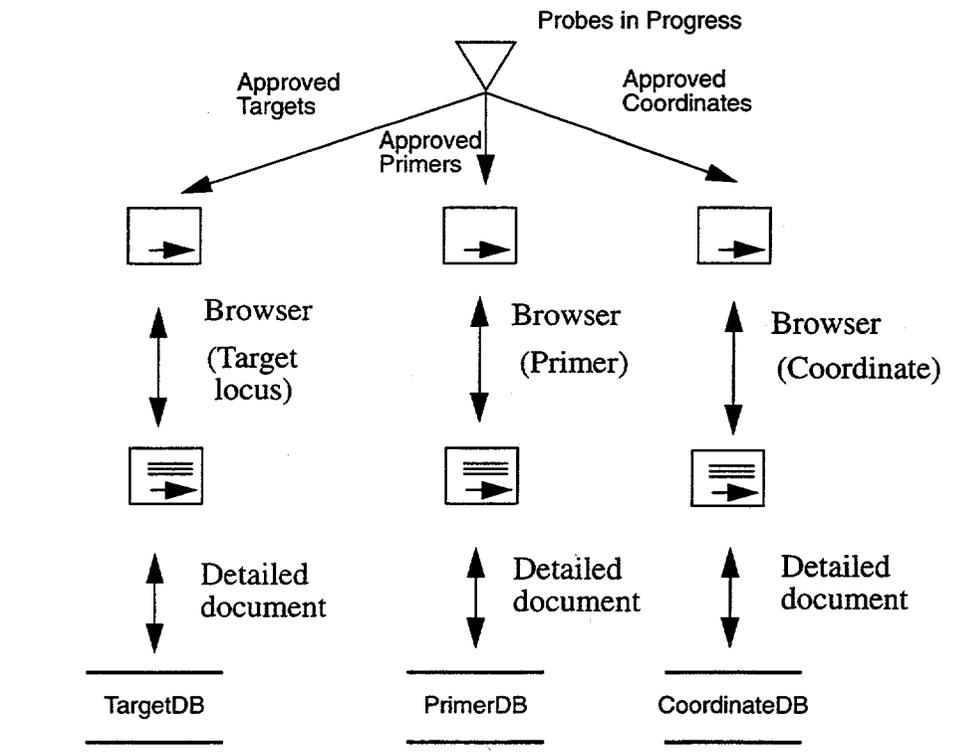


Figure 9 describes the main menu selection from which navigation through the system is possible. Each of the submenu selections is described in separate figures.

FIGURE 10.

Navigational Design: Probes in Progress



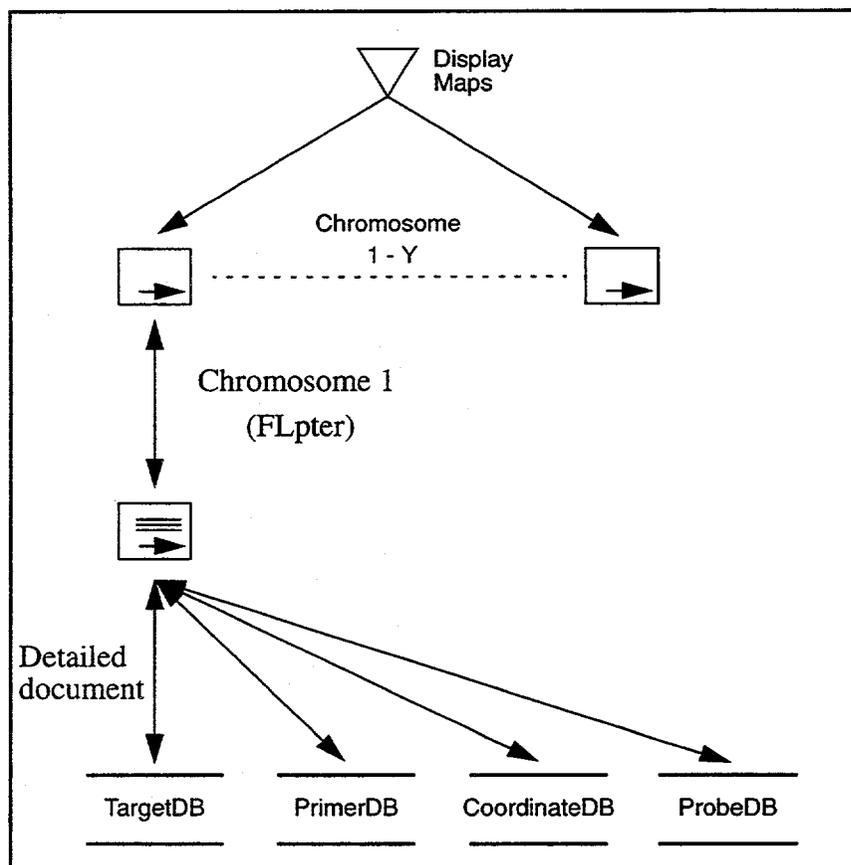
The first submenu selection, *Probes in Progress*, is depicted in Figure 10. The selection of this link gives the user access to another menu-like mechanism with three choices: Approved targets, approved primers, and approved coordinates produced at the Resource. This indexed table of each choice provides the user with direct access to the individual document.

In the case of the approved target selection each line of the Target locus table or browser includes the approval data and the name of the target locus. The indexed table browser of the approved primers produces for each instance of this entity a date of approval, the PCR conditions, the name of the target locus and the name of the primerpair. The selection *approved coordinates* retrieves an indexed table of instances of the coordinate entity. Each line includes the approval date, the target locus and

primerpair name and the coordinate. Essentially, the browsers and the detailed documents look the same as for the data entry section.

FIGURE 11.

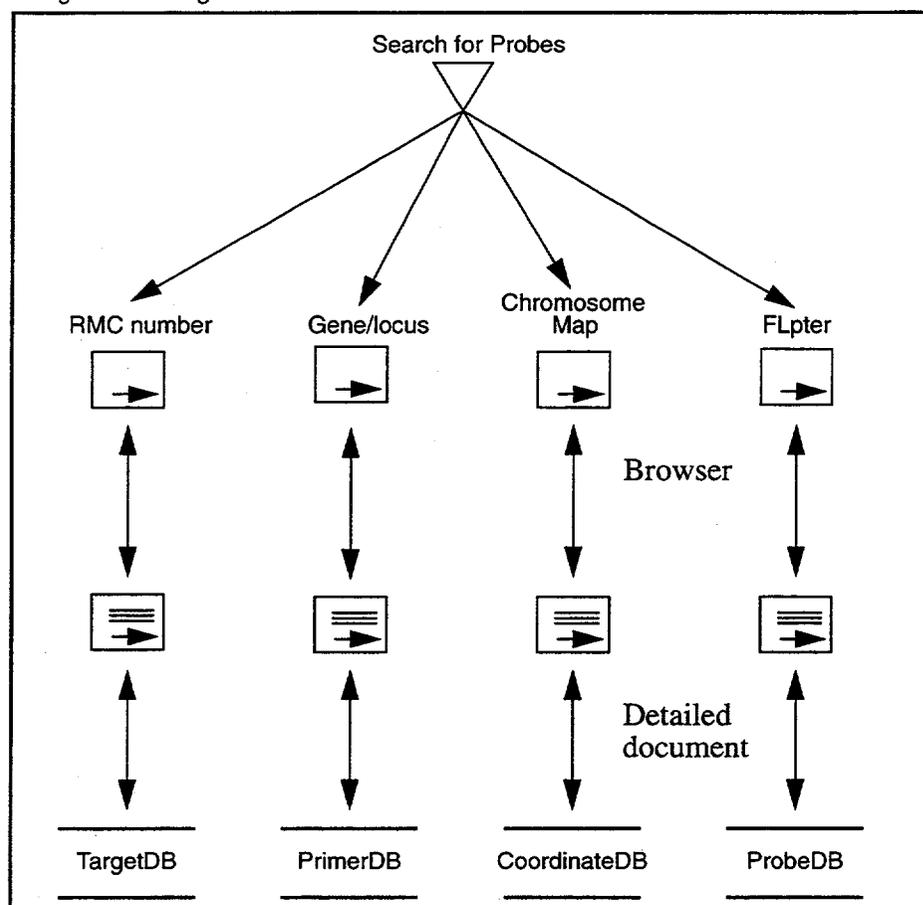
Navigational Design: Display Maps



The next submenu selection, *Display Maps*, takes data for one chromosome mapped at the Resource and displays the information according to the fractional length measurement obtained. The submenu links the user to another HTML document (`~/rmc/rmc-browser/P-Browse.html`) to select the chromosome of which the user wants to see an ordered map display. Clicking on the chromosome displays the list of existing documents. Each RMC number is a link to a detailed document about the probe. If the target/locus and primer pairs were obtained with a valid GDB number this GDB number is given as a link to the GDB database. Again, the detailed Resource document is retrieved using indexed condition. Figure 11 depicts the navigational design of the selection.

FIGURE 12.

Navigational Design: Search for Probe

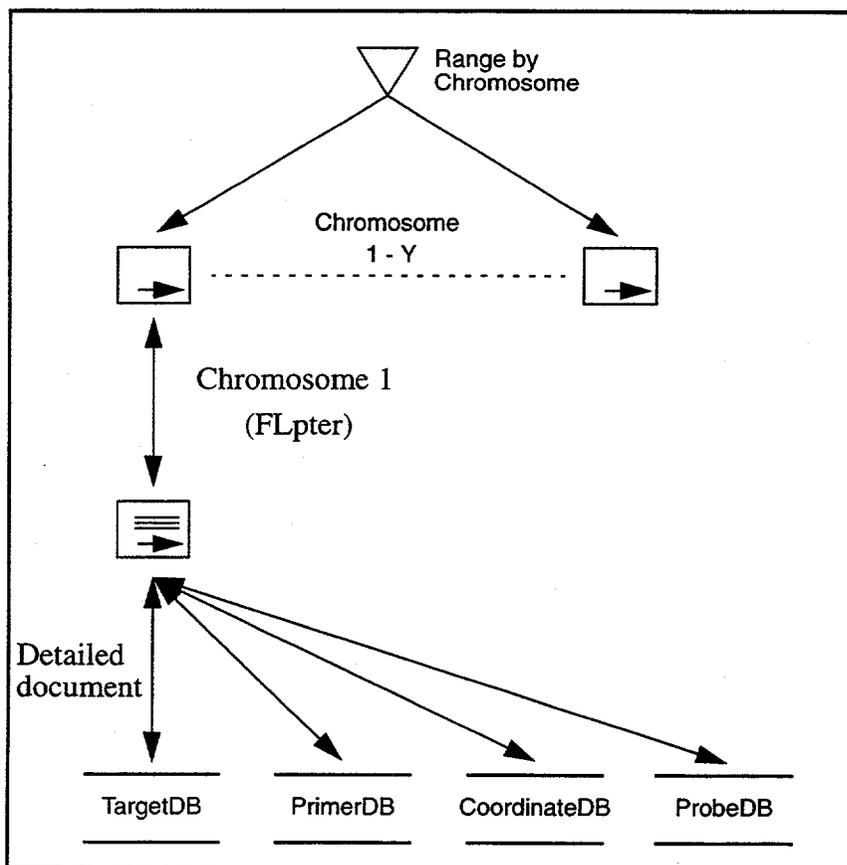


The third submenu selection, *Search for specific probe*, is a search engine for the Resource data. Based on either the RMC number, the gene/locus name, chromosome map and/or the FLpter, and various wildcard implementations, a specific document or a category of documents may be retrieved. On-line help is provided for this search engine. Figure 12 depicts the navigational design of the search engine.

The last selection, range by chromosome, links to an HTML document, (`~/rnc/public/BROWSER.html`) from which the user may choose the chromosome number he wants to view. All available documents of this chromosome will be displayed. Several links are provided from each line of the browser. The RMC number retrieves the detailed document, the chromosome map links to GDB based on that chromosome, and the gene/locus name links to the detailed document in GDB for that gene locus.

FIGURE 13.

Navigational Design: Range by Chromosome



### 3.0 World Wide Web Server - Public Use

---

In order to distribute information generated at the Resource, we built a public web interface that allows outside researchers to view our data. The data can be accessed through a search engine that returns a list of probes. Each probe is a hypertext link to detailed information about the probe, the target, primer pairs used, etc. Whenever a pointer to a public database, e.g., GDB, is present in the database, a link into the public database is created. The detailed probe form features a request button that forwards a probe request to the Resource. The public RMC web server integrates local data with FISH mapping results from collaborators. Request forms are mailed directly to the persons responsible for physically distributing the probes. In addition to probe mapping information, the web server also features sample images for users of Resource imaging software, pages about policies, and procedures employed in the Resource. A summary usage statistic presents a graphical view of the history of access to the server. For security reasons, the public web server runs on a separate workstation on its own data set that is physically distinct from the data used in the laboratory notebook. The following URL provides access to the Resource for Molecular Cytogenetics Web Home Page:

<http://rmc-www.lbl.gov/>

The user has the following choices:

- Probe Mapping Data
- Images
- Policies-Procedures-Projects
- Usage Statistics

#### 3.1 Probe Mapping Data

This section is essentially the same as the browser section in the laboratory notebook. However, probes displayed here include third party probes, such as the ones available through VYSIS, and probes from Dave Ward's laboratory at Yale, in addition to all completed probes from the Resource. Data may be viewed based on following criteria:

- Range by chromosome
- Search for specific probe
- Display map

All data is stored on white.lbl.gov in the directory

```
/usr/local/www/htdocs/supports/testkey
```

The Unix dbms files rmcdb.dir and rmcdb.pag contain all publicly available probes. Currently the only way to display the public probes on white is to ftp the rmcdb.\* files from jenny.lbl.gov to /usr/local/www/htdocs/supports/testkey. However, some of the probes displayed on the development server are restricted. To delete the restricted probes from the public server use the script:

```
/usr/local/www/htdocs/supports/adm/delete-from-rmcdb
```

This script reads each probe in the rmcdb and checks if there is an entry in the restriction column. To run the script type

```
/usr/local/www/htdocs/supports/adm/delete-from-rmcdb > publicProbes.<month>.<day>
```

The directory /usr/local/www/cgi-bin/rmc-browser contains all scripts necessary to display the data on the public web server.

### **3.2 Images**

Users may view some selected images produced at the Resource. The images depict hybridization of FISH applications. They have a short description and may be downloaded for reference.

### **3.3 Policies-Projects-Procedures**

This part of the public server displays the general policy and statement of purpose of the Resource for Molecular Cytogenetics. It also contains abstracts of projects completed and/or pursued at the Resources.

### **3.4 Usage Statistics**

The access of the public server is charted as a monthly usage chart and a by county usage chart. Each month the chart is updated. The user has access to all collected charts and can run comparisons.

Currently, the usage statistics are manually updated each month. The README file in the directory

```
/usr/local/www/htdocs/wwwstats
```

---

## Acknowledgments

---

contains instructions on how to generate the necessary files in order to update the display.

### 4.0 Acknowledgments

---

This work was supported by the Director, Office of Energy Research, Office of Health and Environmental Research, UC-408, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

---

## Acknowledgments

---

---

---

---

## Appendix A.

### README Files

Each of the ~/cgi-bin subdirectories contains scripts that respond to the clients request. Each directory contains a README file, which is reprinted here and briefly describes what the main scripts do.

---

FIGURE 14.

/home/zorn/local/httpd/cgi-bin/primer/README

This directory contains all scripts for the data entry section of the Locus/Primerpair, Optimization and Screening part of the server. The main scripts are:

**primer.pl** - all data submission, search, parse, display subroutines.

**TargetForm.pl** - first data entry form, from here the search engine for GDB or GSDB is called and data retrieved from these databases. The data is then displayed with the new LocusForm to be stored in the database.

For the retrieval from GDB,GSDB the **URL.pl** etc. are needed.

**LocusForm.pl** - text and data entry Locus Form

**PrimerPairForm.pl** - text and data entry for the PrimerPair Optimization form.

**HitForm.pl** - text and data entry for the Screening Form.

**AllHeader.pl** - all headers to be displayed and search engines to retrieve data from db for browsers.

**DisplayData.pl** - routines to correctly format data for browser display.

---

FIGURE 15.

/home/zorn/local/httpd/cgi-bin/probe/README

This dir contains all scripts for the UCSF part of the mapping process.

---

Different dir because of access script.

**ProbeSubmit.pl** - submits and retrieves data from the db.

**ProbeDisplay.pl** - arranges data for browser display.

ProbeForm.pl - just that.

**ProbeHeader.pl** - header and format of browser lines, etc.

---

FIGURE 16.

/home/zorn/local/httpd/cgi-bin/rmc-browser/README

All scripts in this directory manipulate the RMC Group browser on the development server.

The dir /home/zorn/local/httpd/cgi-bin/rmc-browser/loading-scripts contains all scripts to take the various input spreadsheets and make them into the rmc database.

The dir /home/zorn/local/httpd/cgi-bin/rmc-browser/inprogress contains all scripts to display the various browser of for probes in progress.

**cgi-too.pl** - This script contains all subroutines to retrieve data from the data base (read only) and format it into the various browsers. It contains the search engines for the browser search.

**format.pl** - This script contains the format for the probe text forms.

the rest of the shell scripts invoke a Perl script by the same name, e.g., **mail-requested**

---

---

## Appendix B.

## Target Data Entry Forms

This appendix contains examples of all browsers and forms associated with target locus information. Figure 17 depicts a new Target data entry form in which the user indicates where the information should be referenced, i.e., GDB, GSDB, or an internal name. Figure 18 shows the new Target form after information from public databases has automatically been retrieved by Perl scripts. If the retrieved information is correct the user can move on to the form shown in Figure 19.

---

FIGURE 17.

Form to retrieve information from public databases

● Retrieve Information

[Home] [RmcGrain] [DataEntry] [New Target/Frame]

Clear Fields [Retrieve Information]

Enter Target Name here if it is a valid GDB Symbol/GDB Locus Name  
More information about the Target Name will be automatically retrieved from GDB  
GDB Target Name: [ ] Type: GDB Symbol

Enter Target Name here if it is a valid GSDB Locus Name/GSDB Accession Number  
More information about the Target Name will be automatically retrieved from GSDB  
GSDB Target Name: [ ] Type: GSDB Locus Name

Enter Target Name here if it is an internal Name  
Internal Name: [ ]

Clicking this button will retrieve data from GDB or GSDB

FIGURE 18.

Approved target locus text form

GDB link →

APPROVED LOCUS INFORMATION	
<a href="#">[Home]</a> <a href="#">[RmcGroup]</a> <a href="#">[DataEntry]</a>	
Locus Name:	D13S152
Created on:	(9/14/94)
Approved on:	(17Aug95)
Type:	GDB Symbol
<b>GDB ACCESS</b>	
Other Names:	DNA Segment, single copy probes AFM037xa
Requester:	Colin Collins
Conditions:	
Expected Map Position:	13q21.2-q21.3
Primer Pair:	F1/R1
Primer Source:	Research Genetics
Comments:	

FIGURE 19. New target locus form

**LOCUS INFORMATION**

[Home] [RmcGroup] [DataEntry]

---

Target Name: D4S1492  
 Type: GDB Symbol  
 Created: 17 Sep 95

---

GDB ID: G00-366-116

Other Names:

FNA segment, single copy probe, CHL6, GATA

---

Requester:

Conditions:

Expected Map Position:

**New Internal Locus Data**

[Home] [RmcGroup] [DataEntry]

Multiple Entries: YES NO

---

Target Name: Internal Name  
 Type: Internal Name  
 Other Names:

---

Requester:

Conditions:

Expected Map Position:

---

Primer Pair:

Primer Source:

Name:

Sequence:

Name:

Sequence:

Do you want to enter additional Primer Pairs at this time? YES NO

Release data for Optimization at this time? YES NO

Comments:

Data and Link GDB

Default Values



## Appendix D.

## Screening

Documents for the library *Screening* are shown here, e.g., the browser for newly released primerpairs and the data entry form.

FIGURE 21.

Screening browser display and new data entry form

**NEW PRIMER PAIRS FOR FURTHER SCREENING**  
[Home] [FuncGroup] [DataEntry] [Screening]

Released	Size/MgCl/Temp	Base Tested	Worked	Locus/PrimerPair
17Aug95	212-236/1.5/52	-	Yes	D13S153/F1/B1
17Aug95	243-277/1.5/52	-	Yes	D13S154/F1/B1
17Aug95	204-218/1.5/52	-	Yes	D13S155/F1/B1
17Aug95	250-264/3/48	-	Yes	D13S157/F1/B1
17Aug95	99-113/3/48	-	Yes	D13S158/F1/B1
17Aug95	169-203/3/52			
17Aug95	117-127/3/48			
17Aug95	158-173/3/48			

Library:

Insert Size:

If Source is "other" please describe here:

Locus:  Expected Map:

PCR Conditions:

Tested	Worked	Base	Size(bp)	Mgc(mg)	Temp(C)
Yes	Yes	-	212-236	1.5	52

Notebook#  NotebookPg#

Screening Run1:

Screening Run2:

Screening Run3:

Coordinates:  Failed:

previously stored data →

Appendix E.

Mapping

New coordinates from the Resource are listed as a browser. Each line of the browser contains the release date, experiment number, the target locus/primerpair name, and the coordinate. Figure 22 displays a line of a browser and the form retrieved when clicking on the coordinate.

FIGURE 22. Mapping browser and new data entry form

previously stored data

Internal information

Hybridization Information

FISH map Information

STARTED	EXP.	CHR. MAP	LOCUS	COORDINATE
11Sep95	12	1q123	N/A	12b3

ADD MAPPING DATA

Submit RMCDB

LOCUS: N/A

Coordinate/Clone: 12b3

LIBRARY: DuPont, B

Hybridization Quality:

Experiment: 12

Date Frozen: 5-9-95

Metaphase: ok Interphase: clean

Submit RMCDB

FISH Map

RMCNUMBER: RMC81P002

Chromosome Map: 1q123

Mapped FLpter: 10-01

+/-S.E.: 0.0001

N: 1

Metaphase Donor: 1

Restricted: YES NO

Discontinue experiment: NO

Figure 23 depicts a data entry form for clones from other sources beside the RMC.

FIGURE 23. Clones from other sources

NEW CLONE FROM OTHER SOURCE  
[Home] [RmcGroup] [DataEntry] [Mapping]

Target locus:  
N/A  
Name1:  
N/A  
Name2:  
N/A

Library  
Source: DuPont.B Type: Pl Species: human  
Vector: pAD10SacBI Growth Condition: 37C kanamycin Site:  
Eam HI

Coordinate:  
Insert Size(kb):  
Experiment:

Hybridization Quality:  
Date Frozen:  
Metaphase: ok Interphase: score clean

Clear Fields Submit to Database

FISH Map  
Chromosome Map:  
Mapped FL pter:  
+/-S.E.:  
N:  
Metaphase Donor:

Restricted: YES NO  
Discontinue experiment: NO  
Additional Experiment: YES NO  
Select RMC number at this time: YES NO

## Appendix F.

## Browsers

The following are examples of browsers and detailed text documents retrieved on either the development or the public web server.

FIGURE 24. Range by chromosome

guided tour

**RESOURCE for MOLECULAR CYT**

SEARCH RESULTS

[Home] [RmcGroup] [PublicBrowsers]

RANGE BY CHROMOSOME PUBLIC BROWSER

[Home] [RmcGroup] [PublicBrowsers]

Chromosome 01 Chromosome 02 Chromosome 03 Chromosome 04 Chromosome 05 Chromosome 06 Chromosome 07 Chromosome 08 Chromosome 09 Chromosome 10 Chromosome 11 Chromosome 12 Chromosome 13 Chromosome 14 Chromosome 15 Chromosome 16 Chromosome 17 Chromosome 18 Chromosome 19 Chromosome 20 Chromosome 21 Chromosome 22 Chromosome X Chromosome Y

**LBL/UCSF Resource for Molecular Cytogenetics**

RMC ID	CHR. NO.	GENE/LOCUS	Flpter	COORD
<b>RMC02P029</b>	3q22	D3S1292	0.673	124F8
RMC03P004	3pter	D3S1311	0.984	109F1
RMC03P005	3q13	D3S1318	0.591	53D3
RMC03P006	3q24	D3S1308	0.754	46C5
RMC03P008	3q26-q27	D1nt2	0.856	76D8
RMC03P009	3q26	D1nt2	0.863	86H9
RMC03P010	3pter	D3S1305	0.979	75A11
		D3S1303	0.605	74D11
			0.552	2BD8

**LBL/UCSF Resource for Molecular Cytogenetics**

RMC02P029

[Home] [RmcGroup] [PublicBrowsers]

Probe Name: RMC02P029 Original Gene/Locus: D3S1292

This probe has been registered in GDB as 1251866

Overlapping Clones:

Flpter : 0.673 S.F. : 0.006 N : 11 Honor : 0

Chr. Map: 3q22 Internal ID# : 4362 Coordinates: 124F8

Source: Du Pont. BType: P1 Species: human Vector: pAd10SacIII

Insert Size: Insert Site: BamHI

Growth Condition: 37 C, Kanamycin

Comments:

References: Restricted:

RMC Member Comments:

This probe may be requested from the Resource for Molecular Cytogenetics

Add Comments

indexed table

detailed document

FIGURE 25.

Non-graphical map display of chromosome 3

**LBL/UCSF Resource for Molecular Cytogenetics**

☐ Chromosome 3

Fl.pter	Gene/Locus	Probe Number
<b>0.0</b>		
0.011	D3S1307	RMC03P064
0.062	D3S1263	RMC03P019
0.099		RMC03P001
<b>0.1</b>		
0.105	D3S1293	RMC03P012
0.133	RAB8	RMC03P060
0.134	D3S1266	RMC03P059
0.166	D3S1284	RMC03P013
0.190	D3S1298	RMC03P037
0.193	D3S1298	RMC03P038

FIGURE 26.

Search for specific probe

☐ SEARCH

[Home] [RmcGroup] [PublicBrowsers]

---

[Help Using the Search Form](#)

---

[Home] [RmcGroup] [PublicBrowsers]

RMC03P004	3qter	D3S1311	0.984	109P1
RMC03P005	3q13	D3S1310	0.591	5303
RMC03P006	3q24	D3S1308	0.754	4605
RMC03P008	3q26-q27	D1S12	0.856	7608
RMC03P009	3q26	D1S12	0.863	8809
RMC03P010	3qter	D3S1305	0.979	75A11
RMC03P011	3q13-q22	D3S1303	0.605	74B11
RMC03P014	3q12		0.552	2008

FIGURE 27.

Probes in progress

● LOCI APPROVED FOR OPTIMIZATION

[Home] [RmcGroup] [PublicBrowsers] [InProgress]

CREATED ON	PRIMEPAIRS	LOCUS
9/14/94	F1/B1	D13S152
9/14/94	F1/B1	D13S153
9/14/94	F1/B1	D13S154

● APPROVED LOCUS INFORMATION

[Home] [RmcGroup] [PublicBrowsers] [InProgress]

9/14/94

Locus Name: D13S152

Created on: (9/14/94)

Approved on: (17Aug95)

---

Type: STR Symbol

Other Names: DNA Segment, single copy probes AFM337a

---

Requester: Colin Collins

Conditions:

Expected Map Position: 13q21.2-q21.3

---

Primer Pair: F1/B1      Primer Source: Research Genetics

Comments:

## Appendix G.

## Guidi Tools

Non-graphical Map Integration Tools on the Web. These tools can be run from a shell and only the first selection has been converted to be run on the web.

FIGURE 28.

RMC FLpter Order

The screenshot shows a web interface for the 'RMC FLpter Order' tool. It includes a navigation menu with options like 'RMC Linkage Data Connections', 'WJ Contig Connections', and 'WJ VAC Connections'. The main area contains instructions for using the tool, a 'Chromosome' input field, and a 'Proceed' button. Below, an 'Output' section displays a table of results.

**RMC FLpter ORDER**

For example: rmlinkage20  
For all the RMC data on the given chromosome, find targets that have been genetically mapped in data available at GDB and CHLC. Present these in experimental sets ordered by genetic distance.

Chromosome:

Click here:

**Output**

d20s113		0.980	0.65	DWT20Y246
d20s115		0.144	0.11	DWT20Y248
d20s95		0.154	0.13	DWT20Y247
d20s169		0.334	0.31	DWT20Y249
d20s190		0.359	0.33	DWT20Y250
d20s197		0.734	0.70	DWT20Y251
d20s183		0.820	0.79	DWT20Y252
d20s162		0.895	0.66	DWT20Y253
cpc25n	P1	0.081	0.096	RMC20P105
d20s117	P1	0.099	0.009	RMC20P081
humbp2a	P1	0.113	0.014	RMC20P124
d20s117	P1	0.123	0.013	RMC20P053

---

---