Volume 12 Number 1 1984 Nucleic Acids Research

Recombinant DNA data management at the restriction and functional site level

David Shalloway and Norman R.Deering

Molecular and Cell Biology Program, Pennsylvania State University, University Park, PA 16802, USA

Received 23 August 1983

ABSTRACT

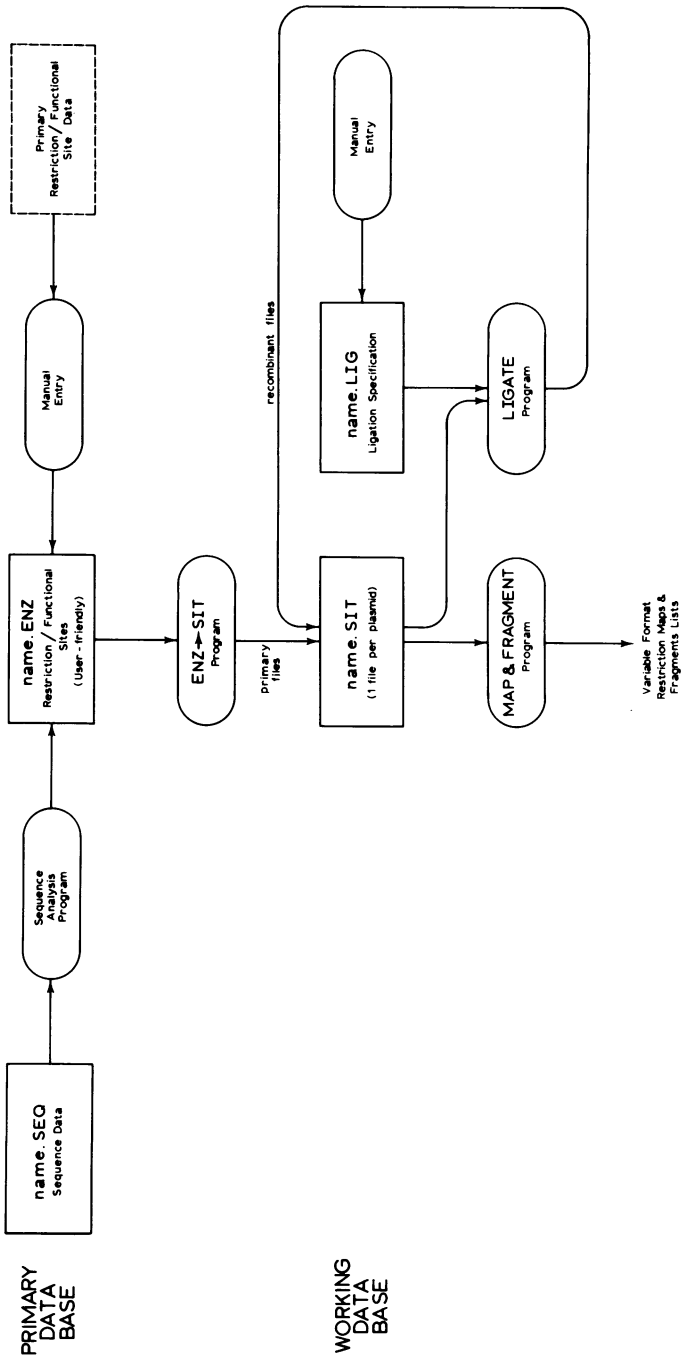
We have created a system to manage in a unified manner the restriction and functional site information required for design and analysis of recombinant DNA experiments. Primary source DNA data and recombinant clone specifications are used to generate recombinant restriction maps and restriction fragment lists. Sequence data, restriction-site data, and functional-site data may be combined in the data base. Interaction and output is user-friendly and versatile.

INTRODUCTION

The design and analysis of experiments involving recombinant DNA technology involve repetitive reference to recombinant plasmid restriction and functional site maps. These recombinant maps are usually constructed from restriction maps of primary (non-recombinant) DNA sources which were determined either by direct analysis of restriction enzyme mapping data or by computer-aided analysis of sequence data. They often also include the positions of important functional sequences such as promoters, exon boundaries, polyadenylation sites, etc. The manual generation of recombinant clone restriction maps using this data base becomes increasingly tedious and error-prone as the number and complexity of clones in the laboratory increases. Updating extensive files of recombinant restriction maps when new restriction enzymes or functional sequences are discovered or when errors are detected in the primary data base also becomes problematic.

Here we describe a computerized housekeeper, MATILDA (Mapping, Accessing and Transforming Information about Ligated DnAs), for integrated management of DNA data at the restriction/functional site level. MATILDA creates restriction/functional maps of recombinant clones constructed from primary or recombinant DNA sources and permits automatic updating of all maps as the need arises.

For example, in our laboratory, DNA fragments from a relatively small


PRIMARY DATA BASE
WORKING DATA BASE
name.SEQ
Sequence Data
Sequence Analysis Program
name.ENZ
Restriction / Functional Sites
(User - friendly)
Manual Entry
Primary Restriction / Functional Site Data
ENZ→SIT
Program
primary files
name.SIT
(1 file per plasmid)
recombinant files
name.LIG
Ligation Specification
Manual Entry
LIGATE
Program
MAP & FRAGMENT
Program
Variable Format Restriction Maps & Fragments Lists

number of primary sources (e.g., pBR322, SV40, the c-src gene, etc.) have been combined to form a relatively large number of plasmids. While the separate cloning steps involved have been relatively simple, multistep cloning procedures (in which recombinant segments from previously constructed plasmids were used in the generation of new plasmids) have resulted in plasmids containing as many as 12 distinct DNA fragments. We use MATILDA to generate for each clone graphic maps and restriction fragment lists which display, in addition to restriction and functional site information, the original sources (e.g., pBR322, SV40, etc.) of the individual DNA segments within each clone. MATILDA requires an updated primary data base for the primary source DNAs and ligation specifications for the recombinants created in the lab. Interaction and output is user-friendly and creates maps and fragment lists containing arbitrary subsets of the available data in variable format.

The system has been developed in BASIC to run on an IBM PC.

SYSTEM STRUCTURE

The structure of the system is shown in Fig. 1. Restriction and functional site data from primary source DNAs are entered manually or by machine into ENZyme files. (MATILDA accepts negative locations and allows the location of the map coordinate origin to be shifted arbitrarily; see below.) An example of a simple ENZ file for a hypothetical primary DNA is shown in Fig. 2. (For the purpose of explication, simple example files are used in the paper. Actual files typically have hundreds of entries.) ENZ format is user-friendly, relatively free-form, and is compatible with machine-generated files from programs that search for restriction sites in sequence data files (such as SEQ [1] or the Conrad and Mount [2, 3] mini-

Fig. 1: Overall system structure. The interrelationship of files (in boxes) and functions (in ellipses) is displayed. Primary source DNA restriction/functional site data are entered either manually (for experimental restriction mapping and functional site data) or by sequence analysis programs (such as SEQ [1] or the Conrad and Mount [2, 3] programs) into ENZyme format. After automatic conversion to SITe format, the file becomes part of the working data base which, along with other SIT and LIGation specification files, is used by the LIGATE function to generate recombinant SIT files. These new SIT files in turn become part of the complete data base which is used to generate further recombinants. They are also used by the MAP and FRAGMENT functions to generate the desired restriction/functional site maps and fragment lists. The UPDATE function (not displayed) regenerates all SIT files whenever the information in the primary source DNA ENZ files is modified.

computer program). Different formats can be combined in one file to permit, for example, the addition of functional site information to a machine-generated restriction site file.

While ENZ format permits relatively unformatted manual data entry, it is not in a form suitable for efficient computation. For this purpose, ENZ files are converted (by MATILDA's ENZ→SIT function) to machine-friendly SITe files which are used as the data base for the other system functions. In practice, SIT files are generated and used only by MATILDA and are invisible to the user. The file, PRMRY.SIT, corresponding to the PRMRY.ENZ file in Fig. 2 is shown in Fig. 3a.

Once SIT files have been generated for the primary DNA sources, they are combined by the LIGATE function to generate SIT files for arbitrary recombinants. LIGATE uses a ligation specification (LIG) file for each clone

```
Sequence              Appears at position

AluI  (AGCT)           271   758   982  1378  1892  1962  2187  2231
                      2302

ApaI  (GGGCCC)         2258


EcoRI  2219,    1787
ALUI  15,20,          192, 325
MboI    29
bglii    29
     BstXI   428
XHOI  57
1975  HinfI
 523  HinfI
ZERO  57
HinfI  1002
LENGTH 2321
*srcATG   571
*srcTAG   1692
$IntronL  593
$IntronR    872
```

Fig. 2: ENZyme file format. The file PRMRY.ENZ for a hypothetical source DNA, PRMRY, is displayed. The format is relatively free-form to permit simple manual data entry and is also compatible with SEQ (1) and Conrad and Mount (2, 3) program output. Different formats can be mixed permitting simple manual addition to machine-generated files. (Excessively heterogeneous data are presented to demonstrate input flexibility.) Entries preceded by special characters (*, $, etc.) are used to designate functional markers which may be selected for printing according to their special character prefix. Parameters used for versatile output control (such as LENGTH, ZERO, and others not displayed in this example) are also embedded in the file.

A

```
(    )Name    :PRMRY
(    )Zero    : 42
(    )Length  : 2321
(    )Records : 32
(  33)AluI    :PRMRY.-42
(   5)AluI    :PRMRY.-37
(   9)MboI    :PRMRY.-28
(   0)BglII   :PRMRY.-28
(  28)XhoI    :PRMRY.0
(   0)ZERO    :PRMRY.0
( 135)AluI    :PRMRY.135
(  80)AluI    :PRMRY.215
(  53)AluI    :PRMRY.268
( 103)BstXI   :PRMRY.371
(  97)HinFI   :PRMRY.468
(  46)$srcATG :PRMRY.514
(  22)$IntronL:PRMRY.536
( 166)AluI    :PRMRY.702
( 113)$IntronR:PRMRY.815
( 111)AluI    :PRMRY.926
(  19)HinFI   :PRMRY.945
( 377)AluI    :PRMRY.1322
( 313)$srcTAG :PRMRY.1635
(  95)EcoRI   :PRMRY.1730
( 106)AluI    :PRMRY.1836
(  70)AluI    :PRMRY.1906
(  14)HinFI   :PRMRY.1920
( 211)AluI    :PRMRY.2131
(  31)EcoRI   :PRMRY.2162
(  13)AluI    :PRMRY.2175
(  28)ApaI    :PRMRY.2203
(  43)AluI    :PRMRY.2246
```

B

```
(    )Name    :pBR322
(    )Zero    : 0
(    )Length  : 4362
(    )Records : 21
( 418)EcoRI   :pBR322.0
(  31)HinDIII :pBR322.31
( 346)BamHI   :pBR322.377
(  89)MboII   :pBR322.466
( 186)SalI    :pBR322.652
(   0)AccI    :pBR322.652
( 281)BglI    :pBR322.933
(  40)NruI    :pBR322.973
( 194)BglI    :pBR322.1167
( 278)BalI    :pBR322.1445
( 802)AccI    :pBR322.2247
( 879)MboII   :pBR322.3126
( 107)AhaIII  :pBR322.3233
(  19)AhaIII  :pBR322.3252
( 234)BglI    :pBR322.3486
( 124)PstI    :pBR322.3610
( 334)AhaIII  :pBR322.3944
```

C

```
(    )Name    :pMAT1
(    )Zero    : 2162
(    )Length  : 6175
(    )Records : 43
( 418)EcoRI   :pBR322.0
(   $)EcoRI   :PRMRY.2162
(  31)AluI    :PRMRY.2131
( 211)HinFI   :PRMRY.1920
(  14)AluI    :PRMRY.1906
(  70)AluI    :PRMRY.1836
( 106)EcoRI   :PRMRY.1730
(  95)$srcTAG :PRMRY.1635
( 313)AluI    :PRMRY.1322
( 377)HinFI   :PRMRY.945
(  19)AluI    :PRMRY.926
( 111)$Intron :PRMRY.815
( 113)AluI    :PRMRY.702
( 166)$Intron :PRMRY.536
(  22)$srcATG :PRMRY.514
(  46)HinFI   :PRMRY.468
(  97)BstXI   :PRMRY.371
( 103)AluI    :PRMRY.268
(  53)AluI    :PRMRY.215
(  80)AluI    :PRMRY.135
( 135)ZERO    :PRMRY.0
(   0)XhoI    :PRMRY.0
(  28)MboI    :PRMRY.-28
(   0)BglII   :PRMRY.-28
(   $)BamHI   :pBR322.377
(  89)MboII   :pBR322.466
( 186)SalI    :pBR322.652
(   0)AccI    :pBR322.652
( 281)BglI    :pBR322.933
(  40)NruI    :pBR322.973
( 194)BglI    :pBR322.1167
( 278)BalI    :pBR322.1445
( 802)AccI    :pBR322.2247
( 879)MboII   :pBR322.3126
( 107)AhaIII  :pBR322.3233
(  19)AhaIII  :pBR322.3252
( 234)BglI    :pBR322.3486
( 124)PstI    :pBR322.3610
( 334)AhaIII  :pBR322.3944
```

Fig. 3: SITe file format. SIT files for primary source DNAs a) PRMRY and b) pBR322 are displayed along with the file for the recombinant c) pMAT1, generated by the LIGation specification in Fig. 4. (Only a subset of the restriction site is displayed for pBR322.) PRMRY.SIT and pBR322.SIT are output from the ENZ→SIT function while pMAT1.SIT is output from the LIGATE function.

The first four lines of the file contain the DNA name, its total length, the location of the coordinate origin (for map printouts), and the number of records in the file. Remaining entries contain the distance between adjacent restriction/functional sites, the restriction enzyme or functional marker name, and the "origin string" giving the name of and location in the primary source DNA from which the site originally derived. The "origin string" is trivial in files for primary source DNAs such as PRMRY and pBR322 but is informative in recombinants such as pMAT1. A ZERO parameter may also be included (see text).

These files are normally "invisible" to the user. The site locations used in the pMAT1 LIG specification were read from maps (as in Fig. 5) of PRMRY and pBR322.

and searches the data base to find the information necessary for the construction of the new recombinant SIT file. LIGate specifications are entered in a simple format (Fig. 4a) that tells which DNA fragments were combined to make the clone. Since most cloning steps involve simultaneous ligation of only 2 or 3 restriction fragments, ligate specifications are typically only 2 or 3 lines in length. Since previously generated recombinant plasmids can be used in the ligate specification (i.e., it is not necessary to refer back to the primary DNA sources in each specification), the specifications need never be complex. However, if desired, recombinants can be specified directly in terms of the primary source DNA fragments they contain (with resultant complexity of the LIG specification).

The interaction of LIG and SIT files through the LIGATE function is demonstrated in Figs. 3 and 4. The LIG specification of Fig. 4a was used by LIGATE to create the SIT file for a new recombinant pMAT1 (Fig. 3c) in which the EcoRI-BglII fragment from PRMRY (Fig. 3a) was ligated between the EcoRI and BamHI sites of pBR322 (Fig. 3b). A record of the original source of the DNA segments within pMAT1 is found in the "origin strings" of the new SIT file (see below). The SIT file for pMAT1 now becomes part of the working data base and can be used in the construction of subsequent recombinant SIT files.

A

```
pMAT1
PRMRY: EcoRI(2162) < BglII(-28)
PBR322: BamHI(377) > EcoRI(0)
```

B

```
pMAT2
PRMRY: EcoRI(2162) < [MboI(-28)] BglII(-28)
XbaI
ZERO
pBR322: BamHI(377) > EcoRI(0)
```

Fig. 4: LIGation file format. a) The LIGation specification to clone an EcoRI (location 2162)-BglII (location -28) fragment from PRMRY DNA between the EcoRI and BamHI sites of pBR322 is shown (BglII and BamHI are cross-ligated). The first line contains the name of the new recombinant, pMAT1. Subsequent lines contain descriptions of the DNA fragments used to construct the new clone. Coordinates following each enzyme refer to the locations printed in restriction maps (such as in Fig. 5) and determine which instance of multiple restriction site occurrences is to be used. The direction of the arrowhead determines which of the two possible fragments obtained from cutting a circular plasmid at the two specified sites is to be used.

b) The LIG file for plasmid pMAT2, a modification of pMAT1. The specification has been changed to reflect the addition of an XbaI linker at the BglII-BamHI junction and the fact that the recognition pattern for the MboI site originally at location -28 in PRMRY is not contained in the new construct (e.g., because of S1 digestion of the BamHI end for the linker addition reaction). In this example the map coordinate origin is fixed to coincide with the XbaI site by an explicit ZERO parameter.

SIT files also provide the data base for the actual production functions, MAP and FRAGMENT. The MAP function (e.g., see Fig. 5) generates restriction and functional site maps of any primary or recombinant DNA in the data base and permits arbitrary subsets of the data to be printed. Origin strings have been printed to display structural content of the recombinants. In practice restriction/functional site maps showing all the available data are so complex as to be useless. We routinely prepare 6-cutter and rare-cutter maps for all our recombinants and, on a day-to-day basis, prepare maps displaying only a few enzyme sites at a time for the design and/or analysis of individual experiments.

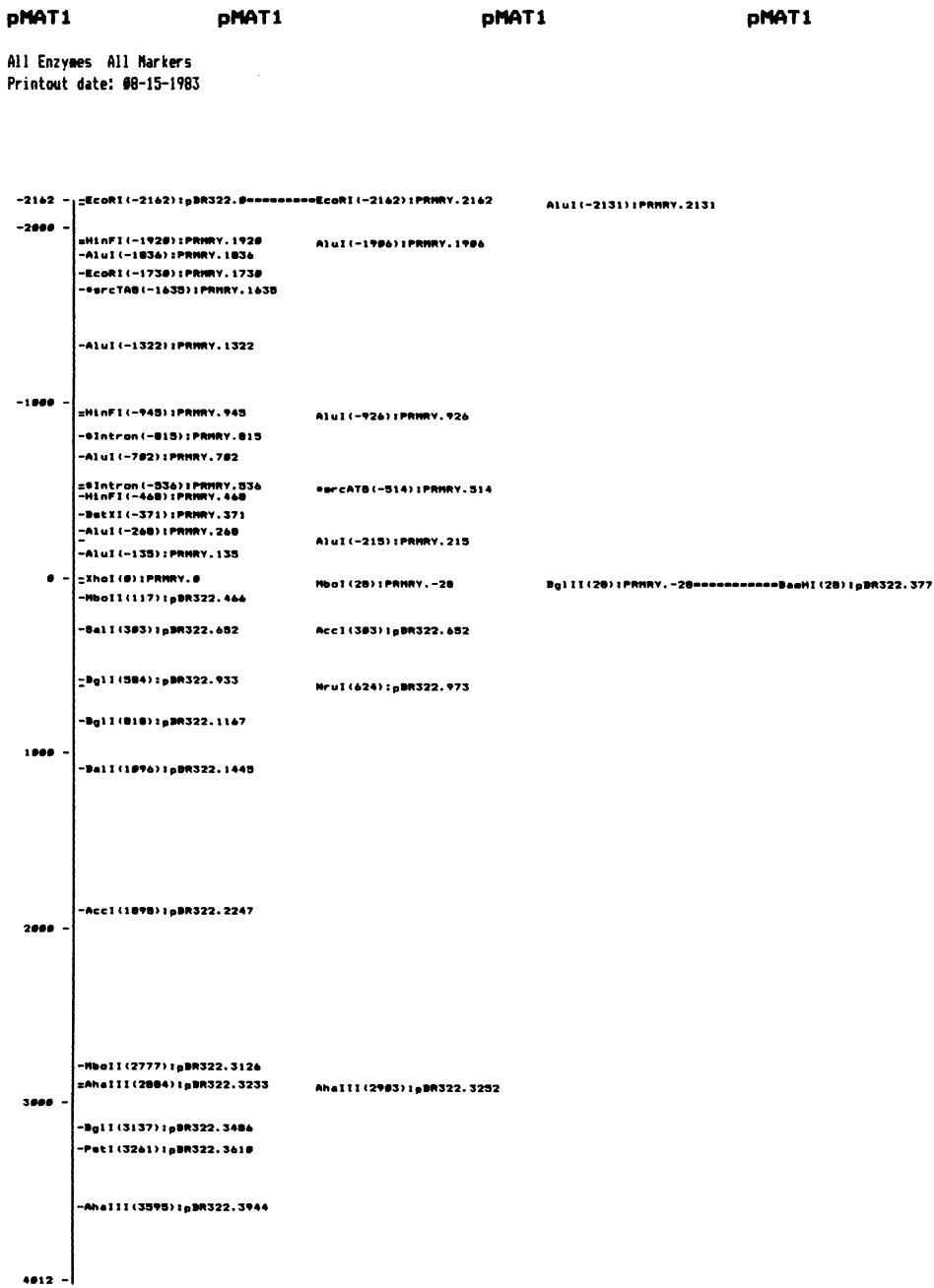
The FRAGMENT program generates lists of the restriction fragments generated by digestion of any DNA (or subfragment) in the DNA data base with an arbitrary set of enzymes (e.g., see Fig. 6). Functional site and origin data (describing the original primary source of the sequences within the fragment) are also included.

The UPDATE function provides simple overall data base maintenance whenever corrections to primary DNA data are made. This function regenerates all SIT files using the updated primary source ENZ and the recombinant clone LIG specification files. Modifications which propagate through a number of sequentially constructed clones are automatically handled by this routine. Since all the required LIG specifications have already been stored in the system data base, no additional input is necessary. It is also possible to update files individually using the ENZ→SIT and LIGATE functions.

All restriction enzyme names that enter the system (whether into ENZ files or as files or keyboard input for the operation of the MAP or FRAGMENT functions) are compared with an ENZymeLIST file to check for typographical or notational errors. Enzyme name notations which are not contained within the ENZLIST checkfile are not accepted. (Variations in capitalization are accepted for input; output capitalization is standardized. Re Figs. 2, 3a, 5 and 6). This facilitates uniform notation and reduces errors in a data base which may be generated from entries by numerous laboratory personnel.

SYSTEM FEATURES

MATILDA is designed to automatically manage most, if not all, of the structural data associated with recombinant clones in the laboratory and to allow, in simple fashion, the creation of graphic restriction maps with the flexibility associated with manual generation of such maps. System features that facilitate this are discussed below.


pMAT1
pMAT1
pMAT1
pMAT1
All Enzymes All Markers
Printout date: 08-15-1983
-2162
-2000
-1000
0
1000
2000
3000
4012
=EcoRI(-2162):pBR322.0
EcoRI(-2162):PRMRY.2162
AluI(-2131):PRMRY.2131
=HinFI(-1920):PRMRY.1920
AluI(-1906):PRMRY.1906
-AluI(-1836):PRMRY.1836
-EcoRI(-1730):PRMRY.1730
-SrcTA0(-1635):PRMRY.1635
-AluI(-1322):PRMRY.1322
=HinFI(-945):PRMRY.945
AluI(-926):PRMRY.926
-0Intron(-815):PRMRY.815
-AluI(-702):PRMRY.702
=0Intron(-536):PRMRY.536
SrcATB(-514):PRMRY.514
-HinFI(-460):PRMRY.460
-BstXI(-371):PRMRY.371
-AluI(-260):PRMRY.260
AluI(-215):PRMRY.215
-AluI(-135):PRMRY.135
=XhoI(0):PRMRY.0
MboI(20):PRMRY.-20
BglII(20):PRMRY.-20
BamHI(20):pBR322.377
-MboII(117):pBR322.464
-BalI(303):pBR322.682
AccI(303):pBR322.682
=BglI(584):pBR322.933
NruI(624):pBR322.973
-BglI(810):pBR322.1167
-BalI(1096):pBR322.1445
-AccI(1098):pBR322.2247
-MboII(2777):pBR322.3126
=AhaIII(2884):pBR322.3233
AhaIII(2903):pBR322.3252
-BglI(3137):pBR322.3486
-PstI(3261):pBR322.3618
-AhaIII(3595):pBR322.3944

Coordinate Systems

Primary data is input into ENZ files with restriction/functional site information given in base pairs. Negative coordinates are allowed. Left and right boundaries of linear DNA fragments can be arbitrarily specified by embedded parameters. "Wraparound" of data from circular DNAs is automatically handled. Restriction enzyme sites are defined as occurring at the center (or just to the left of center on the 5'→3' strand) of the recognition sequences. The ENZ→SIT function automatically recognizes output from the SEQ (1) or Conrad and Mount (2, 3) programs (which generate coordinates using the left ends of recognition sequences) and modifies it to correspond to the center notation using data from ENZLIST. These changes, while insignificant in terms of map coordinates and fragment sizes, are required for the correct representation of restriction sites at the ends of fragments.

The location of the coordinate origin in a printed restriction map need not be the same as that used in the ENZ file. Both ENZ and SIT files may contain a ZERO parameter which is used to shift the coordinate origin away from the left end (the coordinate origin default location) of the primary source map. ZERO parameters are passed on to all subsequently generated SIT files which permits the coordinate origins of a family of recombinants to be automatically set to a specified location within a sequence of interest. For example, in our laboratory, almost all clones contain the c-src gene. MATILDA permits us to automatically set the coordinate origin of each clone to the EcoRI site within this gene by placing a ZERO parameter in the c-src primary source ENZ file. All restriction/functional site locations used by the system (i.e., as printed by the MAP or FRAGMENT functions or as used by LIGATE) are relative to the specified coordinate origin. Alternatively, coordinate origins can be individually set for each recombinant by an explicit ZERO parameter in the LIG specification.

The pMAT1 and pMAT2 examples display the use of ZERO parameters. The

Fig. 5: MAP output. A restriction/functional site map of the recombinant plasmid pMAT1 is displayed. The map coordinate origin is set to coincide with the XhoI site by the ZERO specification found in the PRMRY.ENZ file (Fig. 2) which has been automatically passed down to the pMAT1.SIT file (Fig. 3c). The EcoRI-EcoRI and BglII-BamHI homologous and heterologous junctions used to construct pMAT1 are indicated by the concatenated markers at locations -2162 and 28. While all data are displayed in this simple example, display of the complete data set for actual recombinants is cumbersome and relatively useless. In that case it is more useful to display only subsets of the restriction/functional sites for either the entire recombinant DNA or for one of its subfragments.

pMAT1 **pMAT1** **pMAT1** **pMAT1**

Circular sequence: All
Cut with: AluI HinFI
Printout date: 08-15-1983

Size	Order	L.End		R.End		Origin DNAs	Functional Markers
4179	1	AluI	-135	AluI	-2131	pBR322 PRMRY	
514	5	AluI	-1836	AluI	-1322	PRMRY	*srcTAG
377	6	AluI	-1322	HinFI	-945	PRMRY	
234	9	AluI	-702	HinFI	-468	PRMRY	$Intron *srcATG
224	8	AluI	-926	AluI	-702	PRMRY	$Intron
211	2	AluI	-2131	HinFI	-1920	PRMRY	
200	10	HinFI	-468	AluI	-268	PRMRY	
80	12	AluI	-215	AluI	-135	PRMRY	
70	4	AluI	-1906	AluI	-1836	PRMRY	
53	11	AluI	-268	AluI	-215	PRMRY	
19	7	HinFI	-945	AluI	-926	PRMRY	
14	3	HinFI	-1920	AluI	-1906	PRMRY	

Fig. 6: FRAGMENT output. All fragments generated by AluI and HinfI digestion of pMAT1 are shown arranged in descending order of size. Each fragment is also described by 1) its order in the pMAT1 restriction map (numbering from left to right), 2) the restriction sites and locations of its left and right ends, 3) the primary source DNAs which it contains, and 4) the functional markers it contains.

entry "ZERO 57" in PRMRY.ENZ (Fig. 2) sets the coordinate origin to coincide with the XhoI site. The origin remains at the location of the XhoI site in the recombinant pMAT1.SIT file (Fig. 3c) and is used in the pMAT1 restriction map (Fig. 5) and fragment lists (Fig. 6). This automatic specification is overridden by an explicit ZERO in pMAT2.LIG (re Fig. 4b).

It may be noted that the SIT files use a difference notation in which the distances between adjacent restriction/functional site entries in the files are used instead of absolute locations relative to the origin location. This notation, which is convenient for computational purposes, is invisible to the user and is irrelevant to the discussion above.

Origin Strings

All restriction/functional sites are identified by an "origin string" that describes the primary source DNA from which they were first derived. Thus, the AccI site at location 1898 in the pMAT1 restriction map (Fig. 5) originally occurred in pBR322 at location 2247. Restriction/functional site origin strings are transferred from SIT file to SIT file as recombinants are generated (e.g., Fig. 3) so each site carries a "birth certificate" along with it. Display of this information by the MAP and FRAGMENT functions (e.g., Figs. 5, 6) mimics the standard procedure of notating the origins of the subfragments of recombinant clones.

Junction Modifications

The composite restriction site pattern in the immediate vicinity of a fragment junction depends on the specific ligation method used and on the particular enzyme recognition site patterns and cut sites involved. The multiplicity of techniques used in ligating fragments (e.g., cross-ligating preexisting sticky or blunt ends, filling-in or digesting overhanging sticky ends, polydeoxynucleotide tailing, etc.) precludes automatic adjustment of the junction region site map using only restriction site level data. The most common cause of alteration is cross-ligation between DNA fragment ends generated by different restriction enzymes. To handle this case, as well as to provide conspicuous markers for the locations of fragment junctions, concatenated restriction site markers are used in the printed restriction maps to represent junctions. The EcoRI-EcoRI and BglII-BamHI junctions in the pMAT1 map (Fig. 5) provide examples of this notation for the cases of homologous and heterologous junctions. The concatenated notation shows the user which sites were used in constructing the junction, from which he can determine which sites are still present. (For example, neither the BglII nor the BamHI site remains in pMAT1.)

Additional modifications may also be required. Accordingly, MATILDA provides convenient "editing" facilities in the junction region that allow the user to make any required modifications to the default recombinant SIT file (which includes all enzymes whose recognition pattern center is located within or on the boundary of the specified fragment). Fragment junctions are edited by including additional sites (e.g., because of linker insertion) or removing extraneous sites which are not actually present in the recombinant DNA because their recognition sequence is destroyed in the ligation procedure. Figure 4b presents examples of the notations which are used in the LIG file to indicate addition (XbaI) or deletion (MboI) of restriction sites. While restriction site additions can usually be determined directly from the experimental protocol, determining whether any sites have been deleted generally requires a detailed analysis of sequence data (if available) in the vicinity of the junction. To facilitate this process, MATILDA provides a listing of the suspect restriction sites and modifies the LIG file to permanently encode deletion of those restriction sites specified by the user.

Fragment junction sites at randomly generated locations (e.g., from random shearing or exonuclease digestion) are accommodated by inserting markers at the appropriate locations and using the markers in the LIG specifications.

MAP Output Control

In addition to creating and generating restriction maps of recombinant DNAs for reference, MATILDA can be used to generate output that is specifically tailored for the design and analysis of individual experiments. The user can specify which restriction/functional sites are to be displayed, either by keyboard entry and/or by reference to preset files (e.g., 6-cutters, cheap enzymes, etc.) and may request output of rare-cutters (those enzymes which cut the recombinant clone n times or less where n is specified by the user), fragment junctions, and/or subsets of functional markers. Either all or part of the recombinant DNA can be displayed, and the length (in pages) of the output map can be set as desired.

HARDWARE REQUIREMENTS AND AVAILABILITY

MATILDA has been developed on an IBM PC with 128 Kbyte memory and an IBM PC printer augmented with an Epson Graftrax 80 chip (an IBM PC Graphics printer without the Graftrax chip can also be used). Diskettes containing source code, compiled code, and user documentation are available.

ACKNOWLEDGEMENTS

We thank IBM Corporation for providing the IBM PC computing equipment used for development of the program system, S. Litwer for programming assistance, and S. Person for his support. This research was supported by PHS Grant 1R01CA32317-01 and an ACS Junior Faculty Research Award to D.S.

REFERENCES

1. Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L.H. (1982) Nucl. Acids Res. 10, 279-294.
2. Conrad, B. and Mount, D.W. (1982) Nucl. Acids Res. 10, 31-38.
3. We use a version of the Conrad and Mount programs adapted to run on the IBM PC by W. R. Pearson.