

Software for Automating Analysis of Encoded Combinatorial Libraries

William L. Fitch¹, Jing J. Zhang*¹, Nikhil Shah¹, Glenn I. Ouchi², Robert L. Wilgus³ and Steven Muskal⁴

¹Affymax Research Institute, 4001 Miranda Ave. Palo Alto, CA 94304, USA

²BREGO Research, San Jose, CA 95124, USA

³GlaxoSmithKline, 5 Moore Dr., Research Triangle Park, NC 27709, USA

⁴molSeek, LLC - 1802 N. Carson St, Carson City, NV 89701, USA

Abstract: This paper describes the applications which are used to automate the analysis of encoded combinatorial libraries. Commercial packages from MDL, Oracle and Agilent are linked with application software written in C/C++, in Microsoft Access and in ChemStation macro language. Encoding correspondence lists for each of up to three synthetic steps are conveniently associated with building block lists using the first application, CodeGen.

The second application Decode allows the user to identify the individual beads picked onto a 96-well plate and the pool number for each bead. The decoding chromatography data for each well is then loaded into the program. The chromatography data is used to identify the tags used in the synthesis. Along with the building block information from ISIS/Host, the building block used in each step of the synthesis can be identified. A third routine, Code-to-Structure, takes the coded library building blocks and creates the connection table in ISIS for each structure found by the decode program.

For quality control of encoded library synthesis, the decoded structures on a set of beads is compared to the LC/UV/MS data for the ligand cleaved from the same bead. CAPTURE, a GlaxoSmithKline proprietary application, is used to display and analyze the decoded structures and associated mass spectral data. This application uses simple isotopic composition and electrospray ionization rule sets to predict mass spectra and judge the concordance of a structure- mass spectrum data set. An ancillary program, EIC, is used to extract predicted single ion chromatograms from the full scan LC/MS data.

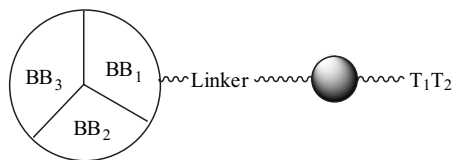
INTRODUCTION

Affymax encoded, split-pool combinatorial libraries are prepared with one compound per bead [1]. Such libraries combine the advantages of split-pool synthesis with those associated with the screening of discrete compounds, since individual beads (therefore single compounds) may be assayed for biological activity. However since the beads are randomized in each pooling operation, tracking the reagents and building blocks to which each bead is exposed is impossible, and the identity of the chemical ligand on an active bead is very hard to determine. Encoding [2] provides a general solution to this problem, whereby a surrogate analyte, or "tag" is attached to the beads which will allow

determination of the reaction history of a given bead and thus the identity of the final product which is attached to it. The Affymax construct is illustrated for a typical 3 step synthesis where 2 steps are encoded. The building blocks are labelled as BB_x and the codes as T_x . The linker is a cleavable chemistry which allows release of the compound for screening.

Our dialkylamine encoding strategy [3-4] has the critical property that the code is much easier to analyze than the associated compounds. We previously described the selection of an optimized set of dialkylamine tags [4]. Table 1 lists the dialkylamine tags and the abbreviations which we will use in this paper. Decoding entails hydrolytic release of the tags, determination of the tags, and thereby (with reference to a coding table) determining the structure of the ligand which was carried by the bead.

A key assumption for the utility of encoded libraries is that the code is predictive of the structure of the attached compound. Decoding must be unambiguous, and the correct compound must have been produced. Obtaining assurance of this presents significant analytical challenges [5-7] but it is critical to obtain, as poor quality libraries can lead to significant wasted time in the screening and re-synthesis of spurious hits.



*Address correspondence to this author at the Affymax Research Institute, 4001 Miranda Ave. Palo Alto CA 94304, USA; E-mail: Jim_Zhang@affymax.com

The analytical methods used for decoding have been greatly improved in recent years [4,8]. These improvements allow for much greater throughput. It is now practical to perform quality control sampling and analysis on a good proportion of the members of an encoded library. This analysis has the benefit of increasing confidence in the encoded library synthesis-and-screening process. The analytical work can have the negative effect of generating a large pile of paper data, the study of which can distract the organic chemist from the more fundamentally important duties of library design and evaluation of biological data. An information management system is needed to streamline this process. Many of these issues have been previously discussed by ourselves and by others [5,9-16]. We describe herein our customized software for automating the decoding of encoded combinatorial libraries.

Table 1. Tag Abbreviations

| Number | Amine Name | Abbreviation |
|--------|----------------------------------|--------------|
| 1 | 2,2,2-trideuteroethylpentylamine | depn |
| 2 | 2,2,2-trideuteroethylhexylamine | dehx |
| 3 | 2,2,2-trideuteroethylheptylamine | dehp |
| 4 | 2,2,2-trideuteroethyloctylamine | deoc |
| 5 | 2,2,2-trideuteroethylnonylamine | denn |
| 6 | 2,2,2-trideuteroethyldecylamine | dcdc |
| 7 | pentylolylamine | pnoc |
| 8 | diheptylamine | hphp |
| 9 | heptyloctylamine | hpoc |
| 10 | dioctylamine | ococ |
| 11 | pentylododecylamine | pndo |
| 12 | hexyldodecylamine | hxdo |
| 13 | heptyldodecylamine | hpdo |
| 14 | didecylamine | dcde |

Experimental

The CodeGen and Decode programs are applications written for Microsoft Access 97, using the programming language Visual Basic for Applications. Proper installation is required because the application uses three different MS Access databases and several Oracle database tables used by MDL (Molecular Design Ltd.) ISIS/Host and ISIS Base. To access data from the Oracle databases, Open DataBase Connectivity (ODBC) drivers are used. CAPTURE (Comprehensive Analysis Program That Ultimately Reveals Everything), is written in Microsoft Visual C++. CAPTURE retrieves plate maps created by CodeGen and the structures from Code-to-Structure using MDL ISIS/Object Library. Mass spectral data is read from files created by the Chemstation macros. Code to structure software is written in C/C++.

LC/MS for Code Reading

We have previously described the analytical methods which are used for decoding. These are LC/fluorescence or

(preferably) LC/MS. In each case the standard ChemStation report output of this analysis is a text file which lists the code abbreviations, the MS signal, the semi-quantitative information on the number of picomoles detected and the expected and measured retention times. Table 2 is an example LC/MS output file.

Table 2 LC/MS Decoding Report

| Tag | MS signal | Area | picomoles | Expected RT | Measured RT |
|------|-----------|---------|-----------|-------------|-------------|
| ococ | MSD1 242, | 0.000 | 0.000 | 1.277 | 0.000 |
| hpoc | MSD1 228, | 0.000 | 0.000 | 1.287 | 0.000 |
| dcde | MSD1 298, | 0.000 | 0.000 | 1.255 | 0.000 |
| hphp | MSD1 214, | 7.510e3 | 1.879 | 1.296 | 1.267 |
| hpdo | MSD1 284, | 0.000 | 0.000 | 1.277 | 0.000 |
| pndo | MSD1 256, | 0.000 | 0.000 | 1.305 | 0.000 |
| pnoc | MSD1 200, | 0.000 | 0.000 | 1.320 | 0.000 |
| hxdo | MSD1 270, | 0.000 | 0.000 | 1.289 | 0.000 |
| denn | MSD1 175, | 0.000 | 0.000 | 1.442 | 0.000 |
| deoc | MSD1 161, | 3.816e4 | 14.586 | 1.460 | 1.409 |
| dedo | MSD1 217, | 0.000 | 0.000 | 1.418 | 0.000 |
| dcdc | MSD1 189, | 0.000 | 0.000 | 1.431 | 0.000 |
| dehp | MSD1 147, | 0.000 | 0.000 | 1.488 | 0.000 |
| depn | MSD1 119, | 5.028e4 | 25.957 | 1.595 | 1.527 |
| dehx | MSD1 133, | 1.258e4 | 6.245 | 1.529 | 1.476 |
| debu | MSD1 105, | 0.000 | 0.000 | 1.695 | 0.000 |

Single bead LC/UV/MS for Library Quality Control

We have previously described the process of picking single beads from encoded library pools and assessing both the codes and the ligand on each bead [5]. This process has been automated through the use of a robotic beadpicker [17-18]. Experiments were conducted on an Agilent LC/MSD instrument controlled by ChemStation Rev. A07.01

Data Transfer to Server

MS acquisition computers should not be encumbered with complex data processing software. We do preliminary processing on the LC/MS computer but subsequent processing on desktop PC's and network servers. Post acquisition, all raw data and text reports are transferred to the decoding server [15].

RESULTS AND DISCUSSION

An Encoded Split Pool Synthesis

To illustrate the issues, a 6x6x6 tripeptide library synthesis will be simulated. We will use the same six amino acid building blocks at each position. Fig. (1) shows the

Table 3.

| Building block | | First position codes | | | Second position codes | | |
|----------------|-----------|----------------------|------|------|-----------------------|------|------|
| Number | structure | depn | dehx | dehp | deoc | denn | dcdc |
| 1 | ala | x | | | x | | |
| 2 | gly | | x | | | x | |
| 3 | glu | | | x | | | x |
| 4 | asp | x | x | | x | x | |
| 5 | phe | x | | x | x | | x |
| 6 | tyr | x | x | x | x | x | |

generic structure of the cleaved product as well as the structures and names of the building blocks. We will use 1, 2 or 3 tags to encode the first building block and 1,2 or 3 tags to encode the second. The third position will not be encoded; the final pools, each containing 36 tripeptides will be kept separate. Table 3 is the building block to tag encoding scheme.

Three tags could have encoded for up to 7 building blocks at each position. We never use more than 3 tags in a single code to maintain good tag detectability. And indeed because we have a total of 14 tags in our tagset, the preferable codes for this simple library would be to use 6 different single tags at each position. (We have not seen the need for introducing parity checking in encoding. In such a scheme only 2 or 4 tags would be used [19].) But we illustrate all of the different 1, 2 and 3 tag combinations in this scheme to explain all of the software issues. Of course the numbers of building blocks do not have to be the same in a library synthesis.

The process steps of encoded library synthesis are as follows (detailed descriptions can be found in the indicated references):

1. Differentiate the resin with BOC protected tag attachment point and cleavable linker [3].
2. Split the resin into 6 reaction vessels
3. Remove the tag protecting group and couple the first code as the Alloc- protected versions [3].
4. Couple building block 1 to the resin
5. Pool and resplit to 6 reaction vessels
6. Remove the Alloc group and couple the second code mixture as succinoyl derivatives [4].
7. Couple building block 2 to the resin

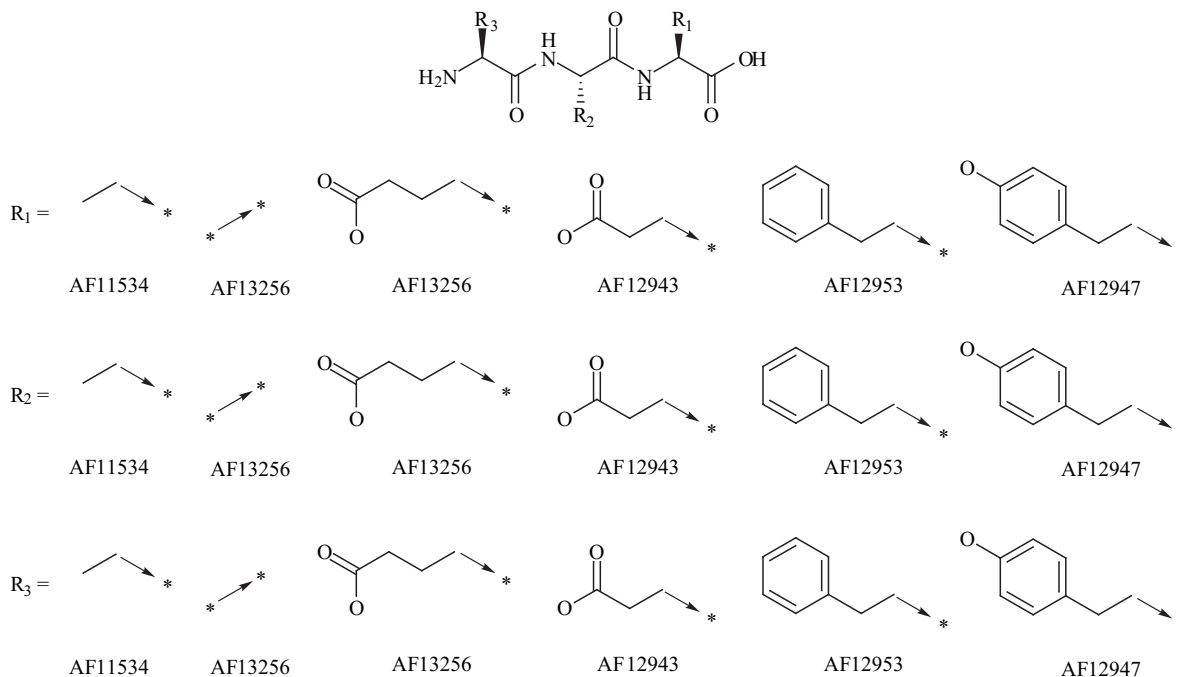
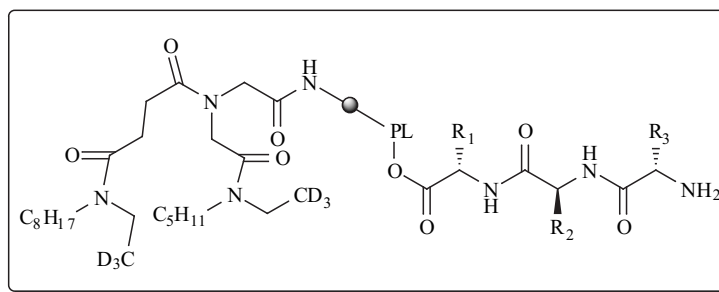


Fig. (1). Generic structure and building blocks for the virtual library.



8. Pool and resplit to 6 reaction vessels
9. Couple building block 3

Here is the structure of the first member of this library. The structure implies a 1:1 ratio of ligand and codes, but remember that we started with differentiated resin so the actual ratio is 10:1. PL is the photo-cleavable linker [20]. Cleavage will yield trialanine. Mineral acid cleavage of the tags will yield a mixture of depn and deoc.

DECODING AND LIBRARY QUALITY CONTROL: THE MANUAL METHOD

Prior to developing the Microsoft Access applications, decoding raw data was converted to structures manually. By reference to the encoding scheme, the detected tags were assigned to either R₁ or R₂ and then the best matches were made. For example if a bead from the 5th peptide pool

yielded codes depn, dehp and deoc, the first two tags would be assigned to R₁ and decoded as structure 5, phe, and the last tag assigned to R₂ and decoded as structure 1, ala. Combining the decoded information with the pool information gives the full structure as consisting of building blocks 5-1-5. For peptides these are easily put together and adequately described as “phe-ala-phe”. For more complex libraries it is necessary to use a chemical drawing program to assemble the chemical structure. Finally the molecular weight of the assembled compound can be calculated for comparison to the ligand LC/MS data.

CODEGEN

The first step in automating decoding is to get the tagging information (e.g. Table 3) into a database. CodeGen was written to generate lists of tag codes to identify each building block in a multi-step synthesis. Fig. (2) shows the user interface for CodeGen. For each library, the user enters

The screenshot shows the Affymax CodeGen user interface. It features three 'Show List' buttons for encoding building blocks 1, 2, and 3. Each button has a 'Number of Building Blocks' and a '# Tags' field. Below each button is a table of tag names and their corresponding tags.

| Tag Name | The Tag |
|----------|---------|
| depn | ✓ |
| dehp | ✓ |
| dehc | ✓ |
| deoc | ✓ |
| deod | ✓ |
| deoe | ✓ |
| deof | ✓ |
| deog | ✓ |
| deoh | ✓ |
| deoi | ✓ |
| deoj | ✓ |
| deok | ✓ |
| deol | ✓ |
| deom | ✓ |
| deon | ✓ |
| deoo | ✓ |
| deop | ✓ |
| deoq | ✓ |
| deor | ✓ |
| deos | ✓ |
| deot | ✓ |
| deou | ✓ |
| deov | ✓ |
| deow | ✓ |
| deox | ✓ |
| deoy | ✓ |
| deoz | ✓ |

Fig. (2). The user interface for CodeGen.

the type of encoding for each step in the synthesis, either full, partial or spatial.

Full Encoding means each building block is identified by one unique set of code tags.

Spatial Encoding means the pool number identifies the building block.

Partial Encoding is a special case where both codes and pool numbers are needed to identify a building block.

After selecting the type of encoding for each step, the user enters the number of building blocks for each synthetic step and clicks on the GO button, starting with step 1. The program computes the minimum number of tags required to encode the building blocks without using more than three tags per building block [4] or the user can select any higher number of tags by checking the boxes next to the tag names. The program will use all of the one-tag assignments first, then the two-tag and finally the three-tag combinations.

The main CodeGen algorithm is called for each synthesis step. All of the tag codes start with the number of tags used, 1,2 or 3. Then the tag codes are listed in alphabetical order separated by commas. For example, legal tags are:

- 1,ococ
- 1,dedc
- 2,dedc,hphp
- 3,dedc,hphp,ococ

The tag assignments are reviewed with the show list button. Once the user is satisfied with the tag lists an Excel spreadsheet formatted as table 1 can be obtained and the tag information is saved to a temporary file.

At this point, the chemist can go to the lab and start generating the combinatorial library. The building block structures are then entered into ISIS/Host in the standard way. These codes can then be loaded into the appropriate Oracle table in the TAG field. When the codes are loaded, the encoding type for each step (1=full, 2=partial or 3=spacial) along with the number of partially pooled building blocks are placed in a second Oracle table.

DECODE

Decode is used to identify the building blocks for each set of tag codes. The program has 3 parts, Beadsetup, Chromread and Decode.

Bead Setup

The first step in the decode process is to place beads on to a 96-well plate. The Beadsetup program allows the user to enter the bead and pool numbers in two ways for quality control (Regular QC) and hit decoding (Irregular) as shown in Fig. (3). For quality control, the number of pools to be placed on the plate(s) is entered along with the number of beads per pool (typically 4 or 5) and the starting pool number.

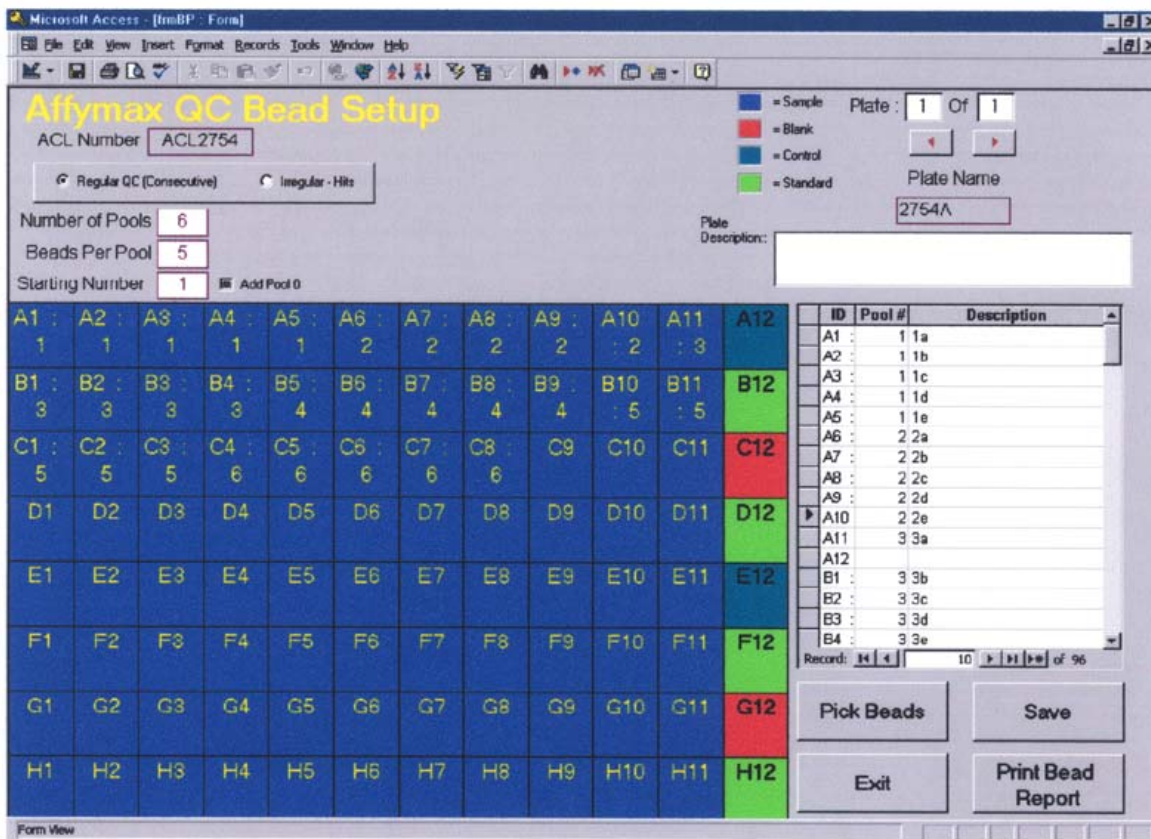


Fig. (3). The user interface for BeadSetup.

By default, the first eleven wells in each row are designated as sample wells. The wells in the 12th column are reserved to hold controls, blanks and standards. The Pick Beads button sets up the wells to each contain one bead for the correct pool. This information is stored with the plate and ultimately can be viewed in the CAPTURE program, so any additional information about a specific well can be entered at this time.

The Print Bead Report button can print the plate information. The Save button saves this plate information.

If Irregular Hits is selected, an entry box and button are displayed. The user enters the number of pools and then each pool number and the associated number of beads for that pool number. The plate is then setup and can be saved.

ChromRead

The decoding analytical method involves acid hydrolysis to release the covalently bonded amines, followed by their LC/MS detection and by reference to a calibration table, their quantitative determination. The LC/MS instrument software reduces the raw selected ion monitoring data into a report.txt file an example of which was shown as Table 2. ChromRead assigns this data to the appropriate well of the 96 well test format as shown in Fig. (4). This program has a list box to select the plate information. The browse button is used to select the files holding the chromatographic data for the first well. Standard Microsoft NT tools are used to navigate to the server where the LC/MS instrument has sent

the datafiles. Next, the number of files to load is entered and the first well number. The get data button populates the Access chromatography data table. Well colors change to reflect the status of each well. The chromatographic data can be viewed either by clicking on the well or as a list by using the View Table of Injections or View Table of peaks buttons. If the chromatographic data is correct, the Save button will save the data and its connection with the selected plate.

Decode

The primary screen of the main Decode application is shown as Fig. (5). When a plate number is selected, the software organizes the plate pool numbers from the Beadsetup step, the building block information from Oracle and the chromatographic results for each well. The left half of this screen shows the numbers of building blocks which were used for each library and the tags which were used in each step. The right side of the screen will change as each well is checked. The top half shows the 96 well graphical user interface(gui). The bottom half shows the chromatographic data for each well. A new column in the chromatographic data now also shows the R group (RG) with which the code is associated: in this case 1 for R1, 2 for R2 and X for the codes which were present in the chromatographic data but not used in the library synthesis. These last can only be present due to lab contamination or the presence of an interfering ligand-derived molecule in the bead hydrolyzate. Note that within each group (1,2 or x) the amines are sorted by the number in the amount column.

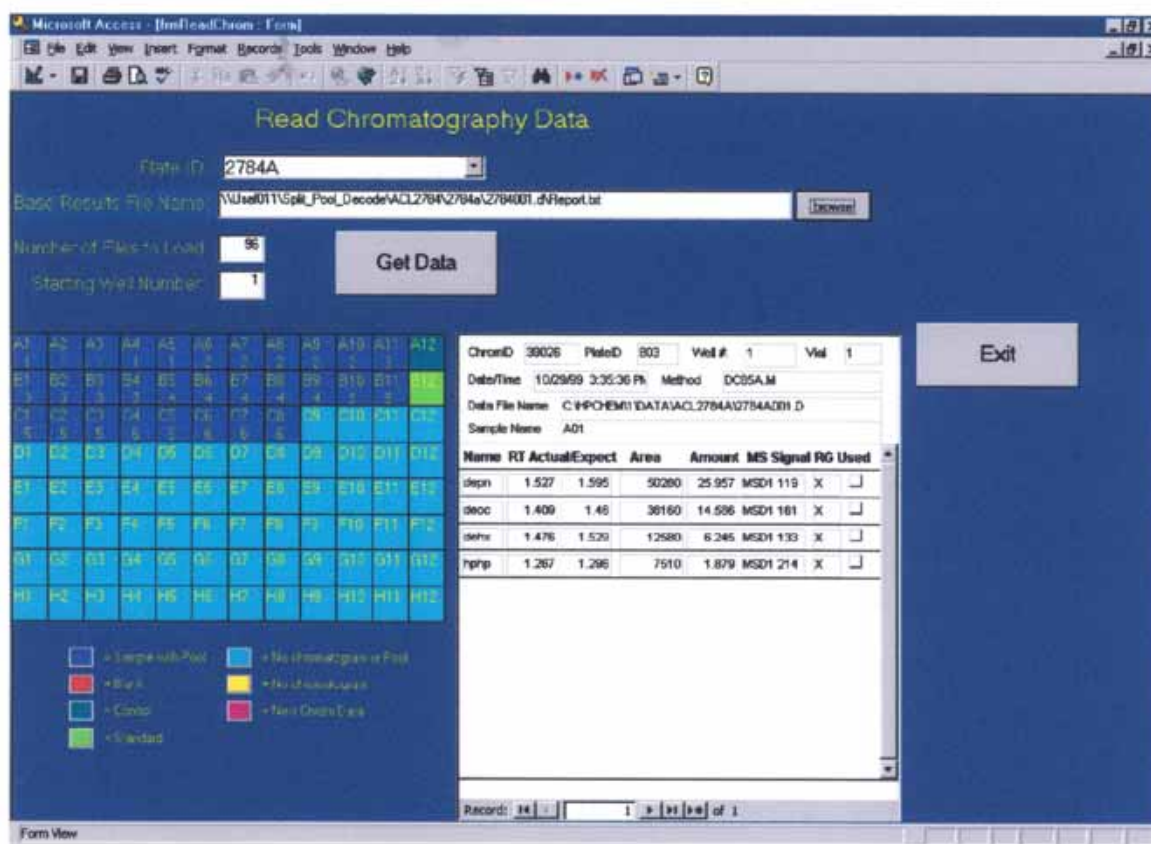


Fig. (4). The user interface for ChromRead.

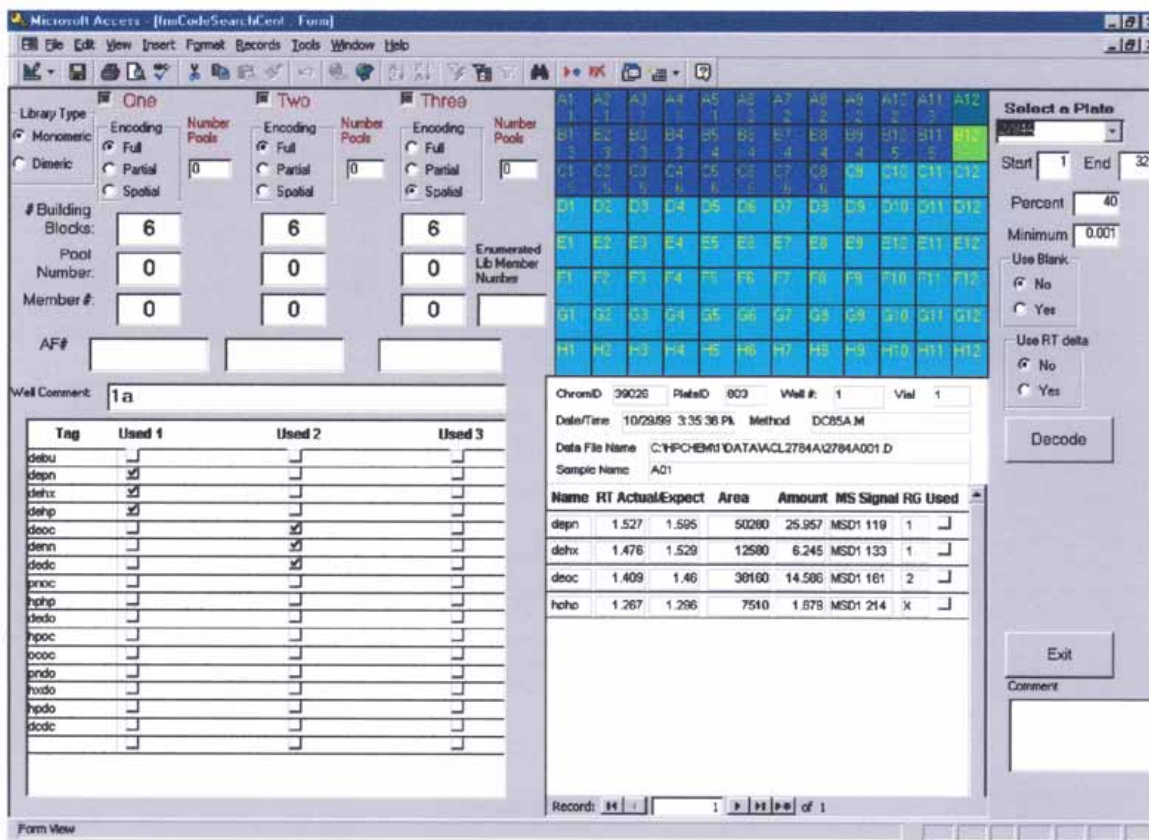


Fig. (5). The primary user interface for Decode. Debu and dedo are two obsolete tags which are including for backward compatibility.

By clicking on the 96 well diagram, the chromatographic data can be viewed. The analyst can now review the analytical quality of the data. For example viewing on the blank wells C12 or G12 will reveal contamination during analysis. Viewing the reference standard wells B12,D12,F12 and H12 will assure the analyst that adequate detection of all tags was maintained across the plate. And viewing the positive control wells A12 and E12 will assure the analyst that the plate was subjected to adequate acid hydrolysis to release the amines.

Two contamination correction tools are provided for the analyst, the blank and the minimum amount buttons. The chromatographic data from a blank can be subtracted from all of the wells by using the Yes radio-button in the Use Blank selection. A “well to use” entry box will appear, with the C12 well used as default. Any well can be entered as the blank well. The peak amounts for each tag in the blank well will be multiplied by 2 and subtracted from all of the same peaks in all other wells as the decode is performed. The minimum button provides another way to assure that very small peaks will not be used for decoding.

The LC/MS decoding chromatography is isocratic and low resolution. The specificity of the method is primarily due to the mass selection of the selected ions. The retention time window in the primary Agilent peak identification protocol is left relatively wide to allow for normal isocratic chromatographic drift across a set of analyses. In some cases interferences are seen in the data. A tighter retention time discrimination can be applied during the Access analysis of Decode by using the “Use RT delta” radio button. An

absolute retention time window is then entered and the Decode software will exclude identifications outside that window.

The user now can select a Percent value which is used to select chromatographic peaks to be used in the decode process. The default is 40%. Empirically this has been found to be a good compromise value to adjust for variations in encoding chemistry and decoding analysis. But more or less stringent percentages can be useful in special cases.

The decode algorithm is run with the Decode button. The key logic is that once the tags used for a specific synthesis step are known, one is simply required to determine if one, two or three of those tags are in the chromatography data. When the chromatography data for a well has been read, the tag peaks, which are used in synthesis step one, are sorted by amount and corrected for blanks or retention time. The first peak amount is then compared to the minimum amount. If the peak amount is greater than the minimum, then there is one tag. If there are no tags with amounts greater than the minimum, then the synthesis step has failed. The next peak is then checked to see if it is at least the Percent of the first peak. If it is, then there is a second code. Then the program looks for a third tag. If there is a fourth peak, which is higher than the Percentage, this is considered a failure for the synthesis step.

If one, two or three tags are found, then the tags are placed in alphabetical order and then the tag code is searched for in the Oracle R-Group table. If there is a match, then the AF number (AF is the Affymax compound numbering

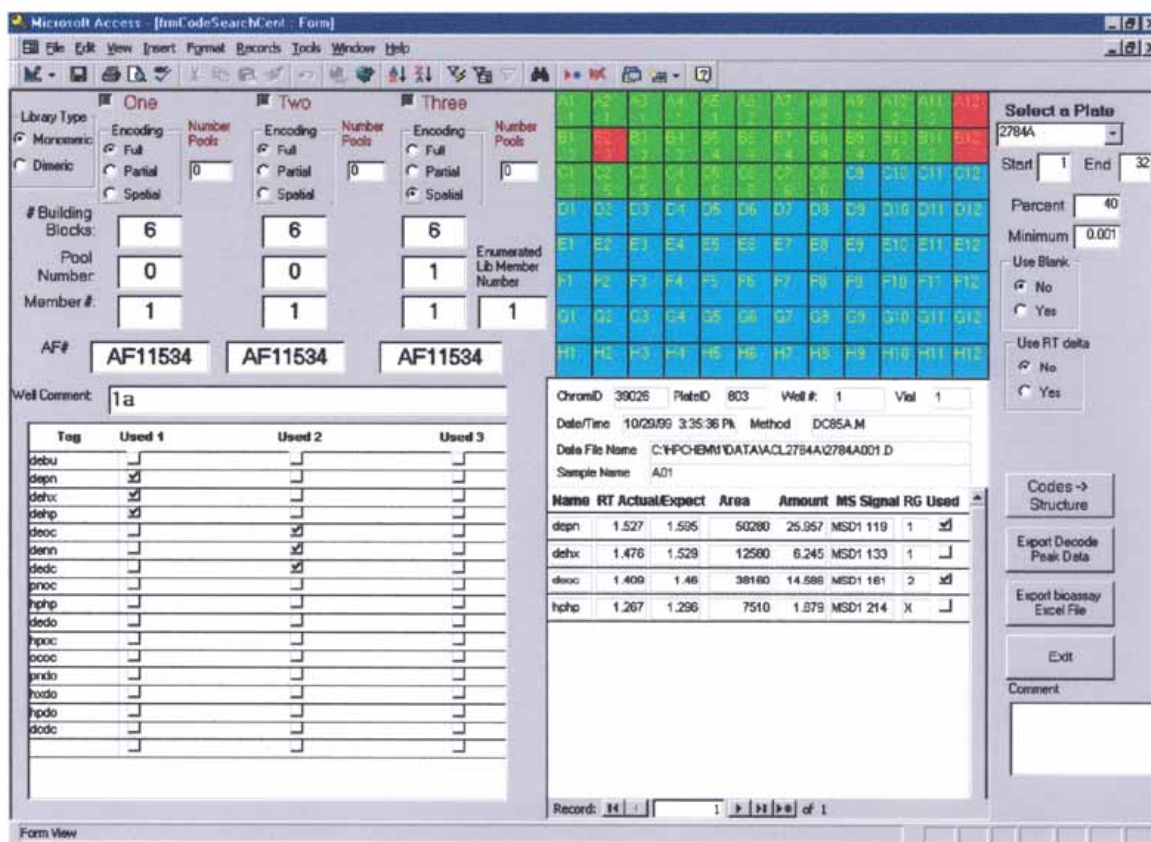


Fig. (6). The Decode user interface demonstrating the Code to Structure step.

system) is retrieved for the R-group along with the member number. These tags are then marked as used. The pool number is then used to retrieve a building for spatially or partially encoded steps. The screen is updated to show the decode information as shown in Fig. (6).

The well colors in the plate gui are adjusted to reflect success or failure of the decode. Obviously, the control wells will always come out red, as they do not represent coded molecules. At this point in the analysis, the user can review the data for completeness and study the failures. Failures are usually due to multiple beads or no beads in a well.

The Export Decode Peak Data feature exports to Microsoft Excel a worksheet which can be used to track quantitative results of the decode process. The worksheet includes building block identifiers and the amounts of tag detected in each well. These results are used in quality control of bead differentiation and encoding.

CODE TO STRUCTURE

Prior to structure generation, it is assumed that a generic structure describing the combinatorial library exists in the Affymax ACL database. The ACL database is an ISIS/Host v3.x RCG database that contains generic structures described with root structures and corresponding named building blocks (see Fig. (1)). Here, each building block has a previously assigned name (e.g. AF11534). After the decode is complete, the Code to Structure button appears as shown in Fig. (6). This button causes the enumeration string for

positive decodes (in this case ACL2784-AF11534-AF11534-AF11534) to be generated for each positive well and appends them to the AFFYBANK_ACL_DECODE table. An Oracle procedure is then launched which generates the structure from each enumeration string. Structures for the positive decodes are generated and placed in a ISIS database.

CAPTURE

We use the Windows NT application, CAPTURE [16], to correlate the decoded structures with the mass spectrum from each well in library quality control. For the example peptide library we would pick 5 beads from each of the 6 pools. Each bead yields a ligand LC/UV/MS datafile and a decoding LC/UV/MS datafile. The user interface in CAPTURE represents a 96 well microtiter plate, as shown in Fig. (7). For the green wells the predicted structure from decoding has matched a spectrum from the LC/MS data. This application uses simple isotopic composition [21] and electrospray ionization rule sets to predict mass spectra and judge the concordance of a structure- mass spectrum data set.

The white well is a case where no structure is available because the decoding failed. The red well is a case with a good decoded structure but the LC/MS data does not match.

The data "behind" the green light of well A01 is shown as Fig. (8). The ligand LC/UV/MS is conducted "full scan", treating the ligand as an unknown. In the figure, the top trace is a base peak chromatogram (BPC), the middle trace is the UV chromatogram and the bottom trace is the spectrum of the major or expected compound. The weakness of these

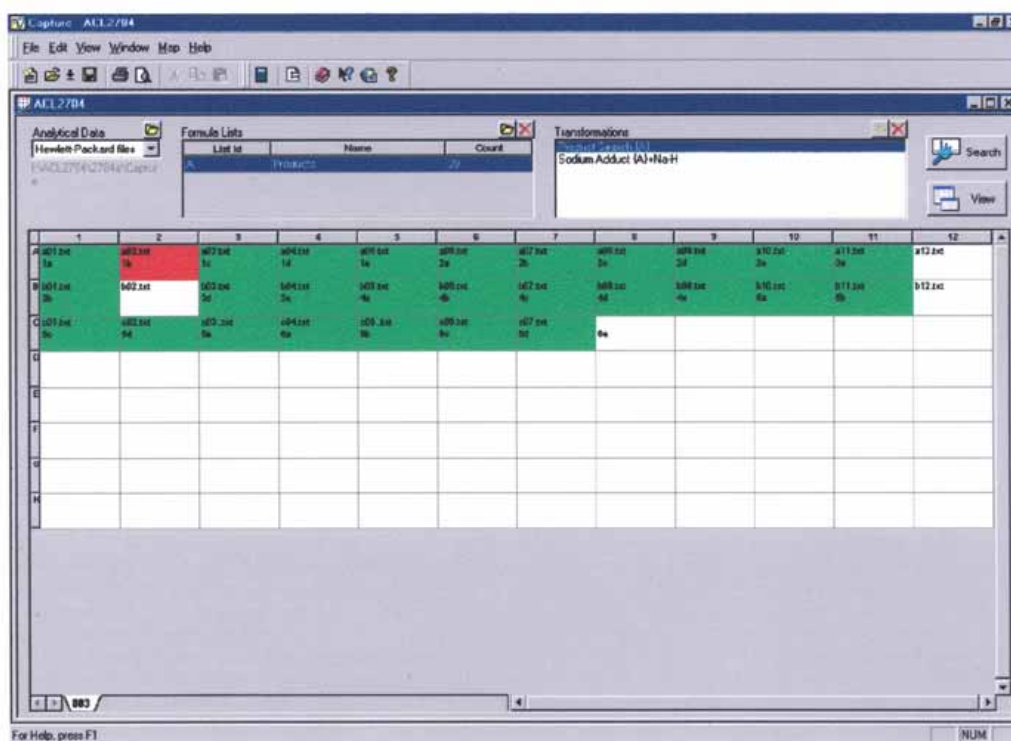


Fig. (7). The main CAPTURE user interface.

signals is typical for single bead mass spectra. Although the nominal amount of ligand cleavable from a single bead is in the hundreds of picomoles, variations in synthetic yield, cleavage yield and ionization efficiency can cause several orders of magnitude differences in actual LC/UV/MS peak

sizes. This simple peptide has insufficient UV absorbance to be seen at this concentration.

As seen in Fig. (8), three of the peaks in the BPC are labeled. These are the three peaks detected by the integration

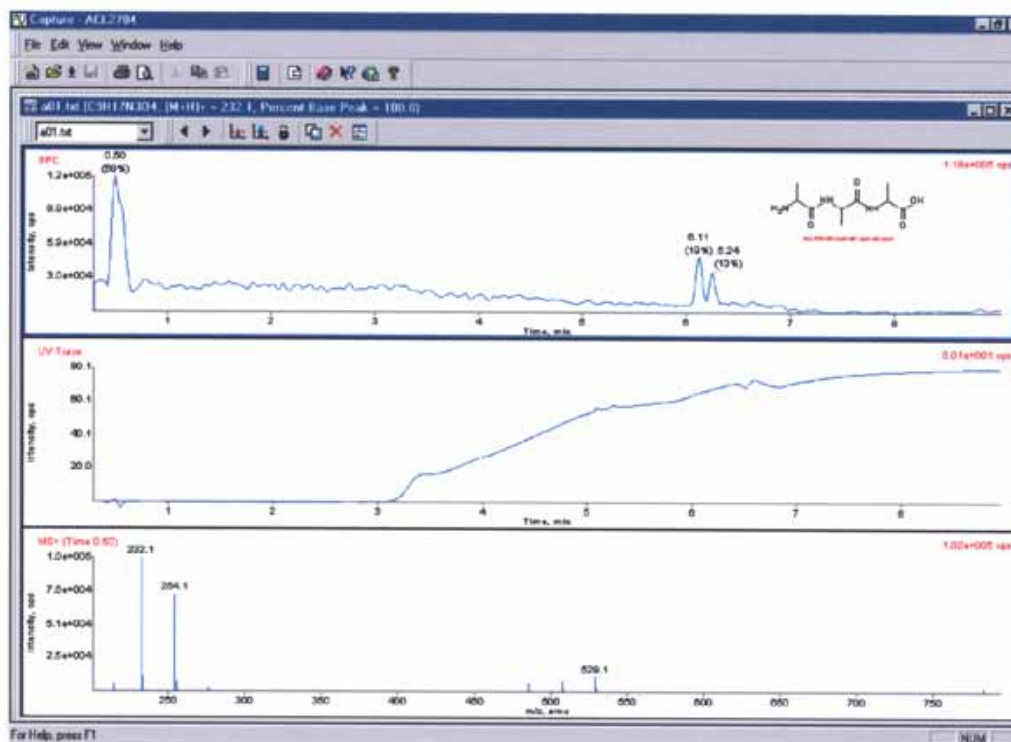


Fig. (8). The chromatogram/spectrum interface in CAPTURE.

parameters set in the Chemstation software. These parameters are optimized for each combinatorial library to minimize the total number of peaks but detect the desired peak in most cases. A slider bar in CAPTURE sets the criteria for what constitutes a green light, such as, predicted M+H must be >50% of the base peak in any of the detected spectra. In the Fig. (8) example, the predicted M+H is 100% of the base peak.

EIC FOR CAPTURE

Encoded library quality control using ligand single bead mass spectrometry is often complicated by a very weak signal observed in CAPTURE. Fig. (9) is an example. Both the UV trace (bottom) and the BPC (middle) show nothing above background; no spectra are available for green/red analysis. This situation arises due to weak UV chromophores, poor electrospray ionization, inadequate cleavage from the bead, a failure of synthesis, using badly contaminated solvents in sample preparation or inadequate sensitivity of the mass spectrometer. An additional data analysis tool in this case is to look for the presence of the predicted mass spectrum in the full scan data. The technique for doing this is called *extracting an ion chromatogram*. And the result is called an extracted ion chromatogram (EIC). For example, if the predicted tripeptide has a monoisotopic molecular weight of 231.1, the predicted positive ion electrospray mass spectrum of this compound would have a dominant ion at M+H=232.1. The three dimensional LC/UV/MS dataset is then queried for the presence of ion current at this mass and plot this current

against time. The top curve of Fig. (9) shows the EIC for this structure in this dataset. You can see that this focused data gives evidence for the presence of a molecule of the predicted properties. Just as important would be an EIC which showed no peak in the predicted EIC. This would provide strong evidence that the predicted molecule is not present.

To manually obtain the EIC for each well of a 96 well analysis can be painful. First the decoded structure is evaluated for molecular weight. Then the M+H is calculated. Then the raw LC/MS datafile is opened in the instrument control software, ChemStation. Finally an EIC is created for the predicted molecule. Doing this for 96 datafiles is prohibitive. A new piece of Access software was created to automate this task. The user interface is shown as Fig. (10). For each plate the user associates the raw datafile of the ligand LC/MS with each well. Then the appropriate type of EIC is chosen depending on the chemistry of the ligands. For positive ion mass spectrometry of small molecules, the default choice is to create M+H and M+Na (sodium). For negative ion mass spectrometry, the default is M-H. An option to look at doubly charged positive ions is offered as is an option to create a user-defined EIC.

The process is then initiated and each structure is queried and a set of target ions is created. ChemStation is then opened and a macro is run which opens each raw LC/MS datafile and extracts the predicted ion. The macro then creates new data transfer text file for each well and places them in a separate EICCAPTURE folder next to the existing CAPTURE text files. This EIC process is of critical

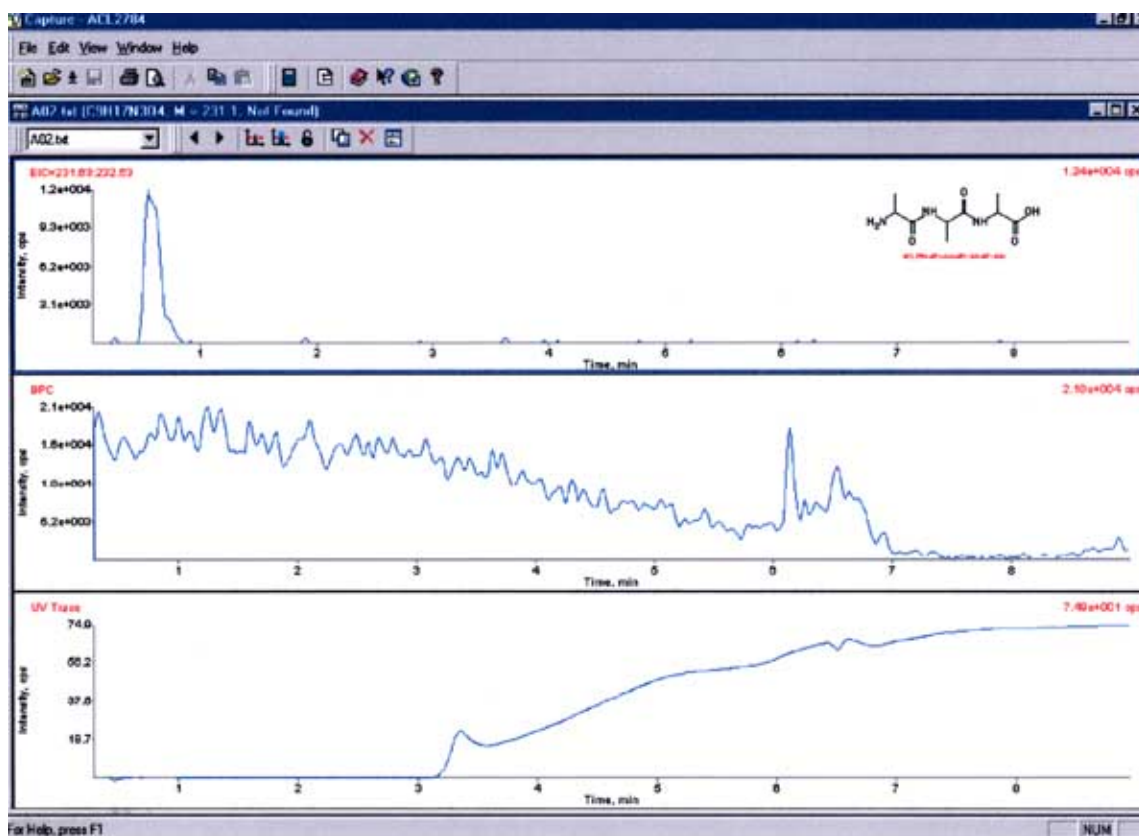


Fig. (9). The chromatogram/structure interface when using EIC for CAPTURE.

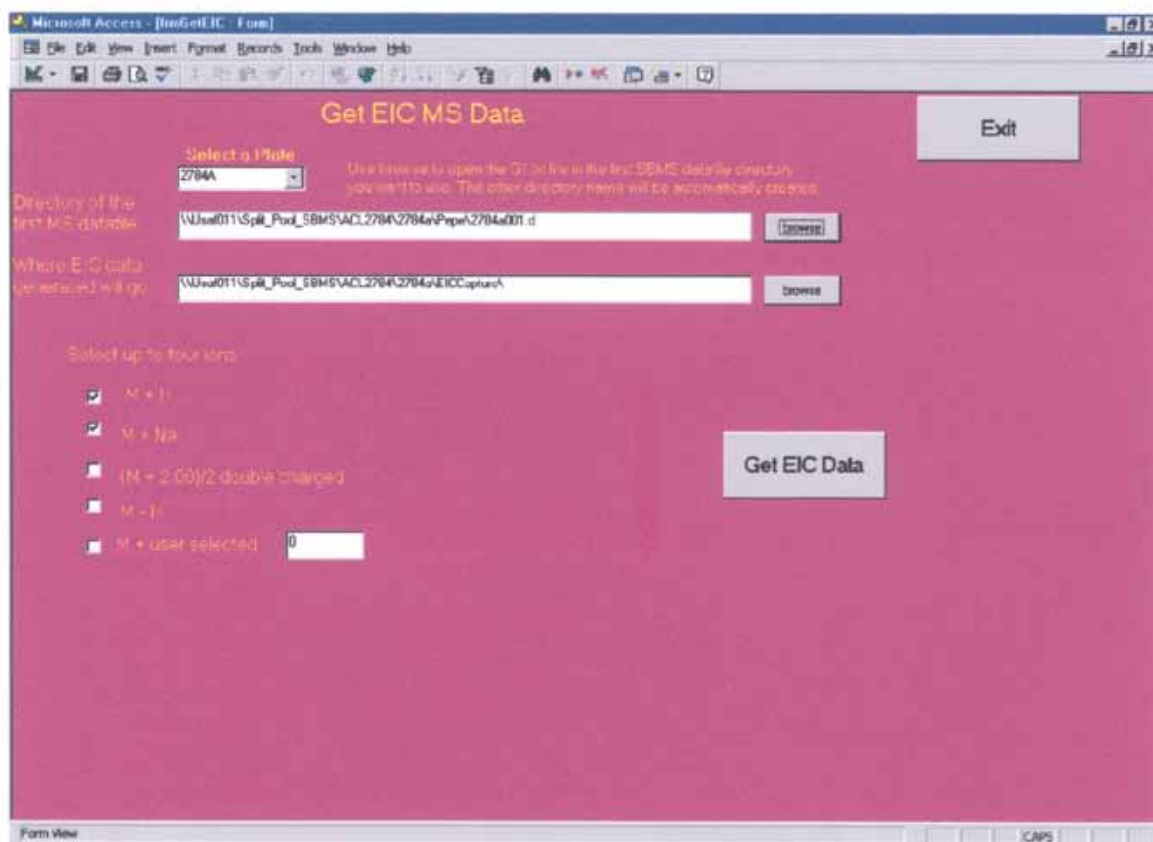


Fig. (10). The user interface for EIC CAPTURE.

importance in encoded library quality assurance [6] and has saved our chemists considerable effort previously expended doing these steps manually.

SUMMARY AND CONCLUSIONS

The new process for analyzing decoding and library quality control information is summarized in Fig. (11). The Access software resides only on the analytical chemist's computer. The structure database resides on the corporate Unix server. The LC/MS data is collected on dedicated data acquisition computers and transferred to NT servers for processing. CAPTURE and IsisBase reside on the end-user's desktops. Starting in the upper left corner, first the synthetic chemist plans and produces a library. The MDL software is used to enter the building blocks in an Oracle database. The Access application CodeGen is used here to set up the building block/code correspondence and load this information into the Oracle tables of the ISIS/Host ACL database.

When the chemist is done with the synthesis, samples are submitted to analytical for quality control. The analyst runs the Decode *Beadsetup* software to arrange single beads from each pool of the library into 96 well plates. Samples are analyzed for ligands using LC/UV/MS and these datafiles are stored on the Decode NT server along with the CAPTURE .txt files created by the Chemstation macro. The same samples are then decoded by LC/MS to generate datafiles and report.txt files which are organized onto the same NT server.

The analyst then starts the Decode *ChromRead* application to associate a decode.report.txt file with each well of each plate. The Decode main application is then run which converts the data for each well into an enumeration string ACLxxxx-Afaaaa-Afbbbb-Afcccc. The Code-To-Structure application then creates an ISIS database of these enumerated structures. As a final step the analyst uses the EICCAPTURE application in Access to generate the EICCAPTURE text files for use in CAPTURE.

The chemist is then informed that the analysis is complete and the data can be viewed. Each chemist has the CAPTURE application on his or her desktop. They first load the appropriate structures from the ISIS database of decoded structures onto their 96 well template for their quality control. Next, they find the LC/UV/MS text files from the server. They have a choice of looking at the data without EIC's (in the CAPTURE folder) or with EIC's (in the EICCAPTURE folder).

For the decoding of hit structures from biological screening, the process is simpler. The beads are delivered to the analytical laboratory with a 96 well hardcopy diagram which shows the library number and the pool number. The analyst uses the Beadsetup software to arrange the plate for analysis. After decoding LC/MS, the ChromRead, Decode and Code-to-Structure applications are used to create the structures. These are then communicated back to the originating scientist using the feature of CAPTURE which creates an Excel spreadsheet of the structures.

Encoded combinatorial chemistry is a young technology. The formats for applying this technology to drug discovery

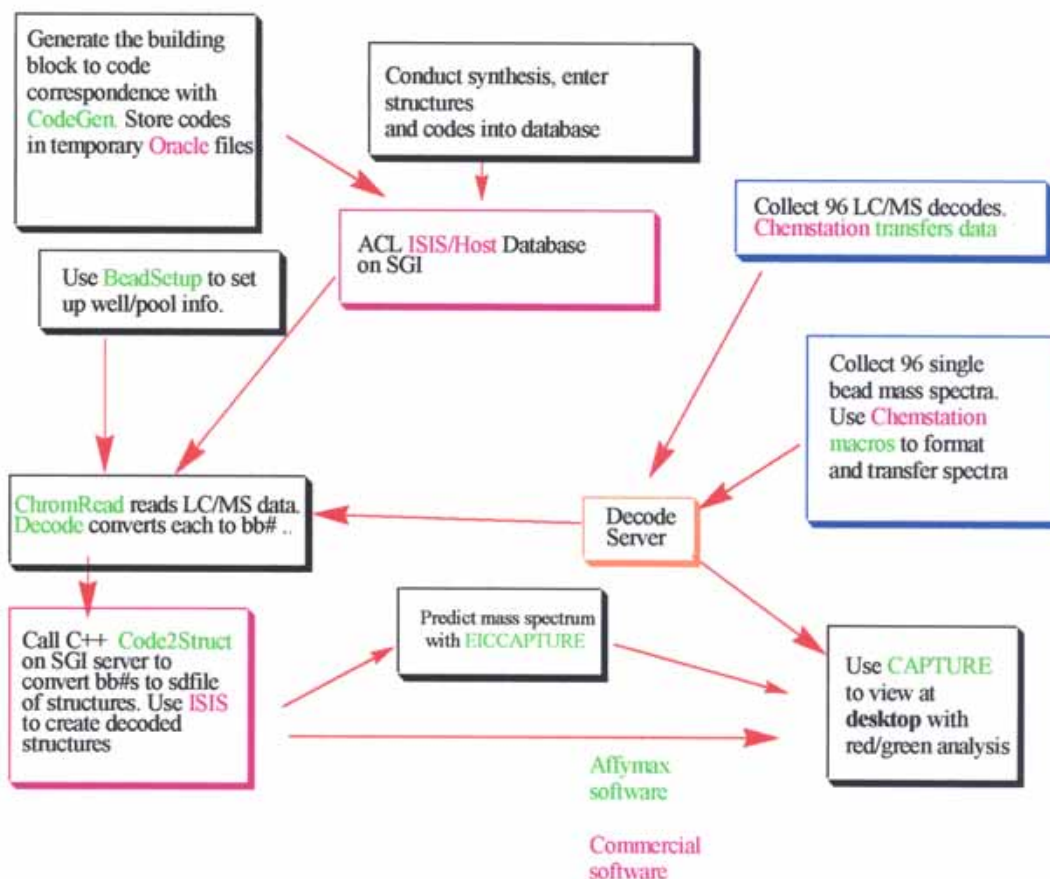


Fig. (11). Summary chart of software applications.

change on a regular basis; chemists can invent new formats much faster than software algorithms can be adapted and tested. We have found it valuable to use flexible Access software for application development in this area.

This software and its application are complex. It requires diligence to avoid errors in mismatching codes and structures or datafile names and beads. However the process has been shown to be a tremendous time saver over manual methods of decoding data handling. Only after this system was put in place, has Affymax been able to realize the full potential of encoded library chemical diversity.

REFERENCES

- [1] Lebl, M.; Krchnak, V.; Sepetov, N.F.; Seligmann, B.; Strop, P.; Felder, S.; Lam, K.S. *Biopolymers*, **1995**, *37*, 177-198.
- [2] Xiao, X.-Y. *Front. Biotechnol. Pharm.*, **2001**, *1*, 114-149.
- [3] Ni, Z.; Maclean, D.; Holmes, C.P.; Murphy, M.M.; Ruhland, B.; Jacobs, J.W.; Gordon, E.M.; Gallop, M.A. *J. Med. Chem.*, **1996**, *39*, 1601-1608.
- [4] Fitch, W. L.; Baer, T.A.; Chen, W.; Holden, F.; Holmes, C.P.; Maclean, D.; Shah, N.; Sullivan, E.; Tang, M.; Waybourn, P.; Fischer, S.M.; Miller, C.A.; Snyder, L.R. *J. of Comb. Chem.*, **1999**, *1*, 188-194.
- [5] Lewis, K. C.; Fitch, W.L.; Maclean, D. *LC/GC*, **1997**, *16*, 644-649.
- [6] Dolle, R. E.; Guo, J.; O'Brien, L.; Jin, Y.; Piznik, M.; Bowman, K.J.; Li, W.; Egan, W.J.; Cavallaro, C.L.; Roughton, A.L.; Zhao, Q.; Reader, J.C.; Orłowski, M.; Jacob-Samuel, B.; Carroll, C. *J. Combinatorial Chem.*, **2000**, *2*, 716-731.
- [7] Nicolaou, K. C.; Pfefferkorn, J.A.; Mitchell, H.J.; Roecker, A.J.; Barluenga, S.; Cao, G.-Q.; Affleck, R.L.; Lillig, J.E. *J. Amer. Chem. Soc.*, **2000**, *122*, 9954-9967.
- [8] Lane, S. J.; Pipe, A. *Rapid Commun. Mass Spectrom.*, **2000**, *14*, 782-793.
- [9] Bauer, B. In *Molecular diversity and combinatorial chemistry: libraries and drug discovery*; Chaiken, I. M.; Janda, K. D. Eds.; American Chemical Society, Washington, D.C., **1996**, pp 233-243.
- [10] Daley, D. J.; Scammell, R.D.; James, D.; Monks, I.; Raso, R.; Ashcroft, A.E.; Hudson, A.J. *Am. Biotechnol. Lab.*, **1997**, *15*, 24-28.
- [11] Ouchi, G. I. *LC/GC*, **1998**, *16*, 362-367.

- [12] Richmond, R.; Gorlach, E.; Seifert, J. *J. Chromatogr. A*, **1999**, 835, 29-39.
- [13] Tong, H.; Bell, D.; Tabei, K.; Siegel, M.M. *J. Am. Soc. Mass Spectrom.*, **1999**, 10, 1174-1187.
- [14] Koch, K. *LC/GC*, **2000**, 18, 500-512.
- [15] Shah, N.; Teeter, S.; Fitch, W.L.; Wilgus, R.L.; Koch, C. *Scientific Computing and Instrumentation*, August **2001**.
- [16] Shah, N.; Tsutsui, K.; Lu, A.; Davis, J.; Scheuerman, R.; Fitch, W.L. *J. Combi. Chem.*, **2000**, 2, 453-460.
- [17] Gavin, R. M.; Sugarman, J. H.; Kedar, H.; Chan, S. Systems and methods for arraying beads. PCT Int. Appl. **1997**, 45 pp.
- [18] Karet, G. *Drug Discovery and Development*, **1999**, 55.
- [19] Dillard, L. W.; Connelly, J.A.; Baldwin, J.J.; Horlbeck, E.G.; Kirk, G.L.; Lauri, G. Selecting codes to be used for encoded combinatorial libraries. PCT Int. Appl. **2000**, 37 pp.
- [20] Holmes, C. P.; Jones D.G. *J. Org. Chem.*, **1995**, 60, 2318-2319.
- [21] Kubinyi, H. *Anal. Chim. Acta*, **1991**, 247, 107.

Received: 5 December, 2001

Accepted: 21 December, 2001