



October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

---

# **User's Manual**

**for**

# **PatentIn 3.2**

---

Approved by:

---

Program Manager

Date

Prepared by:

Computer Sciences Corporation  
2611 Jefferson Davis Highway, Suite 10,000  
Arlington, VA 22202-4016

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

<b>Document Identification</b>	
User's Manual for PatentIn 3.2	PTO-OP-07
USPTO Contract 50-PBPT-7-00004	System Development and Maintenance
Task Title	PatentIn Maintenance
Task Type	Discrete
Task Order Number	CSC-04-04

### Version Control Log

<b>Version Control Number</b>	<b>Date Change Effective</b>	<b>Nature of Change</b>	<b>Pages Affected</b>	<b>Signature Authority</b>
N/A	11/10/98	Initial document prepared by Lockheed Martin Corporation	All	J. Paccassi
Draft 1.0	06/03/99	Initial document prepared by CSC	All	J. Paccassi
Version 1.0	09/16/99	Incorporate USPTO comments and updates from FQT	i, ii, 1v, 3-5, 3-6, 4-2, 4-3, 4-5, 4-6, 5-2, 5-3, 5-5, 6-2, 6-4, 6-7, 7-2, 7-3, 8-3 through 8-9	J. Paccassi
Version 1.1	06/02/00	Update document to reflect change from JAVA to C++	All	J. Paccassi
Version 1.2	06/22/00	Incorporate USPTO updates from FQT	3-5, 3-10, 5-4, 5-10, 6-2, 8-4	J. Paccassi
Version 2.0	01/26/01	PatentIn 3.1 release	All	J. Paccassi
Version 2.1	04/27/01	Updates to PatentIn 3.1 release	All	J. Paccassi
Version 2.2	10/16/03	Updates to PatentIn 3.2 release	All	J. Paccassi

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

## **EXECUTIVE SUMMARY**

This User's Manual does not conform to the Data Item Description for the User's Manual, PTO-OP-07. Instead, as required by Section 2.4.4.4 of the WPI Task Management Plan (TM02), it is being updated using the format in which the document was originally prepared.

This User's Manual is being updated at PTO's direction under the task order number CSC-04-04 to include changes for PatentIn 3.2.

## TABLE OF CONTENTS

<b>SECTION 1 INTRODUCTION.....</b>	<b>1-1</b>
1.1 PURPOSE .....	1-2
1.2 CONVENTIONS.....	1-2
1.3 OVERVIEW .....	1-2
<b>SECTION 2 SYSTEM REQUIREMENTS AND PATENTIN 3.2 ACCESS.....</b>	<b>2-1</b>
2.1 SYSTEM REQUIREMENTS .....	2-2
2.2 PATENTIN ACCESS .....	2-2
<b>SECTION 3 GETTING STARTED .....</b>	<b>3-1</b>
3.1 SEQUENCE SCREEN .....	3-2
3.2 PROJECT MENU .....	3-3
3.3 CREATING AND SAVING A NEW PROJECT .....	3-4
3.4 OPENING A PROJECT .....	3-5
3.5 SAVING A PROJECT.....	3-6
3.6 VIEWING A WORK FILE.....	3-7
3.7 VIEWING WORK IN PROGRESS .....	3-7
3.8 VIEW A SEQUENCE LISTING .....	3-8
3.9 VIEWING ERROR REPORTS .....	3-9
3.10 RENAME SEQUENCE.....	3-10
3.11 EXIT PATENTIN.....	3-10
3.12 HOW TO USE ONLINE HELP.....	3-10
3.13 MESSAGE DIALOG.....	3-11
<b>SECTION 4 PROJECT AND APPLICANT DATA.....</b>	<b>4-1</b>
4.1 APPLICATION STEPS MENU .....	4-2
4.2 PROJECT DATA.....	4-3
4.3 PRIOR APPLICATION INFORMATION .....	4-3
4.4 APPLICANT DATA .....	4-5
4.4.1 Individual Applicants.....	4-6
4.4.2 Organization Applicants .....	4-7
<b>SECTION 5 SEQUENCE DATA .....</b>	<b>5-1</b>
5.1 SEQUENCE.....	5-2
5.1.1 Selecting a Standard Organism.....	5-3
5.1.2 Creating default explanations for "n"s and "Xaa"s .....	5-4
5.1.3 Searching for a Sequence.....	5-4
5.1.4 Clearing the Screen.....	5-4
5.2 ADDING A SEQUENCE.....	5-5
5.3 IMPORTING A SEQUENCE.....	5-6
5.3.1 Format for Multi-Sequence Data Files to be Imported by PatentIn 3.x .....	5-7
5.3.1.1 Sequence Header.....	5-7
5.3.1.2 Sequence Data.....	5-8
5.3.2 Format for Single Sequence Data Files to be Imported by PatentIn 3.2 .....	5-8
5.3.3 Importing Sequences from a Project.....	5-10
5.3.4 Importing a 2.1 Project .....	5-11
5.4 COPYING A SEQUENCE .....	5-12
5.5 PASTING A SEQUENCE .....	5-13
5.6 DELETING A SEQUENCE .....	5-13
5.7 SKIP A SEQUENCE .....	5-13

## TABLE OF CONTENTS (CONTINUED)

5.8	RESTORING A SEQUENCE .....	5-14
5.9	REORDERING SEQUENCES .....	5-15
5.10	VALIDATING SEQUENCES .....	5-16
5.11	SAVING SEQUENCES.....	5-16
5.12	RELOADING A SAVED PROJECT .....	5-16
5.13	ADDING CUSTOM CODONS.....	5-16
5.14	ADDING A CUSTOM ORGANISM.....	5-18
5.15	ARTIFICIAL SEQUENCE OR UNKNOWN ORGANISM .....	5-20
<b>SECTION 6 FEATURE DATA .....</b>		<b>6-1</b>
6.1	SEQUENCE FEATURES .....	6-2
6.1.1	Feature Key Selection.....	6-4
6.1.2	Additional Information Required for Modified_Base.....	6-5
6.1.3	Additional Information on CDS.....	6-5
6.1.4	Further Definition of “n” or “Xaa” .....	6-6
6.1.5	Selecting an Amino Acid.....	6-6
6.1.6	Additional Information for MOD_RES .....	6-7
<b>SECTION 7 PUBLICATION DATA .....</b>		<b>7-1</b>
7.1	PUBLICATION TYPE SCREEN .....	7-2
7.2	JOURNALS PUBLICATION INFORMATION .....	7-3
7.3	DATABASES PUBLICATION INFORMATION.....	7-5
7.4	PATENTS PUBLICATIONS INFORMATION.....	7-6
7.5	THESES PUBLICATION INFORMATION .....	7-8
<b>SECTION 8 CREATING A SEQUENCE LISTING PROJECT FILE .....</b>		<b>8-1</b>
8.1	SEQUENCE LISTING FILE .....	8-2
8.2	GENERATING A SEQUENCE LISTING FILE .....	8-2
8.3	VIEWING A SEQUENCE LISTING FILE.....	8-4
8.4	COPYING THE SEQUENCE LISTING TO A DISK .....	8-5
<b>Appendix A LIST OF ACRONYMS.....</b>		<b>A-1</b>
<b>Appendix B FIELD IDENTIFIERS, LENGTHS AND TYPES .....</b>		<b>B-1</b>
<b>Appendix C COUNTRY CODES.....</b>		<b>C-1</b>
<b>Appendix D CONVERSION TABLE BETWEEN NUCLEOTIDE TRIPLETS (CODONS) AND ONE- AND THREE-LETTER AMINO ACID CODES .....</b>		<b>D-1</b>
<b>Appendix E NUCLEOTIDE SEQUENCE FEATURES.....</b>		<b>E-1</b>
<b>Appendix F AMINO ACID SEQUENCE FEATURES.....</b>		<b>F-1</b>
<b>Appendix G DATA TABLES FOR MOD_RES SEQUENCE FEATURES .....</b>		<b>G-1</b>
<b>Appendix H ADDITIONAL LIPID SEQUENCE FEATURES .....</b>		<b>H-1</b>

**TABLE OF CONTENTS (CONTINUED)**

**Appendix I ACCEPTABLE CHARACTERS IN THE SEQUENCE DESCRIPTION**  
**FIELD ..... I-1**

**Appendix J ADDITIONAL MODIFIED\_BASE SEQUENCE FEATURES ..... J-1**

**Appendix K Technical Notes..... K-1**

    K.1 MICROSOFT® ACCESS NOTES..... K-2

    K.2 GENERAL HINTS AND TIPS ..... K-3

    K.3 INSTALLATION AND TESTING NOTES ..... K-3

## LIST OF TABLES

Table 1-1: Document Conventions .....	1-2
Table 5-1: Header for Sequence .....	5-7
Table B-1: Field Name, Identifier, Size, and Type.....	B-2
Table C-1: Country Codes .....	C-2
Table D-1: Exchange Nucleotide Characters and Amino Characters.....	D-2
Table E-1: Nucleotide Sequence Features .....	E-2
Table F-1: Amino Acid Sequence Features .....	F-2
Table G-1: First Data Table for MOD_RES Sequence Features.....	G-2
Table G-2: Second Data Table for MOD_RES Sequences .....	G-2
Table H-1: Additional Lipid Sequence Features .....	H-2
Table I-1: Acceptable Characters in the Sequence Description Field .....	I-2
Table J-1: Modified_base Sequence Features.....	J-2

## LIST OF FIGURES

Figure 3-1: Sequence Screen .....	3-2
Figure 3-2: Project Menu .....	3-3
Figure 3-3: Save As Screen.....	3-4
Figure 3-4: Open Screen .....	3-5
Figure 3-5: Save As Screen.....	3-6
Figure 3-6: View Work in Progress Window .....	3-7
Figure 3-7: View Sequence Listing Window.....	3-8
Figure 3-8: View Error Report Window.....	3-9
Figure 3-9: Rename Sequence Screen .....	3-10
Figure 3-10: Exit PatentIn Screen.....	3-10
Figure 3-11: Help Screen .....	3-11
Figure 3-12: Message Dialog Screen .....	3-11
Figure 4-1: Application Steps Menu.....	4-2
Figure 4-2: Project Data Screen.....	4-3
Figure 4-3: Prior Application Information Screen.....	4-4
Figure 4-4: Applicant Data Screen .....	4-5
Figure 4-5: Individual Applicants Screen.....	4-6
Figure 4-6: Organization Applicants Screen.....	4-8
Figure 5-1: Sequence Screen .....	5-2
Figure 5-2: Selecting an Organism Screen .....	5-3
Figure 5-3: Add A Sequence Screen.....	5-5
Figure 5-4: Import Sequence(s) Screen .....	5-6
Figure 5-5: Sample ASCII Sequence Data .....	5-8
Figure 5-6: Sequence Type Selection Screen .....	5-8
Figure 5-7: Sequences Being Imported Screen.....	5-9
Figure 5-8: Validation Errors Screen.....	5-10
Figure 5-9: Select Sequences From Project Screen .....	5-11
Figure 5-10: Browse Window for 2.1 Project Import.....	5-12
Figure 5-11: PatentIn 2.1 Projects List .....	5-12

Figure 5-12: Edit Menu.....	5-13
Figure 5-13: Sequence Recovery Screen .....	5-14
Figure 5-14: Reorder Sequences Screen.....	5-15
Figure 5-15: Custom Codons Input Screen.....	5-17
Figure 5-16: Amino Acid Drop Down List Screen.....	5-18
Figure 5-17: Custom Organism Input Screen .....	5-19
Figure 5-18: Artificial Sequence/Unknown Organism Comment .....	5-20
Figure 6-1: Features Screen .....	6-2
Figure 6-2: Feature Names/Key Selection Screen .....	6-4
Figure 6-3: Feature Names/Key Selection Screen with Modified Base .....	6-5
Figure 6-4: Feature Names/Key Selection Screen with LIPID Selected .....	6-6
Figure 6-5: Feature Names/Key Selection Screen with MOD_RES Selected.....	6-7
Figure 6-6: First MOD_RES Pull Down List .....	6-8
Figure 6-7: Second MOD_RES Pull Down List.....	6-9
Figure 7-1: Publication Type Screen .....	7-2
Figure 7-2: Journals Publication Information Screen .....	7-3
Figure 7-3: Databases Publication Information Screen .....	7-5
Figure 7-4: Patents Publication Information Screen.....	7-6
Figure 7-5: Theses Publications Information Screen.....	7-8
Figure 8-1: Sequence Generation Screen.....	8-2
Figure 8-2: Second Sequence Generation Screen.....	8-3
Figure 8-3: View Results Window .....	8-4
Figure 8-4: Copy to Disk Screen .....	8-5

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

## **SECTION 1 INTRODUCTION**

## SECTION 1 INTRODUCTION

### 1.1 PURPOSE

PatentIn facilitates the creation of sequence listings for inclusion in patent applications sequences. It accepts data about the sequences, validates the data, creates a sequence listing file and a mechanism for printing out and saving to removable medium for submission. This manual describes how to use PatentIn.

### 1.2 CONVENTIONS

Consistent visual cues and standard keyboard operations are used throughout this manual. These conventions are listed in Table 1-1 below.

**Table 1-1: Document Conventions**

Notation	Represents	Example
	Process the command with a mouse action.	 Select Open from the file menu.
	Process the command with the keyboard.	 Enter changes into the Edit field.
<b>Bold type</b>	Name of a function, file, menu item, or programming construct.	 Click on <b>Exit</b> .
	An informative message to the user.	 NOTE: These fields are mandatory.

### 1.3 OVERVIEW

PatentIn is a computer program designed to expedite the preparation of United States Patent and Trademark Office (USPTO) patent applications containing nucleic acid and polypeptide sequences.

PatentIn complies with all format requirements specified in World Intellectual Property Organization Standard (WIPO) ST.25 and the related U.S. final rule, "Requirements for Patent Applications Containing Nucleotide Sequence and/or Amino Acid Disclosures." It runs on Windows 95, Windows 98, Windows NT, Windows 2000 and Windows XP. Screen displays are in English. Since the sequence listings generated by PatentIn are in compliance with ST.25, this program has worldwide applicability.

For ease of use, the design follows standard Windows user interface conventions.

PatentIn includes the following tools:

- **A Sequence Editor**

The primary tool within PatentIn is the sequence editor, which enables you to enter and modify both nucleic acid and protein sequence listings, as well as import sequence listing files created by another editor or word processor (provided they are stored as ASCII text files).

When working in PatentIn you may enter data in any order, and also add, remove, or revise sequence listing data at any time. You may also save a partially completed project and finish it at a later time. PatentIn does not require that a project exist or remain on a particular machine or device. Users are free to e-mail files to clients or each other so that they might review/update them.

- **A Sequence Generator**

After you have entered all the data necessary for your patent application, PatentIn enables you to generate your application. The application consists of a computer-readable, ST.25 compliant file containing a sequence listing file.

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 2**  
**SYSTEM REQUIREMENTS AND PATENTIN 3.2 ACCESS**

## SECTION 2 SYSTEM REQUIREMENTS AND PATENTIN 3.2 ACCESS

### 2.1 SYSTEM REQUIREMENTS

PatentIn is a self-contained application that can be downloaded from the United States Patent and Trademark Office (USPTO) web site. It operates in a Windows 95/98/NT/2000/XP environment. A minimum of 64 MB of memory is recommended. Additional memory may be required for large patent applications. Very large projects, projects with 100,000 sequences or a sequence approaching 1M, will need a minimum of 128 MB of memory. The disk space required to install PatentIn 3.2 is 2.5 MB. Additional disk space is required to store project files and sequence listing files. Microsoft® Access, with some of the extensions, is required to use the PatentIn 2.1 project load feature. (The load of the sequence data files does not have this restriction.)

For PatentIn to work correctly, the “TMP” environment variable must point to a valid directory, and the “PATH” environment variable must include the DOS backup command in the path. Most Windows installations will meet these requirements.

 SPECIAL NOTE FOR USERS WITH VERY LARGE SEQUENCES AND LARGE NUMBERS OF SEQUENCES: PTO has located a viewer that works for very large text files. A 60 day evaluation version is downloadable at [www.fileviewer.com](http://www.fileviewer.com) <<http://www.fileviewer.com>> . The viewer is named "V" and version 2000 SR-1. PTO tested it with 60MB and 120MB files and it “worked great.” It should manage files of just about any size. PTO tested it on a laptop with Win98. (LocalAdmin may be required for installation.)

### 2.2 PATENTIN ACCESS

PatentIn was designed to be installed on individual computers; it can be downloaded onto your personal computer and, if desired, the project files stored remotely. The program can be downloaded from the USPTO web page. The Universal Resource Locator (URL) from which the PatentIn 3.2 application can be obtained is:

<http://www.uspto.gov/web/offices/pac/patin/patentin32rel.htm>. Follow the instructions found on the Web page to download PatentIn 3.2 and install it on your PC. Upon completing the installation, an icon will be placed on your desktop. Access to the PatentIn 3.2 application program occurs when you double-click on the PatentIn 3.2 icon.

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

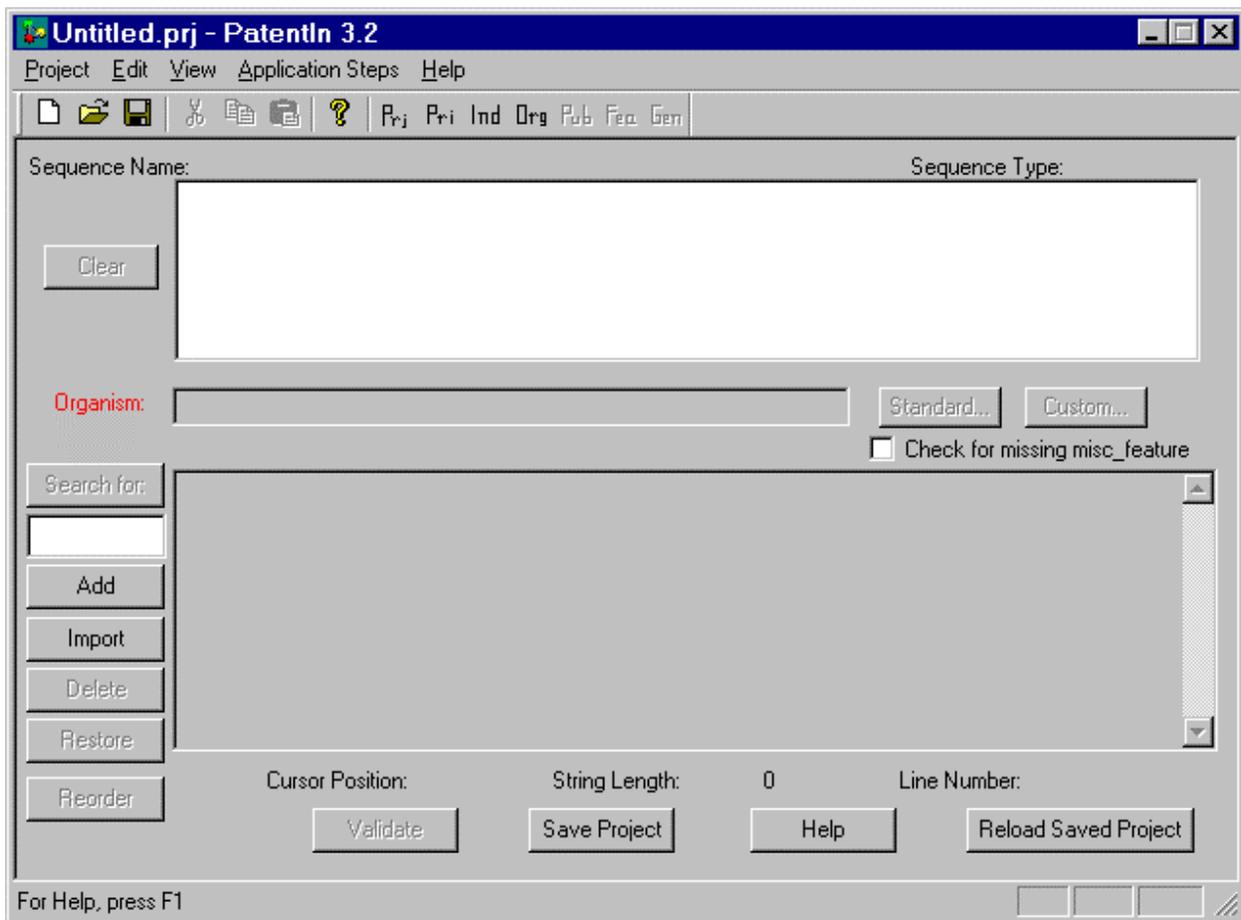
**SECTION 3**  
**GETTING STARTED**

## SECTION 3 GETTING STARTED

### 3.1 SEQUENCE SCREEN

When you first access PatentIn by double-clicking on the PatentIn 3.2 icon on your desktop, you have immediate access to the Sequence Screen.

The Sequence Screen (Figure 3-1) is the main screen.



**Figure 3-1: Sequence Screen**

The Sequence Screen (Figure 3-1) provides the user with five drop-down menus, three of which provide access to the real-time system interface. They are Project, Application Steps, and Help. The remaining two drop-down menus, Edit and View, are general Microsoft® Windows-type menus. The user may select any one of the three drop-down menus when a project is begun. PatentIn presents an empty project upon startup entitled “Untitled.” The user can open an existing project with the Project Menu (Figure 3-2).

### 3.2 PROJECT MENU

The Project Menu (Figure 3-2) enables you to create and save a project. Selecting “Save” displays the Save As Screen (Figure 3-3) where a new project is given a name and is saved. Selecting “Open” displays the Open Screen (Figure 3-4) where the user can select a previously saved project to open. The “Exit PatentIn 3.2” selection closes the application. Menu items that require a project to be opened, or an output file to be present, are grayed out until those conditions are met.

The Project Menu (Figure 3-2) selections are shown below. The user will see a list of selectable Menu Items under the Project menu and the active project name on the upper left-hand corner of the screen. In this case, Untitled is shown as no project has yet been opened or saved.

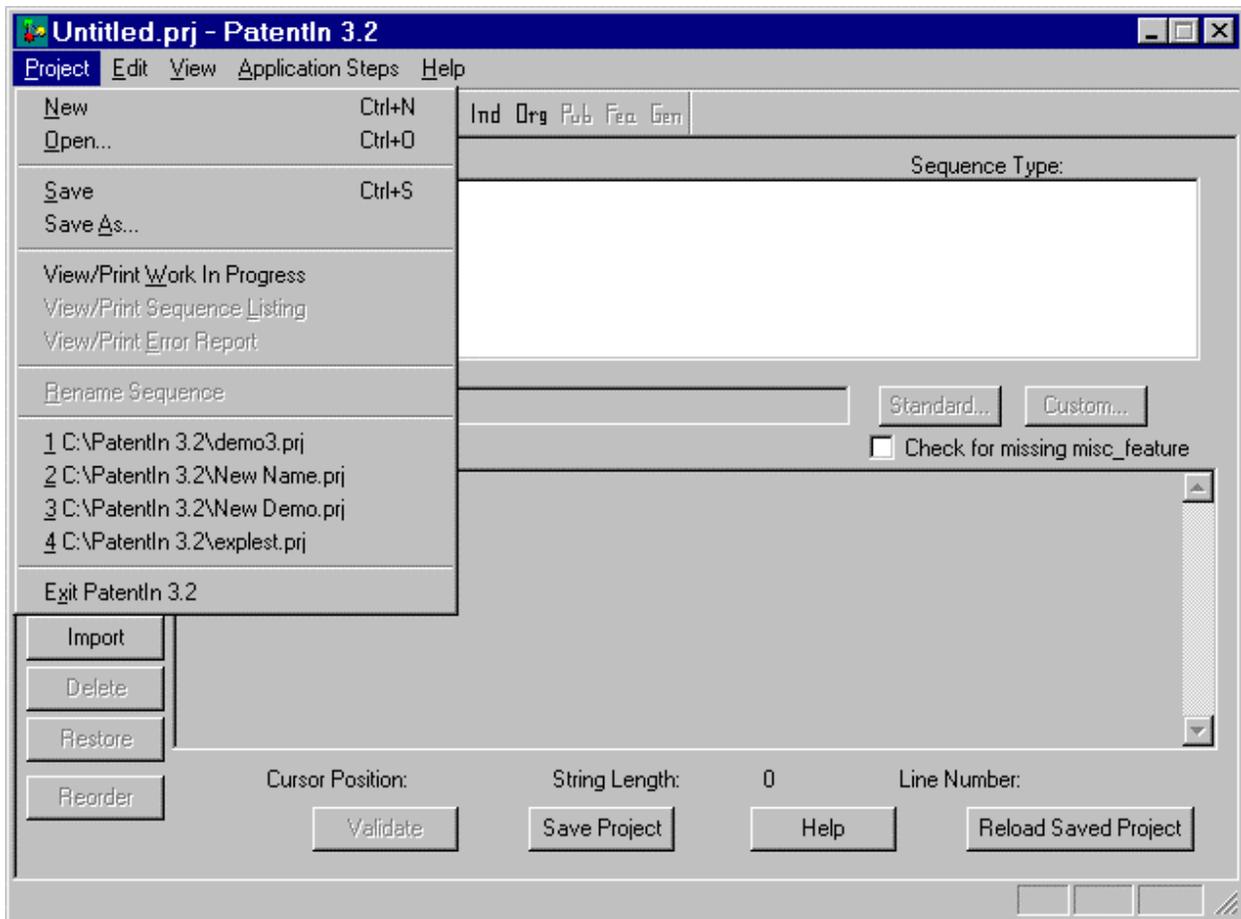


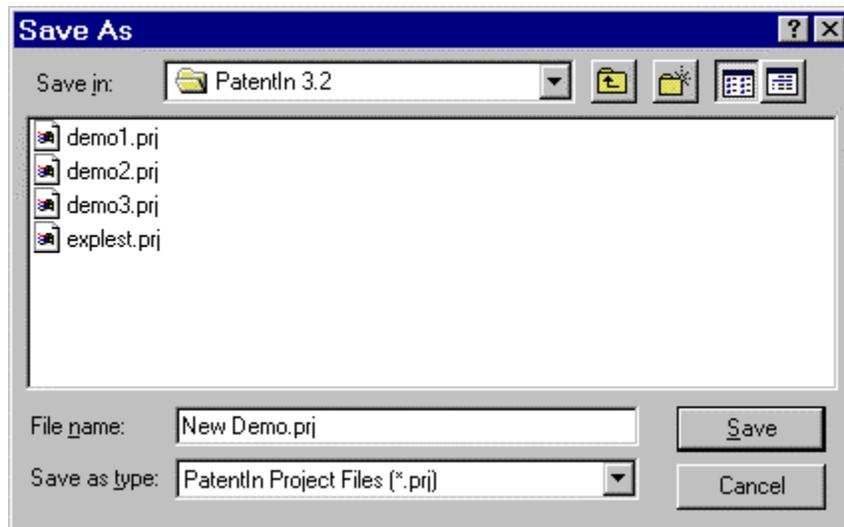
Figure 3-2: Project Menu

### 3.3 CREATING AND SAVING A NEW PROJECT

#### To create and save a new project:

Begin building the new file upon opening the main screen, Sequence Screen, or:

1.  Select **New** from the Project menu. This clears all of the current project information.
2.  Select **Save** from the Project menu. The Save As Screen (Figure 3-3) appears.
3.  Enter the new file name into the **File Name** dialog box in a desired folder.
4.  Click on **Save** to create the new file.
5. The name for the new project will be displayed on the upper left-hand corner of the screen.



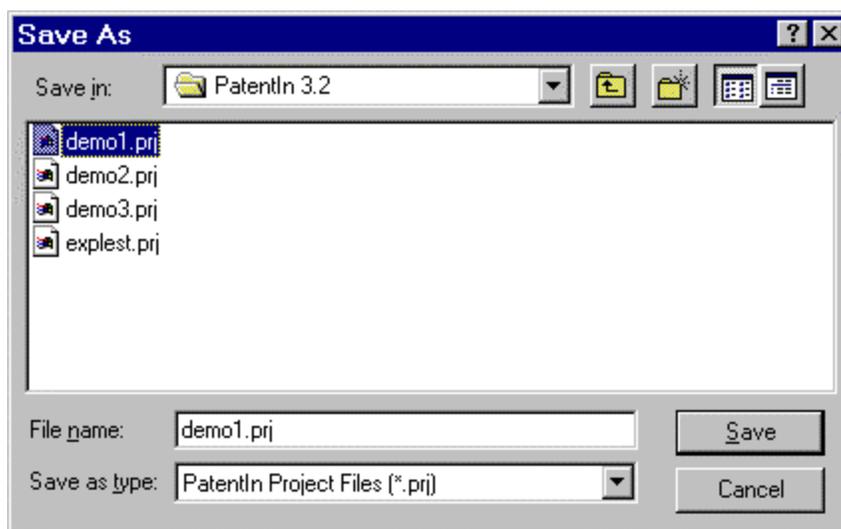
**Figure 3-3: Save As Screen**

 **NOTE:** Both **Save** and **Save As** save only the project (\*.prj) file. The generated listing is saved, as a text file when the project is generated. It is saved as a file with the same file name as the project and has the extension “ST25.txt”.

### 3.4 OPENING A PROJECT

#### To open an existing project:

1. Select **Open** from the **Project** menu. The Open Screen (Figure 3-4) appears.
2. Open the directory where the file is located.
3. Double-click on the file name to open the file.
4. You are returned to the main screen. The name of the opened project is displayed in the upper left-hand corner of the PatentIn screen, indicating the project is active.



**Figure 3-4: Open Screen**

**i** SPECIAL NOTE FOR USERS WITH VERY LARGE SEQUENCES AND/OR LARGE NUMBERS OF SEQUENCES: it takes some time for a large project to clear from memory. This is especially noticed when immediately reopening the project.

### 3.5 SAVING A PROJECT

#### To save a Project:

When a project is saved for the first time, the user is automatically prompted to enter a file name.

1.  Select **Save** from the **Project** menu. The Save As Screen (Figure 3-5) will appear if the project has not previously been named. Otherwise, the project will be saved as the previously opened or created name.
2.  Select the directory where you want to save the file.
3.  Type the new file name in the **File Name** dialog box.
4.  Click on the **Save** button to save the project with the new file name.
5. PatentIn returns you to the main screen. The new name is displayed in the upper left-hand corner of the screen, indicating that the project is active.

#### To save under a different file name:

1.  Select **Save As** from the **Project** menu. The Save As Screen (Figure 3-5) will appear.
2.  Select the directory where you want to save the file.
3.  Type the new file name in the **File Name** dialog box.
4.  Click on the **Save** button to save the project with the new file name.
5. PatentIn returns you to the main screen. The new name is displayed in the upper left-hand corner of the screen, indicating that the project is active.

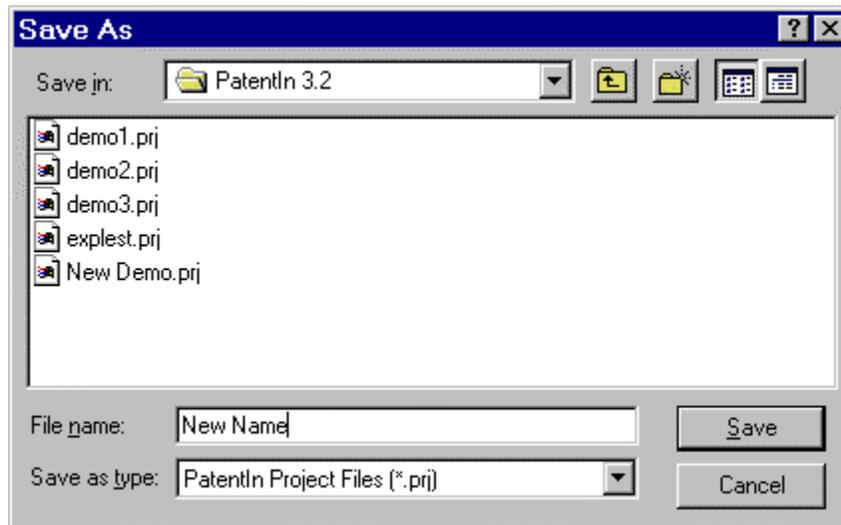


Figure 3-5: Save As Screen

### 3.6 VIEWING A WORK FILE

The user can view the current work in progress by creating a work file. This work file provides a vehicle for the user to view the data for the entire project in a single place instead of reviewing each individual screen. Use caution not to confuse the work file with the sequence listing.

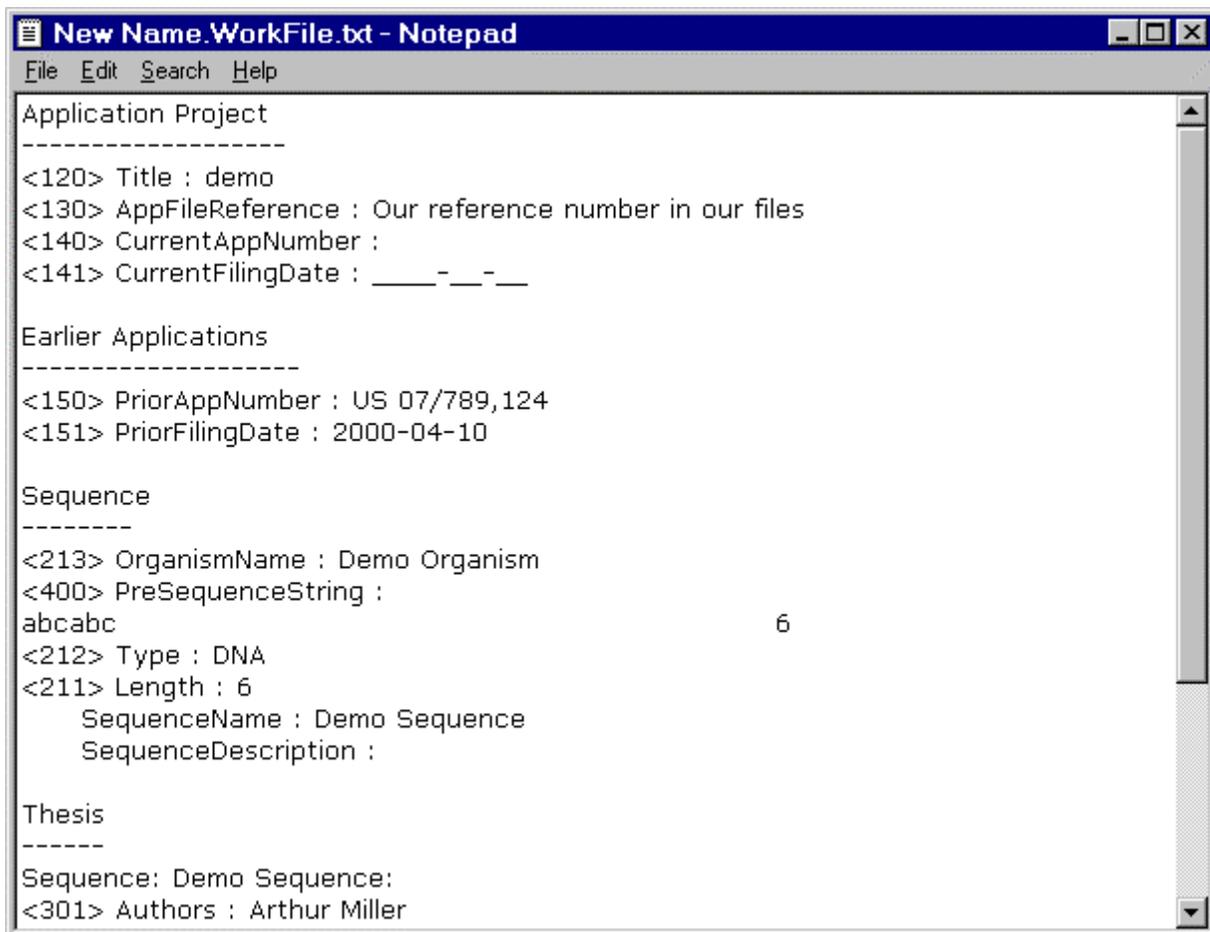
#### To see the Work File:

From the Project Menu,  select **View/Print Work in Progress**.

 NOTE: This work file was built by selecting **Create Work File**.

### 3.7 VIEWING WORK IN PROGRESS

PatentIn provides the user with an on-screen display of the Patent Application with the View Work in Progress Window (Figure 3-6).



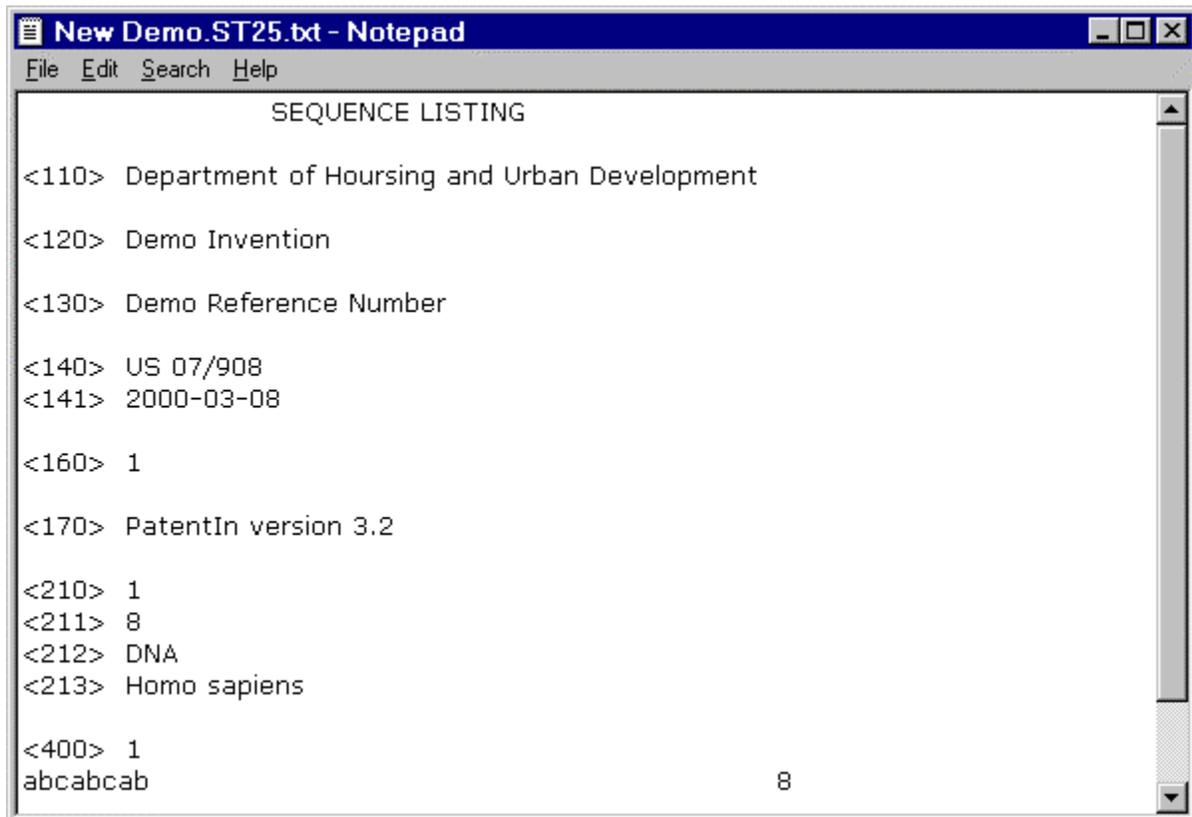
**Figure 3-6: View Work in Progress Window**

**To view the current patent application:**

1. From the Project menu,  select View Work in Progress.
2. To print the error report,  click on the File then Print.
3. To exit the screen,  click on File Exit.

**3.8 VIEW A SEQUENCE LISTING**

PatentIn provides the user with an on-screen view of the Sequence Listing with the View Sequence Listing Window (Figure 3-7).



**Figure 3-7: View Sequence Listing Window**

**To view the sequence listing:**

1. From the Project menu,  select **View Sequence Listing**.
2. To print the error report,  click on the **File** then **Print**.
3. To exit the screen,  click on **File** then **Exit**.

 **NOTE:** The sequence must first be generated.

### 3.9 VIEWING ERROR REPORTS

PatentIn provides the user with an on-screen Error Report with the View Error Report Window (Figure 3-8), if one exists, for the opened project.



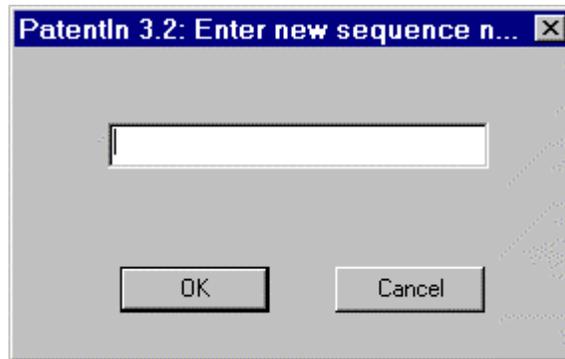
**Figure 3-8: View Error Report Window**

**To view an error report:**

1. From the **Project** menu,  select **View Error Report**.
2. To print the error report,  click on the **File** then **Print**.
3. To exit the screen,  click on **File** then **Exit**.

### 3.10 RENAME SEQUENCE

A new feature of PatentIn is the ability to change the name of a sequence (Figure 3-9).

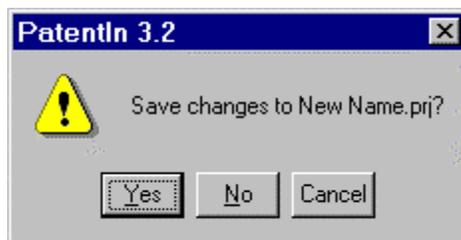


**Figure 3-9: Rename Sequence Screen**

1. To open the **Rename Sequence** screen,  click on the Sequence Name.
2. From the Project Menu,  select **Rename Sequence**.
3.  Type the new sequence name in the **Rename Sequence** dialog box.
4.  Click on the **OK** button.

### 3.11 EXIT PATENTIN

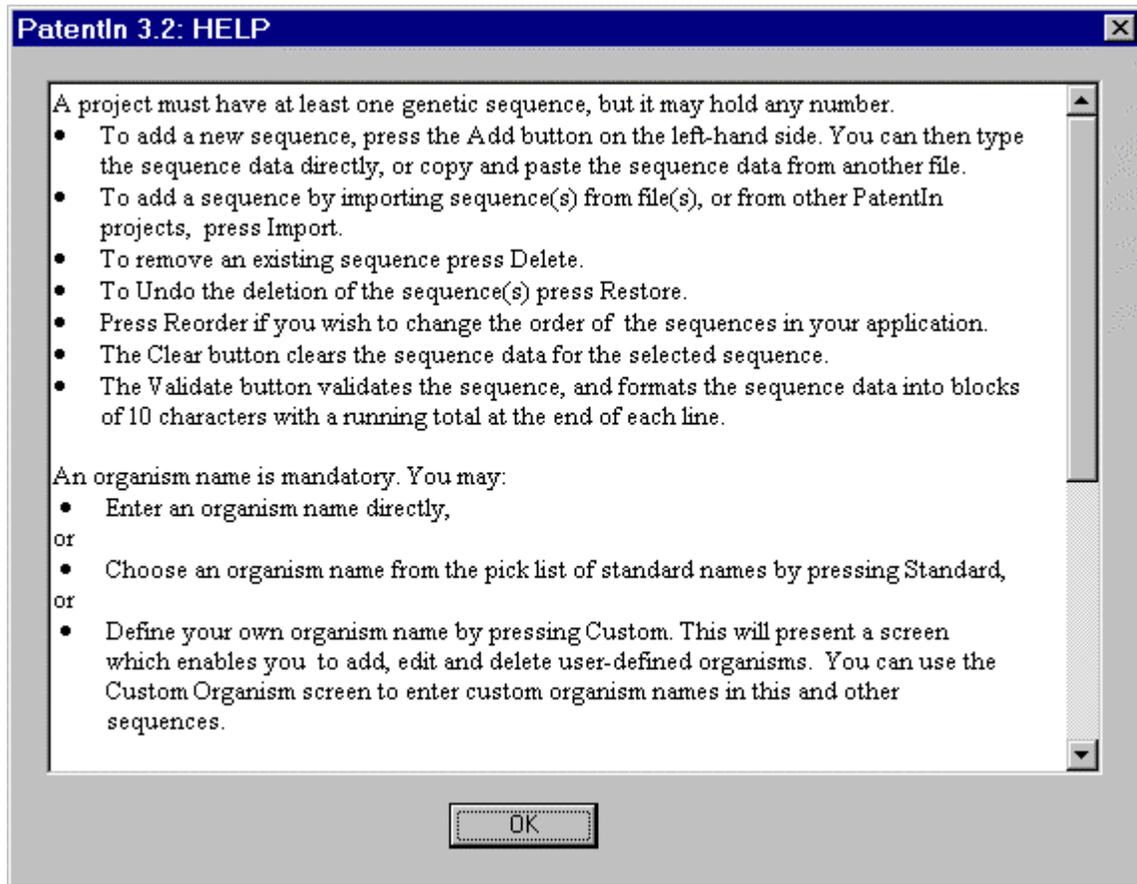
If the project does not have a current save, the user will be queried with the Exit PatentIn Screen (Figure 3-10) about whether the project should be saved:



**Figure 3-10: Exit PatentIn Screen**

### 3.12 HOW TO USE ONLINE HELP

On-line help is available for most of the PatentIn screens. The Help Screen (Figure 3-11), presents a typical help screen retrieved by pressing F1 or the help button. This example is the help screen access from the Sequence Screen (Figure 3-1).

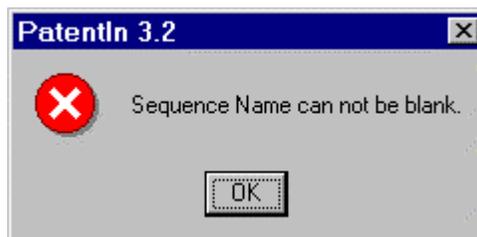


**Figure 3-11: Help Screen**

1. To exit the **Help** screen,  click on the **OK** button.

### 3.13 MESSAGE DIALOG

The Message Dialog Screen (Figure 3-12) is a screen that appears if one of the action buttons (Add, for example, described in Section 4, Project and Applicant Data) is pressed and an entry has not been made to the input area of the screen.



**Figure 3-12: Message Dialog Screen**

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 4**  
**PROJECT AND APPLICANT DATA**

## SECTION 4 PROJECT AND APPLICANT DATA

Once the Sequence Listing data file has been created, the user can add information to the application.

### 4.1 APPLICATION STEPS MENU

The Application Steps Menu (Figure 4-1) selections are available when a project is begun, has been created or selected. The project name is visible at the upper left corner of the screen. In this example the project has opened a previously existing file. New Name is shown for the project name.

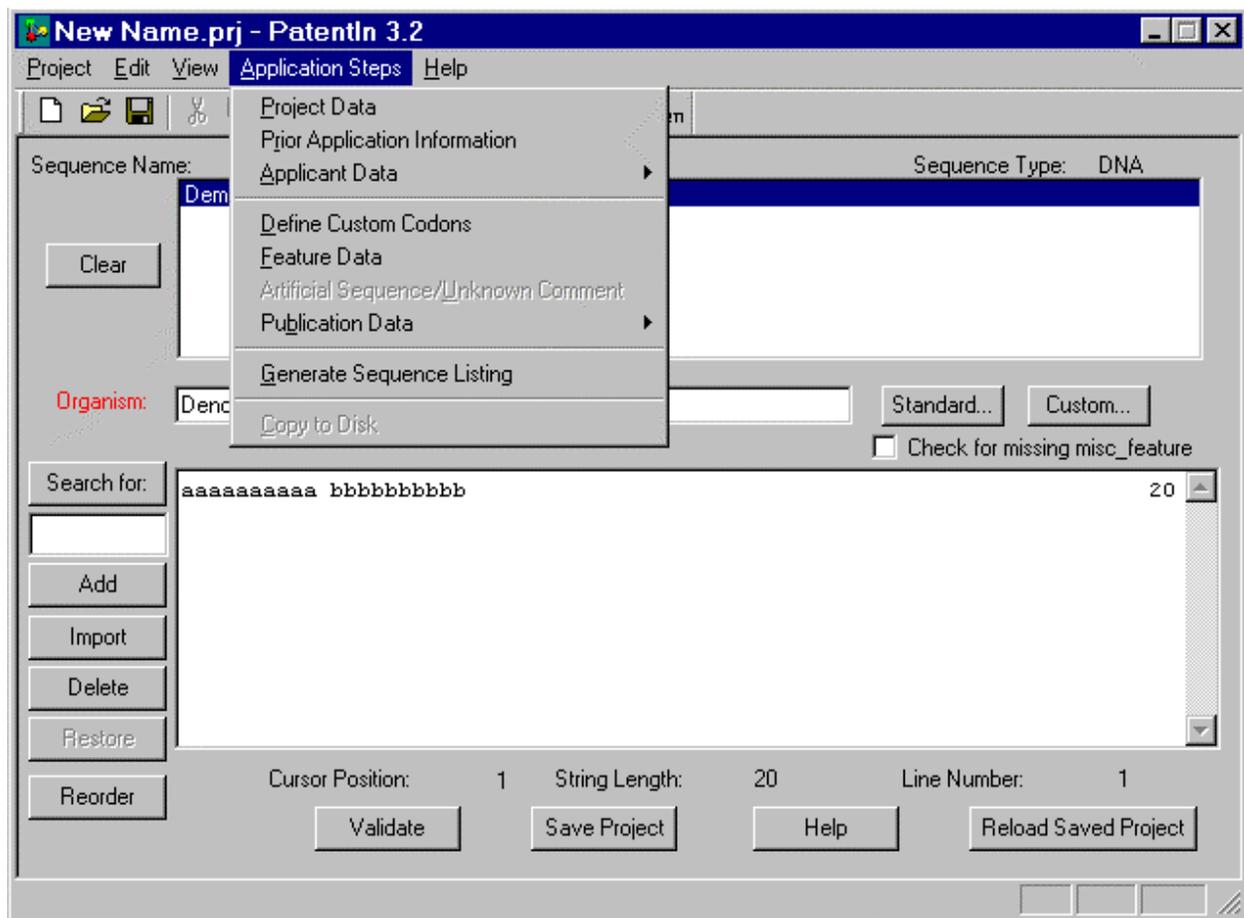


Figure 4-1: Application Steps Menu

## 4.2 PROJECT DATA

The Project Data Screen (Figure 4-2) provides the user with input fields to establish the identifying information for the new invention. This information is the key that establishes the title of the invention and the filing date.

**i** NOTE: The mandatory information fields (Title Of Invention and Application File Reference) are in red.

PatentIn 3.2: Project Data - New Name.prj

Title Of Invention: Demo Title

Current Application Number: \_\_\_\_\_  
(US 07/999,999 or PCT/US 96/999999)

Current Filing Date: \_\_\_\_-\_\_-\_\_  
(YYYY-MM-DD)

Application File Reference: Demo Reference Number

Validate Save Project OK Cancel Help

Figure 4-2: Project Data Screen

### To enter Project Data:

1. Enter the **Title of the Invention**. This information is mandatory.
2. Enter the **Current Application Number**, if one exists. If an application is entered the current filing date becomes mandatory.
3. Enter the **Current Filing Date**. The date format is numeric: YYYY-MM-DD.
4. Enter the **Application File Reference**.
5. To validate the information entered, click on **Validate**.
6. To save the information, click on the **Save Project** button.

## 4.3 PRIOR APPLICATION INFORMATION

Entering information about prior applications is optional since such information is available to the examiner elsewhere in the application file wrapper. Any number of prior applications may be

included on the Prior Application Information Screen (Figure 4-3). They will be displayed in the table in the order entered and may be selected for editing or deleting.

Item	Application Number	Filing Date
1	US 07/789,124	2000-04-10

**Figure 4-3: Prior Application Information Screen**

**To enter information about a Prior Application:**

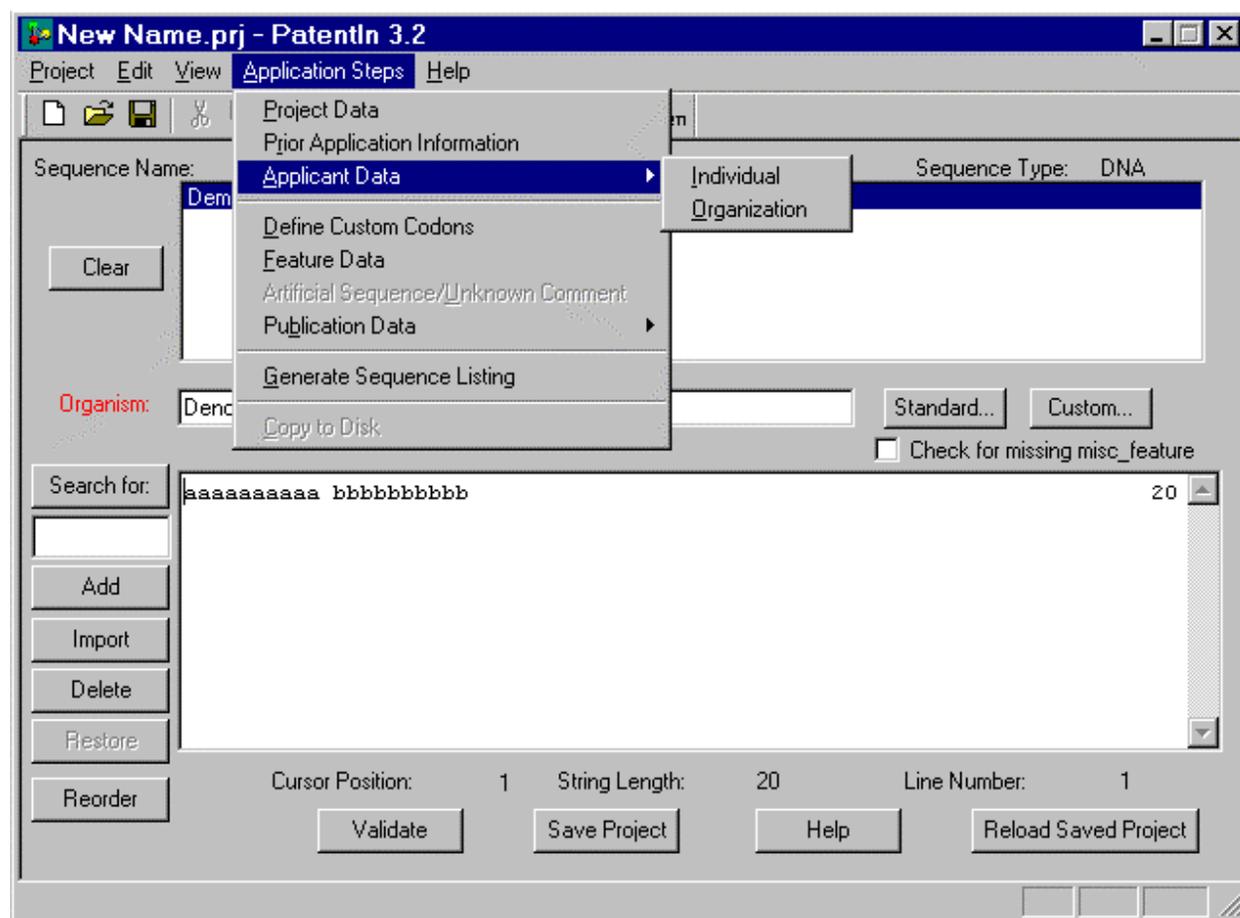
1. Enter the **Prior Application Number**. If a prior application number is entered then the prior application date becomes mandatory.
2. Enter the **Prior Application Filing Date**. The date format is numeric: YYYY-MM-DD.
3. To clear the information in the **Edit Prior Application** area, click on **Clear**.
4. To insert the information to the list, select the item you want the information to follow, enter the **Prior Application Number** and the **Prior Application Filing Date**, and then click on the **Insert** button.
5. To replace an entry from the list, select the item, enter the **Prior Application Number** and the **Prior Application Filing Date**, then click **Replace**.
6. To delete an entry from the list, select the item from the list, and then click on the **Delete** button.
7. To validate the information entered, click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.

8. To save the information,  click on the **Save Project** button.
9. To validate and close,  click on the **OK** button.

 **NEW TO PATENTIN:** In PATENTIN 3.2, when **OK** is clicked the data in the edit field(s) will be inserted into the list when they are different from the selected row.

#### 4.4 APPLICANT DATA

The Applicant Data Screen (Figure 4-4) allows the user to input information for an Individual or Organizational applicant. Select Applicant Data from the Application Steps menu, then select either Individual or Organization from the next menu. If Individual is selected, the Individual Applicants Screen (Figure 4-5) will appear. If Organization is selected, the Organization Applicants Screen (Figure 4-6) will appear.



**Figure 4-4: Applicant Data Screen**

#### 4.4.1 Individual Applicants

The Individual Applicants Screen (Figure 4-5) allows the user to enter information about an individual applicant. Only the name(s) of the applicant(s) will appear on the sequence listing, the spaces for the other information are for the user's convenience.

**i** NOTE: The field names in red (Last Name and First Name) are mandatory information.

Item	Individual Name	Phone Number
------	-----------------	--------------

**Figure 4-5: Individual Applicants Screen**

#### To enter information about an Individual Applicant:

1. From the Application Steps menu, Select Applicant Data, then Select Individual.
2. Enter the User's Last Name (surname).
3. Enter any Suffix the user has on his/her name (e.g., Jr., III).
4. Enter the User's First Name.
5. Enter the User's Middle Initial.

6.  Enter the User's Street Address, City, State/Province, Country, Zip/Postal Code, Phone Number, Fax Number, and Electronic Mail Address.
7. To clear the information about the Individual Applicant,  click on **Clear**.
8. To insert the information to the list,  select the item you want the information to follow,  enter the **Edit Individual Applicant** information, then  click on the **Insert** button.
9. To replace an entry from the list,  select the item,  enter the **Edit Individual Applicant** information, and then  click Replace.
10. To delete an entry from the list,  select the item from the list, and then  click on the **Delete** button.
11. To validate the information entered,  click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has yet been inserted, is not validated.
12. To validate and close,  click on the **OK** button.
13. To add another applicant to the list, repeat Steps 2-12.

 **NOTE:** The phone number, fax number and zip code are no longer validated.

 **NEW TO PATENTIN:** In PATENTIN 3.2, when **OK** is clicked the data in the edit field(s) will be inserted into the list when they are different from the selected row.

#### 4.4.2 Organization Applicants

The Organization Applicants Screen (Figure 4-6) allows the user to enter information about an organization applicant. As with individual applicants, only the name of the organization appears on the sequence listing and the spaces for the other information are for the user's convenience.

PatentIn 3.2: Organization Applicants - New Name.prj

Edit Organization Applicant

Organization:

Street Address:

Clear City:  State / Province:

Country:  Zip Code:

Phone Number:  Fax Number:

Electronic Mail Address:

Applicant List:

Item	Organization Name	Phone Number
------	-------------------	--------------

Insert

Replace

Delete

Validate Save Project OK Cancel Help

**Figure 4-6: Organization Applicants Screen**

**To enter information about an Organization Applicant:**

1. From the **Applicant Steps** menu, Select **Applicant Data**, then Select **Organization**.
2. Enter the **Organization's** Name.
3. Enter the Organization's Street Address, City, State/Province, Country, Zip/Postal Code, Phone Number, Fax Number, and Electronic Mail Address.
4. To clear the information about the Edit Organization Applicant portion of the screen, click on **Clear**.
5. To insert the information in the list, select the item you want the information to follow, enter the **Edit Organization Applicant** information, then click on the **Insert** button.
6. To replace an entry from the list, select the item, enter the **Edit Organization Applicant** information, and then click **Replace**.
7. To delete an entry from the list, select the item from the list, and then click on the **Delete** button.
8. To validate the information entered, click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.
9. Repeat steps 2 through 9 until all applicant information has been included.

10. To validate and close,  click on the **OK** button.

 **NOTE:** The phone number, fax number and zip code are no longer validated.

 **NEW TO PATENTIN:** In PATENTIN 3.2, when **OK** is clicked the data in the edit field(s) will be inserted into the list when they are different from the selected row.

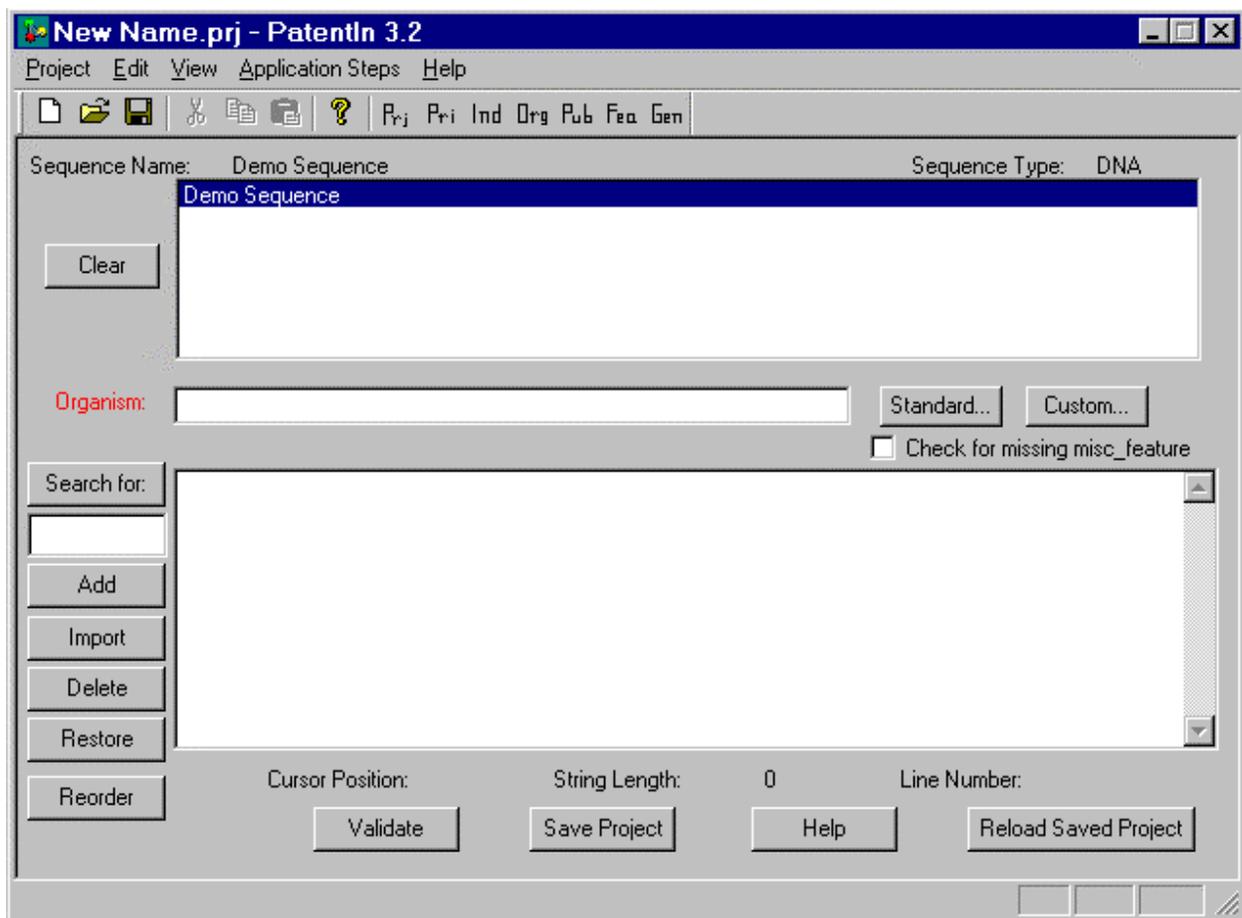
October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 5**  
**SEQUENCE DATA**

## SECTION 5 SEQUENCE DATA

### 5.1 SEQUENCE

The Sequence Screen (Figure 5-1) is where you create and modify sequences. You can create and edit custom codons and custom organism names from this screen. It also provides a search function where a genetic sequence may be entered and searched for in the files for this project. The user will access this screen immediately after PatentIn is started.



**Figure 5-1: Sequence Screen**

**i** NOTE: To begin entering a new sequence you must first have a sequence name. See Section 5.2.

#### **To select a sequence for editing:**

**h** Select a sequence name from the list of sequences.

**The following sequence characteristics are displayed:**

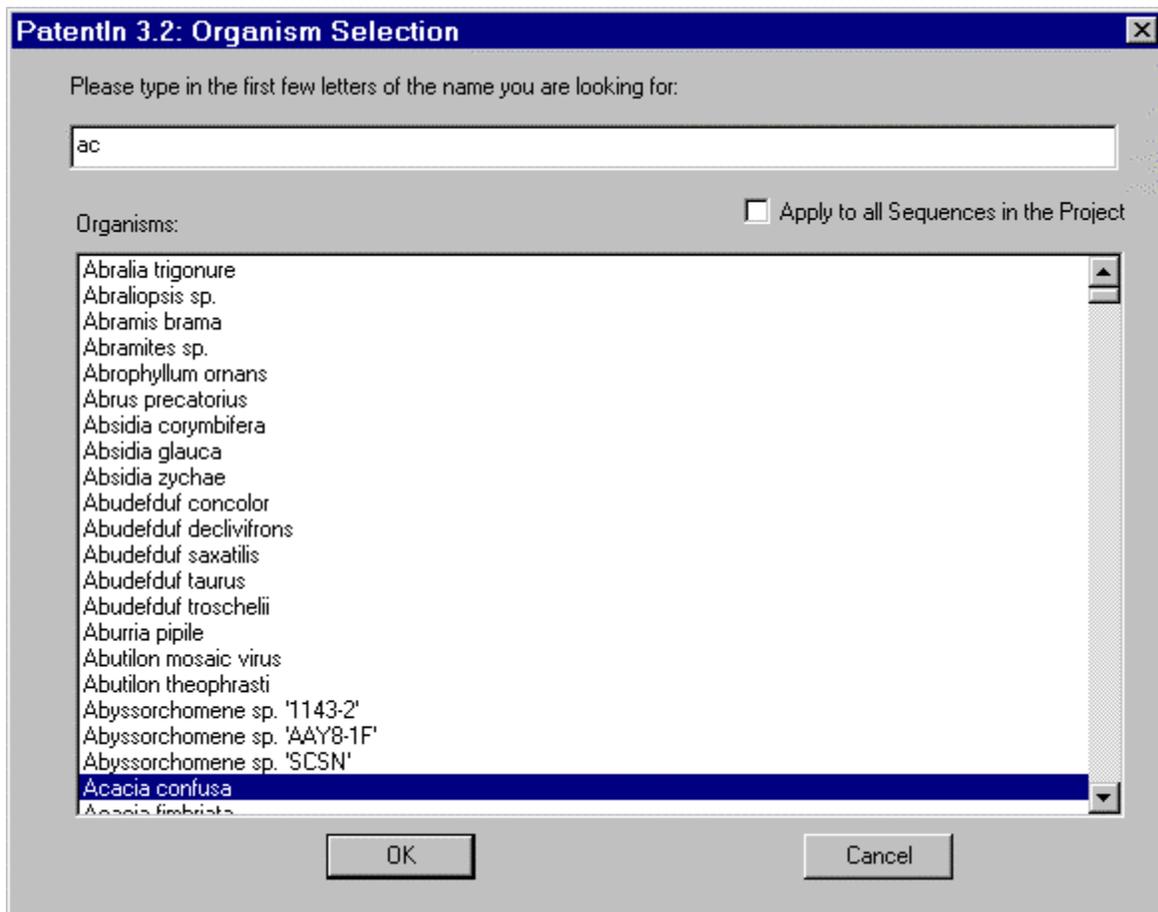
**Cursor Pos** - Shows the current cursor position. This field is blank when there is no sequence.

**String Length** - Shows the length of the sequence string on the line.

**Line Number** - Indicates the line number of the cursor position relative to the beginning of the string.

### 5.1.1 Selecting a Standard Organism

This screen, Figure 5-2, enables the user to select an organism name from common organisms. Included in its capabilities is an attempt to match on partially input names.



**Figure 5-2: Selecting an Organism Screen**

**To select an Organism Name:**

1. Click on the **Standard** button (Figure 5-1)
2. Begin entering characters for the organism you are looking for.

3.  Click on the **Apply to all Sequences in the Project** checkbox to enable/disable assigning this organism name to all sequences currently in the project.
4.  Click on the **OK** button to enter the selected organism name.

### 5.1.2 Creating default explanations for "n"s and "Xaa"s

PatentIn 3.2 can create default explanations for "n"s and "Xaa"s in the sequence listing that have no user-supplied explanations. The default explanations take the form of misc\_features containing the location of the variable characters, and the message, "n is a, c, g, or t" for nucleic acid sequence or "Xaa can be any naturally occurring amino acid" for protein sequences. The **Check for missing misc\_feature** box on the Sequence Screen (Figure 5-1) can be used to turn this feature on or off. Users may wish to check the box and press Validate to see if there are any variable characters not defined by a misc\_feature. If the box is unchecked, PatentIn 3.2 will create any missing misc\_features. Users may create selected definitions by hand and PatentIn 3.2 will provide the remaining ones.

### 5.1.3 Searching for a Sequence

**To search for a specific sequence:**

1.  Enter a particular substring (a feature, for example) in the edit field below the **Search for** button (Figure 5-1).
2.  Click on the **Search** button. The cursor will move the first instance of that sub-sequence beginning with its current position.

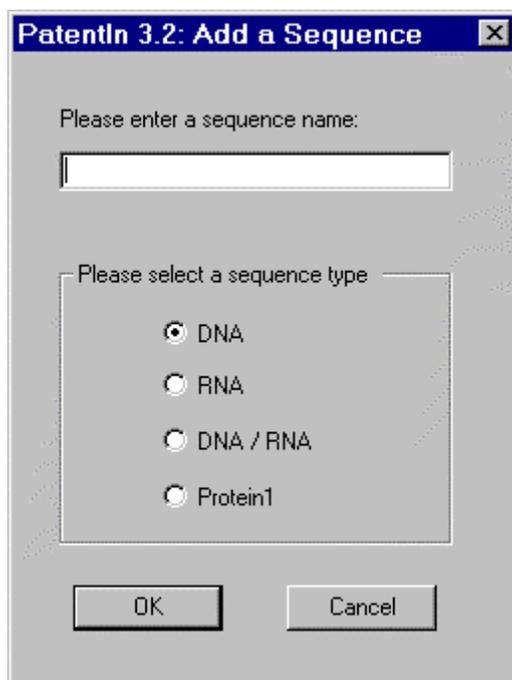
 **NOTE:** Each search is limited/truncated to 60 characters.

### 5.1.4 Clearing the Screen

1. To clear all of the screen about a specific selected sequence,  click on the **Clear** button (Figure 5-1).

## 5.2 ADDING A SEQUENCE

The **Add** button on the Sequence Screen (Figure 5-1) provides a means to enter a sequence name and to select a sequence type from a list of radio buttons (Figure 5-3).



**Figure 5-3: Add A Sequence Screen**

### To add a sequence:

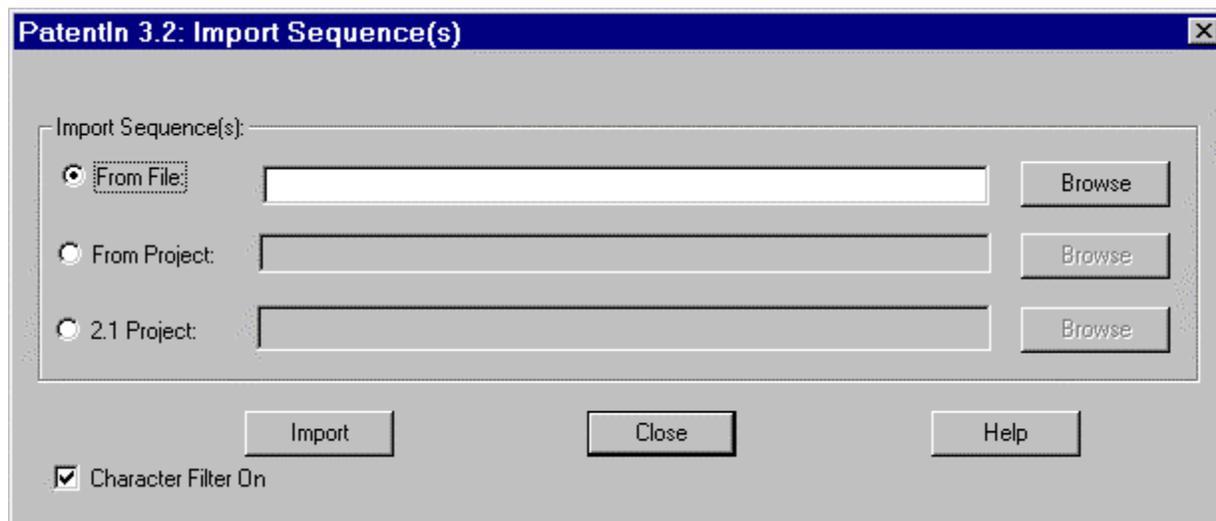
1. From the Sequence Screen,  select the **Add** button. The Add A Sequence Screen appears (Figure 5-3)
2.  Enter the sequence name into the dialog box.
3.  Select the sequence type by clicking on the  radio button next to the appropriate sequence type.
4.  Click **OK**.

You can now enter sequence strings in the edit field at the bottom of the screen. If you are using Windows 95, a maximum of 64K characters will display in the field. In Windows NT, 98, 2000 and XP, the upper limit is over 1 million sequence characters.

You can work around these limitations by using import files, rather than the Sequence Editor, to create and edit sequences.

### 5.3 IMPORTING A SEQUENCE

The **Import** button on the Sequence Screen (Figure 5-1) provides a means to import sequences from either a file, a project, or a 2.1 project by selecting one of three radio buttons.



**Figure 5-4: Import Sequence(s) Screen**

1. To use a sequence from a text file, click on the **Import** button on the Sequence Screen, and then click on the **From File** radio button (See 5.3.1).
2. To use a sequence from another project, click on the **Import** button on the Sequence Screen, and then click on the **From Project** radio button (See 5.3.3).
3. To use a project from a PatentIn 2.1, click on the **Import** button on the Sequence Screen, and then click on the **2.1 Project** radio button (See 5.3.4).
4. A **Browse** button is provided for each radio button to assist the user in providing the file folder and file name, and formatting the input for multiple file selections. Folder and file names with imbedded spaces are accepted.
5. The **Character Filter On** checkbox can be used in conjunction with the first radio button: “From File:”. The default state for this box is checked, turning on the function. Unchecking this box will allow only files that contain no extraneous errors to be imported. Leaving the box checked will still give the user a list of characters found to be in error, but the valid characters will be imported and placed in the project. Because PatentIn 3.x allows multiple sequences in a single file and thus uses the “<” characters for its header, this character will be understood to be a valid character, which will most likely cause a “missing header” error message and thus cannot be removed as an extraneous character. For best results it is recommended that the sequence files contain only the sequence, with spaces and numbering, and that any title or other text be deleted prior to importing. For example, if the title of the sequence is Genomic DNA, PatentIn 3.2 would filter out the characters e, o and i and so the first seven characters of your sequence would be “gnmcdna”.

**NOTE:** Importation of a 2.1 Project requires additional software to be installed. See Section 5.3.4.

**i** NOTE: For the Protein/3 selection, the data must be imported from a text file that contains only amino acid abbreviated names as shown in Appendix D: Conversion Table Between Nucleotide Triplets (Codons) And One- And Three-Letter Amino Acid Codes, PRT/3 Column. The PRT/3 strings are converted to PRT/1 characters for subsequent use in the Sequence Editor. A Protein/3 file without a header can also be imported.

### 5.3.1 Format for Multi-Sequence Data Files to be Imported by PatentIn 3.x

A sequence file is an ASCII text file containing one or more sequences. Each multi-sequence data file must begin with a header having the following format described in Section 5.3.1.1. The header must be the first non-blank text on its line.

#### 5.3.1.1 Sequence Header

The entire header must be on a single line (Table 5-1).

<*SequenceName*;*SequenceType*;*OrganismName*>

**Table 5-1: Header for Sequence**

<i>SequenceName</i>	the name of the sequence
<i>SequenceType</i>	one of the following: <ul style="list-style-type: none"> <li>• DNA</li> <li>• RNA</li> <li>• DNA/RNA</li> <li>• Protein/1</li> <li>• PRT</li> <li>• PRT/1</li> <li>• PRT1</li> <li>• Protein/3</li> <li>• PRT/3</li> <li>• PRT3</li> </ul>
<i>OrganismName</i>	the name of the organism, and is optional. If it is omitted, the header looks like: < <i>SequenceName</i> ; <i>SequenceType</i> >.

**i** NOTE: Notice that there are semi-colon separators. They are always required.

### 5.3.1.2 Sequence Data

The sequence data begin on the line following the header. The sequence data are a string of letters appropriate to the sequence type. The sequence data may span multiple lines. Sequence data may not contain spaces. A space signifies the end of the sequence data.

The sequence data are terminated by one or more spaces, or by the start of the next header. There may be one or more empty lines between the end of a sequence and the start of the next.

A two-sequence file might look like this Sample ASCII Sequence Data (Figure 5-5):

```
< First Sequence;DNA;Abies alba>
ttttcttattgtttctcctactgcttatcataatgattgtcgtagtggcttcctcatcgt
ctccccaccgcctaccacaacgactgccgcagcggattactaatagtatcaccaacagc
ataacaaaaagaatgacgaagaggggttgctgatggtgtcgcgcagcggcgtagcagaagga
gtggcggagggg

< Second Sequence;RNA; >
uuuucuuauuguuucuccuacugcuuaucauaaugauugucguaguggcuuccucaucgu
cuccccaccgccuaccacaacgacugccgcagcggauuacuaauaguaucaccaacagc
auaacaaaaagaaugacgaagagggguugcugauggugucgccgacggcguagcagaagga
guggcggagggg
```

Figure 5-5: Sample ASCII Sequence Data

### 5.3.2 Format for Single Sequence Data Files to be Imported by PatentIn 3.2

A sequence file is an ASCII text file containing one or more sequences. A single sequence data file does not require a Sequence Header. If the header is missing, the user is queried for the sequence type (Figure 5-6) and the file is assumed to be a single sequence data file.

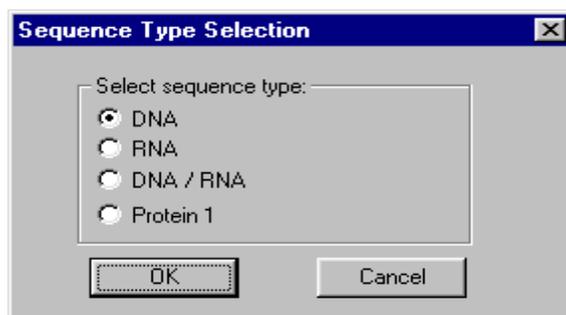
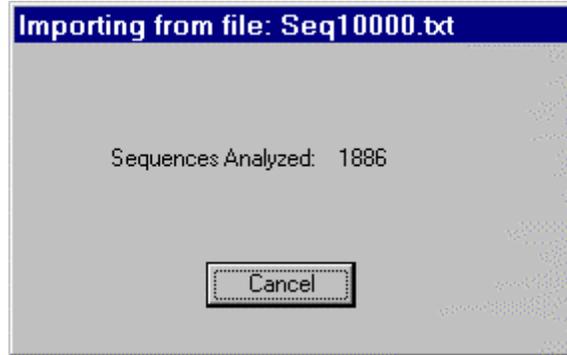


Figure 5-6: Sequence Type Selection Screen

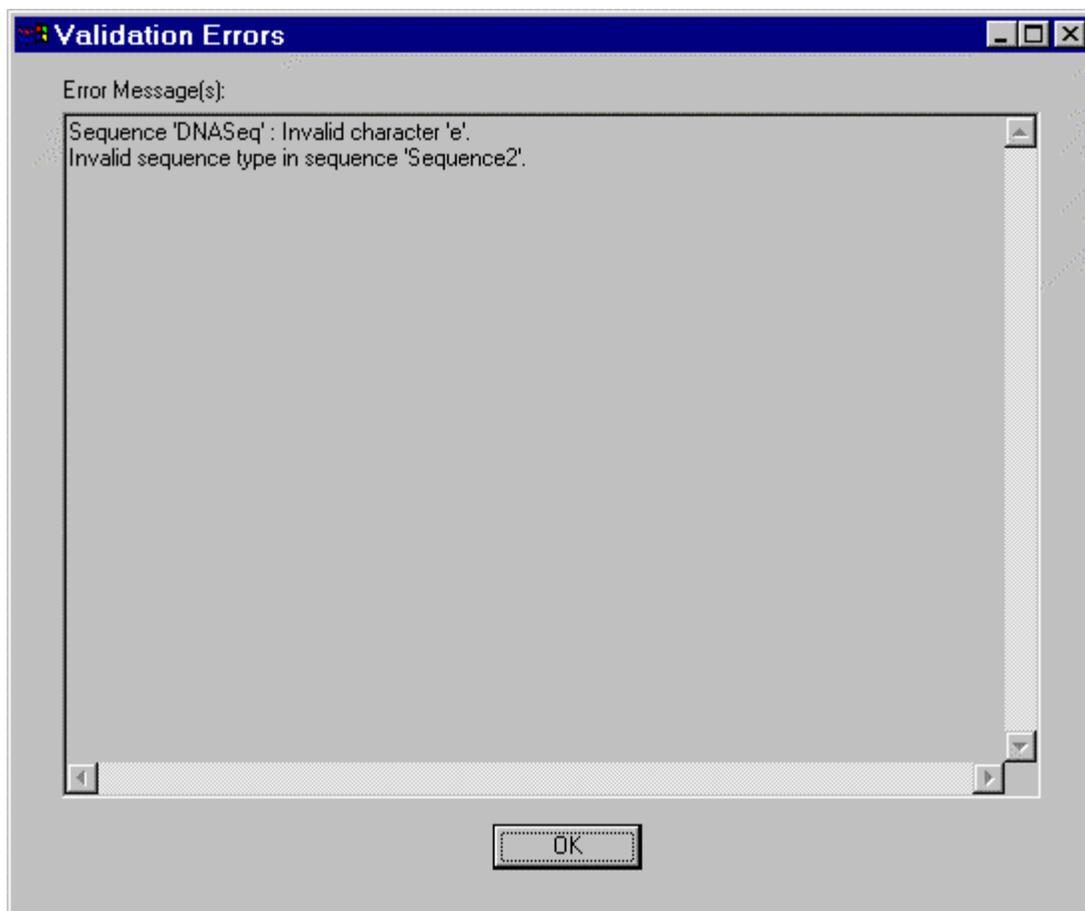
**i** NOTE: PatentIn 2.1 ASCII text files used the extension “.gbs” for its sequence files.

While the import feature is running, a screen displays the total number of sequences that have currently been analyzed. Figure 5-7, below, shows that a multi-sequence file named SEQ1000.txt has analyzed 513 sequences at the time this screen was captured.



**Figure 5-7: Sequences Being Imported Screen**

A validation errors screen is displayed if validation errors occur. In the next example (Figure 5-8), the file singlefilenoheader was specified as DNA instead of Protein 1. Nucleo is actually a file with the names of nucleotides, not the sequences.



**Figure 5-8: Validation Errors Screen**

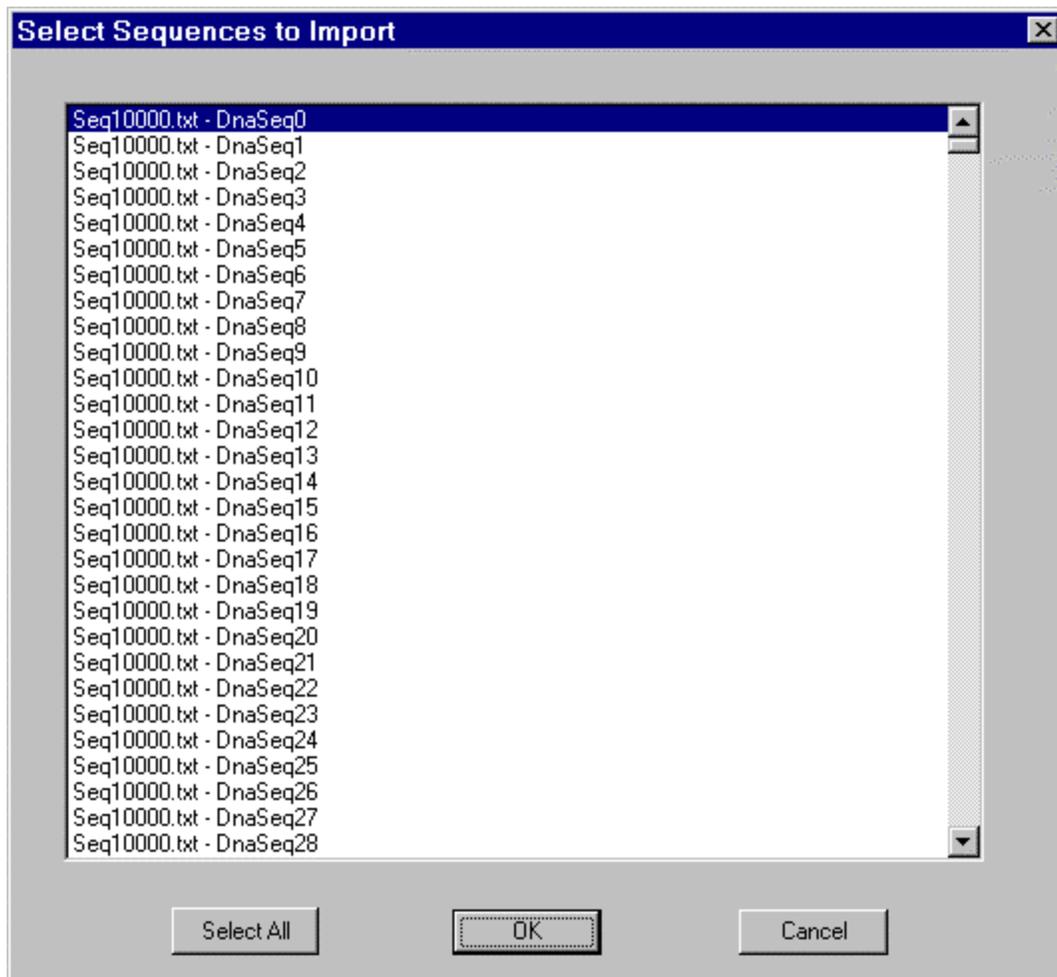
### 5.3.3 Importing Sequences from a Project

PatentIn provides a mechanism to import sequences from a PatentIn 3.x project (Figure 5-4: Import Sequence(s) Screen).

1. To use a sequence from a project file,  click on the **Import** button on the sequence screen (Figure 5-1) then  click on the **From Project** radio button.
2. The **Browse** button is provided to assist the user in providing the file folder and file name.
3. When a project has been selected a list of sequences in the project is displayed.
4.  Click on the sequence(s) to be imported.

 **NOTE:** By holding the Ctrl key down multiple sequences may be selected.

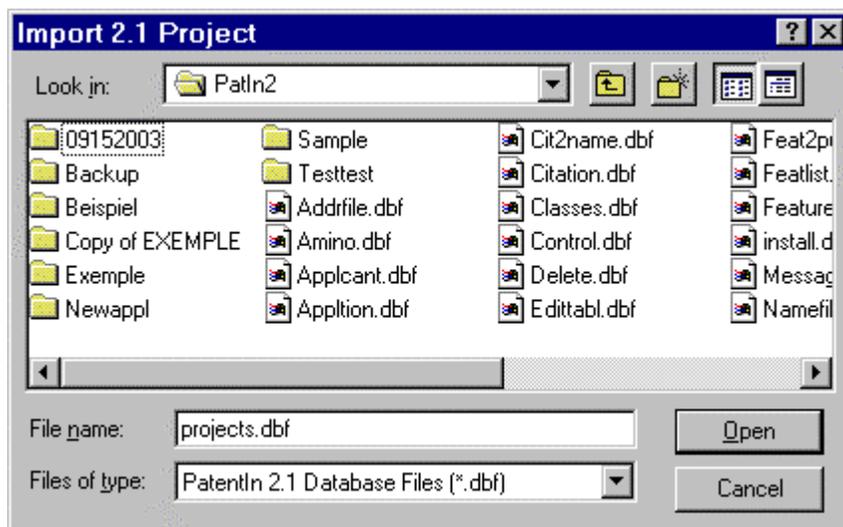
5. Or  select the **Select All** button to select all of the sequences (Figure 5-9).



**Figure 5-9: Select Sequences From Project Screen**

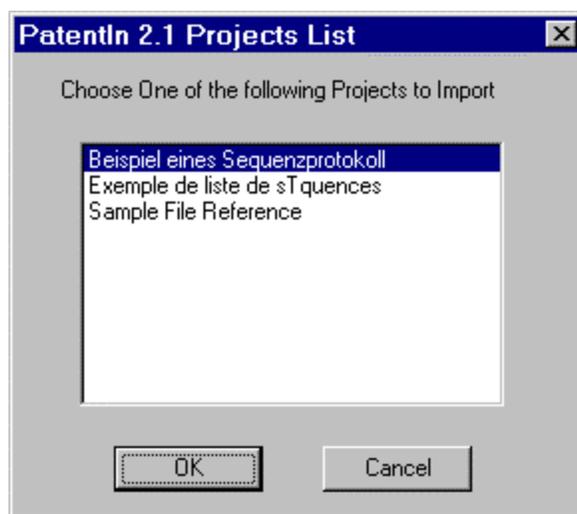
### 5.3.4 Importing a 2.1 Project

1. To use a sequence from a PatentIn 2.1 project file,  click on the **Import** button on the sequence screen, then  click on the **2.1 Project** radio button.
2. The **Browse** button is provided to assist the user in providing the file folder and file name. Notice that in this case a “.dbf” file is selected, not a project or text file. See Figure 5-10.



**Figure 5-10: Browse Window for 2.1 Project Import**

3. After the PatentIn 2.1 Project file is identified, the list of project long names is displayed. See Figure 5-11.



**Figure 5-11: PatentIn 2.1 Projects List**

4. Click on the project to be imported.
5. Click on the **OK** button.

#### 5.4 COPYING A SEQUENCE

PatentIn uses standard Windows-type edit features.

##### To copy a sequence:

1. Highlight the sequence to be copied.
2. Click on the **Edit** menu, then click **Copy** (Figure 5-12).

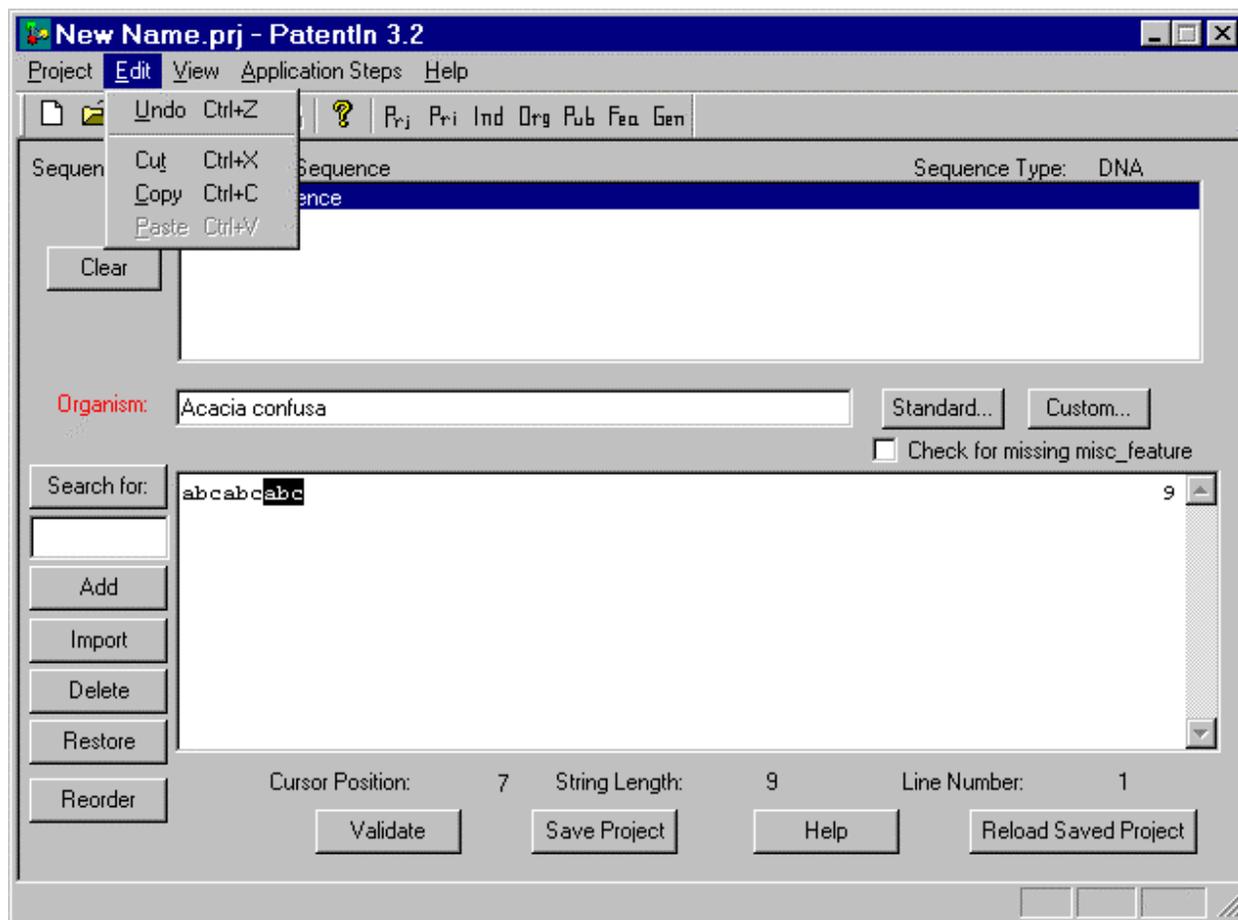


Figure 5-12: Edit Menu

## 5.5 PASTING A SEQUENCE

### To paste a sequence:

1. Position the cursor where the copied material is to be inserted.
2.  Click on the **Edit**, then  click **Paste**.

## 5.6 DELETING A SEQUENCE

### To delete a sequence:

1. Position the cursor where the copied material is to be deleted.
2.  Click on the **Delete** button.

## 5.7 SKIP A SEQUENCE

When a sequence is skipped, PatentIn 3.2 will not generate anything between the <210> and the <400> for this sequence.

### To skip a sequence:

1. Select the sequence name to be skipped.

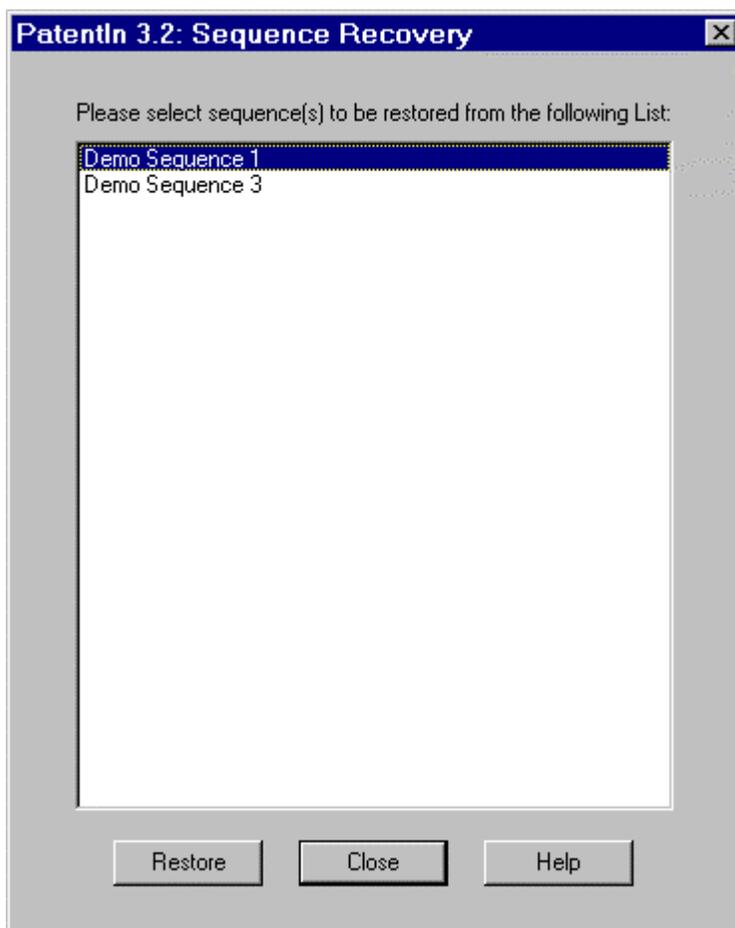
2. Delete all of the text from the sequence edit text box as well as all of the feature and publication data associated with the sequence. PatentIn 3.2 will automatically generate the sequence number and "000" for the <400> in the sequence listing.

## 5.8 RESTORING A SEQUENCE

When a sequence has been deleted, it can be restored until the current project update is terminated.

### To restore a sequence:

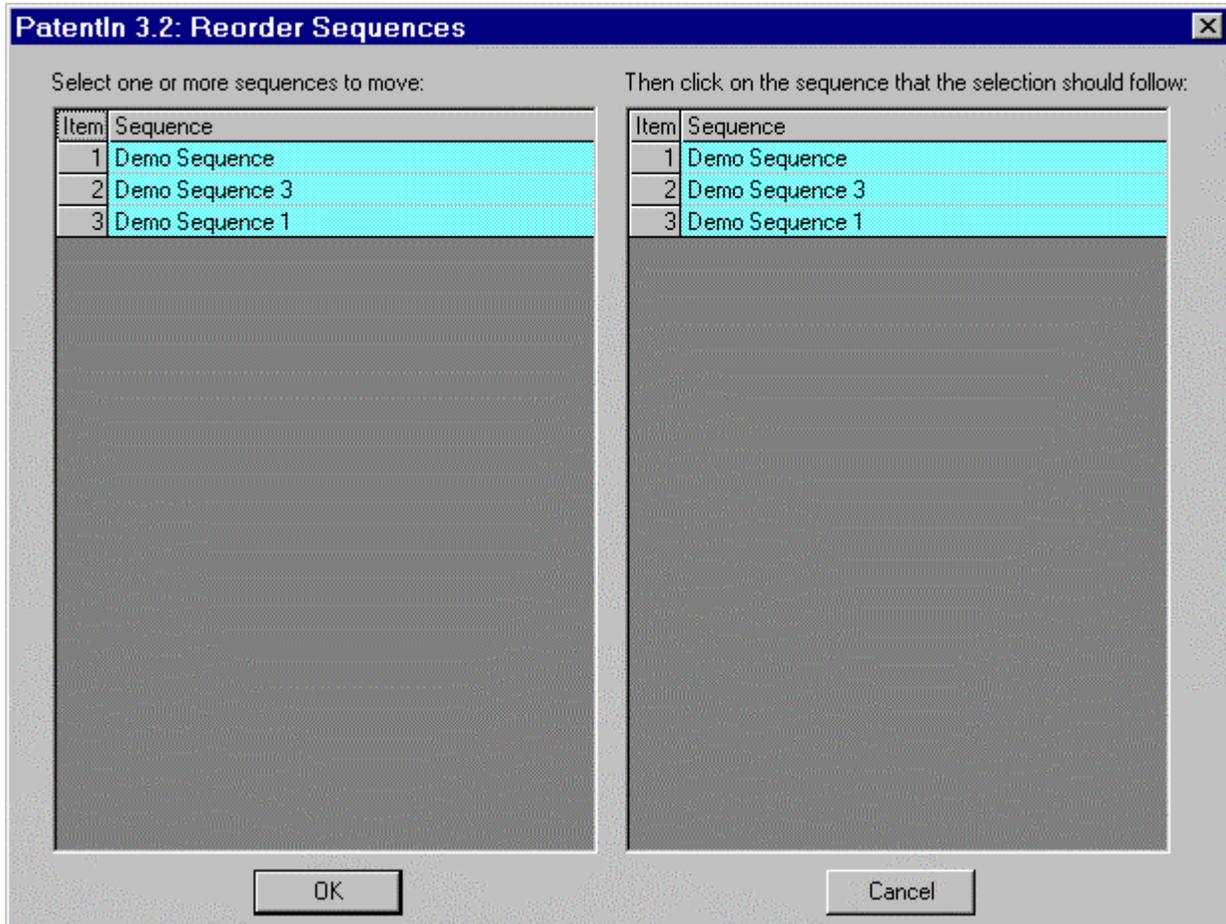
3.  Click on the **Restore** button on the main screen. The Sequence Recovery Screen (Figure 5-13) appears.
4.  Select the sequence(s) to be restored.
5.  Click on the **Restore** button.



**Figure 5-13: Sequence Recovery Screen**

## 5.9 REORDERING SEQUENCES

The Reorder Sequences Screen (Figure 5-14) provides the user with the means to compare the current sequence order and the new sequence order. The Current Sequence Order is displayed on the left of the screen. It displays the sequences in the order that the sequences were entered into the application. The New Sequence Order, displayed on the right, displays the sequences in the order that the user specifies by selecting a contiguous group of sequences from the left side and selecting a single sequence on the right side that they are to be placed after.



**Figure 5-14: Reorder Sequences Screen**

### To reorder sequences:

1.  Select the sequence row(s) from the menu on the left-hand side.
2.  Click on the row the sequence will follow on the right-hand side.
3. Repeat steps 1-2 until the sequences are in the desired order.

## 5.10 VALIDATING SEQUENCES

### To Validate the Sequences:

1. On the Sequence Screen (Figure 5-1),  click on the **Validate** button. A message screen will inform you if there was an error, otherwise Validation OK will appear on the status bar.

 NOTE: For the Sequence Data, validation is done for the selected sequence name.

## 5.11 SAVING SEQUENCES

### To Save a Sequence:

1. On the Sequence Screen (Figure 5-1),  click on the **Save Project** button. Your work will be saved in its current state.

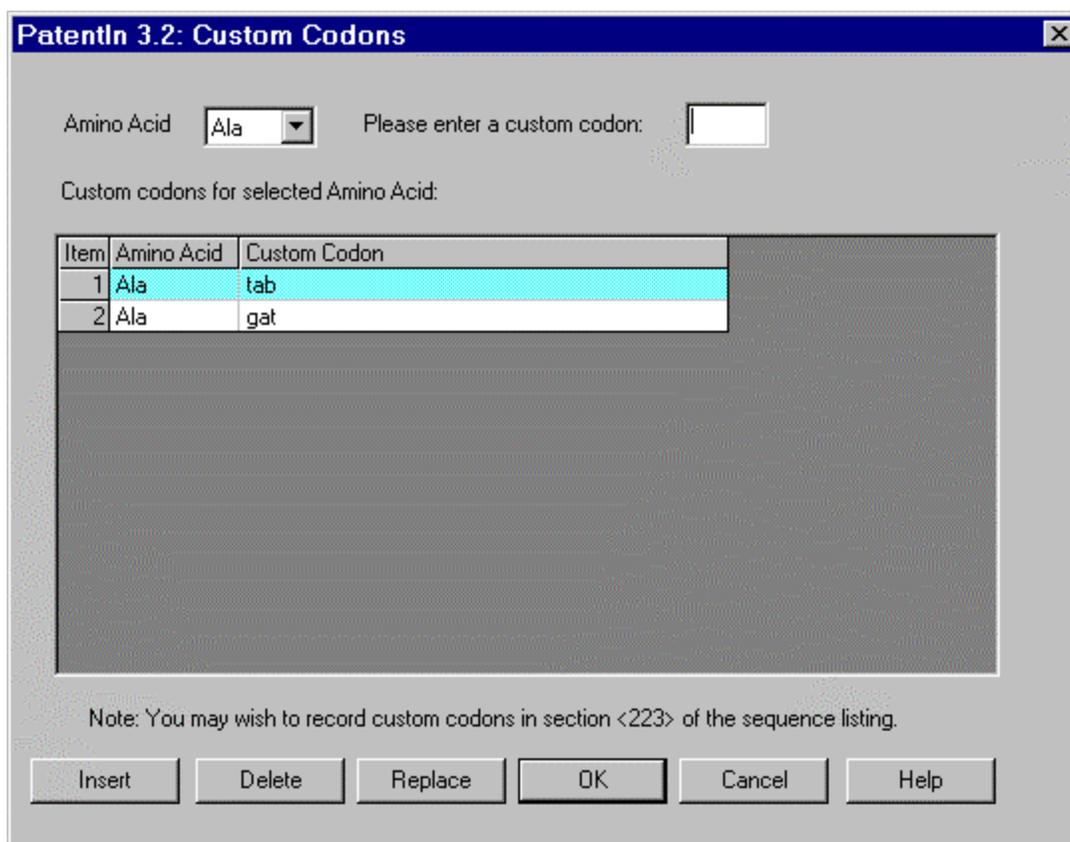
 NOTE: It is important to remember while working with very large or complex projects that saving often can save hours of rework, especially when there is a system failure.

## 5.12 RELOADING A SAVED PROJECT

The **Reload Saved Project** button is supplied, at the request of some of the users, to quickly load the current project from its last saved state (Figure 5-1).

## 5.13 ADDING CUSTOM CODONS

The Custom Codons Input Screen (Figure 5-15) provides the means to add Custom Codons to the list of standard codons on the user workstation. This screen is accessed from the Application Steps menu (Figure 4-1) by selecting the Define Custom Codons item.



**Figure 5-15: Custom Codons Input Screen**

**To add a Custom Codon:**

1. Select the **Define Custom Codons** item from the **Application Steps** menu (Figure 4-1).
2. Select an **Amino Acid** from the drop-down list (Figure 5-16).
3. Enter the **Custom Codon**.
4. Click on the **Insert** button.

**NOTE:** The format of this screen has changed. It now allows the user to see all of the custom codons on a single screen instead of having to select each amino acid individually.

**To delete a Custom Codon:**

1. Click on the Custom Codon in the list.
2. Click on the **Delete** button.

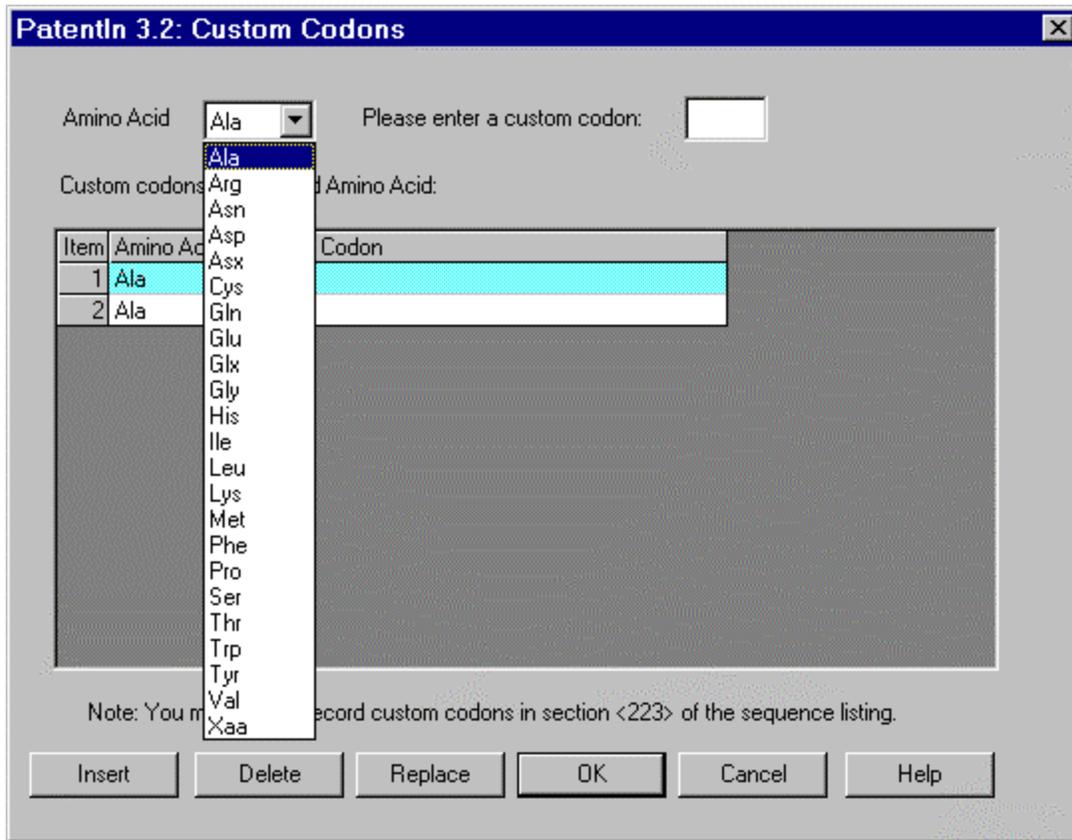
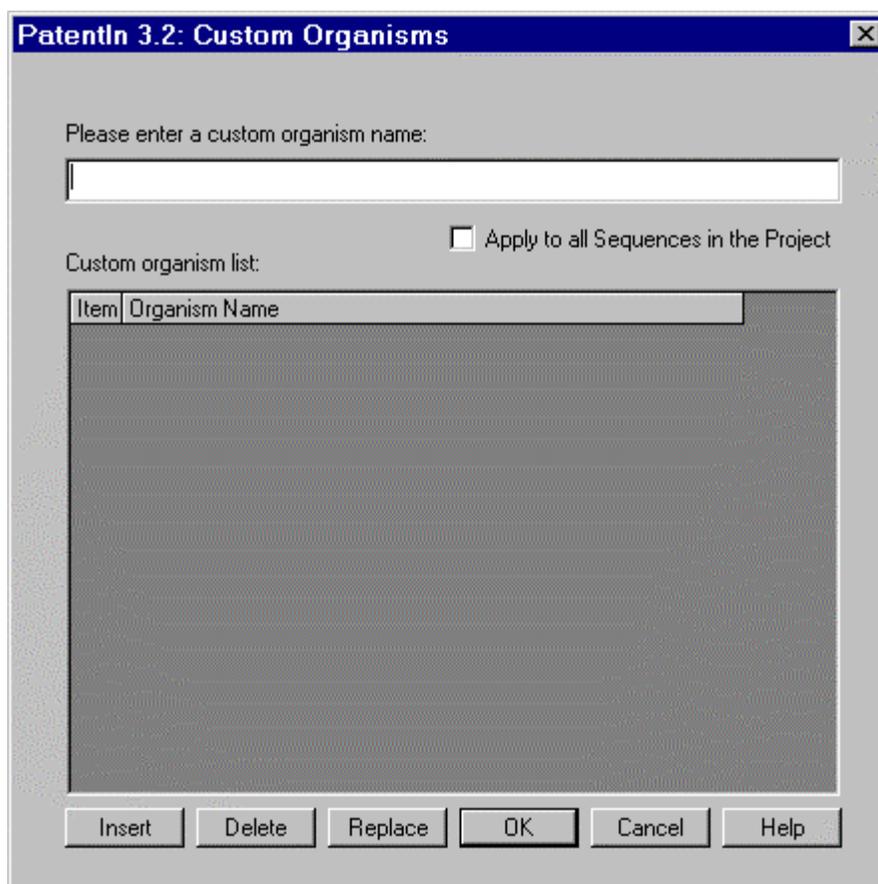


Figure 5-16: Amino Acid Drop Down List Screen

#### 5.14 ADDING A CUSTOM ORGANISM

The Custom Organism Input Screen (Figure 5-17) provides the means to add a custom organism to the list of custom organisms. It also enables you to select a custom organism to enter on the Sequence Screen (Figure 5-1). This screen is accessed from the Sequence Screen by selecting the **Custom** button. The user enters the custom organism into the screen, and then can manipulate the list by adding or deleting organisms.



**Figure 5-17: Custom Organism Input Screen**

**To add a Custom Organism:**

1.  Select the **Custom** button from the Sequence Screen.
2.  Enter the Custom Organism.
3.  Click on the **Insert** button.

**To delete a Custom Organism:**

1.  Click on the Custom Organism in the list.
2.  Click on the **Delete** button.

**To replace a Custom Organism:**

1.  Click on the Custom Organism in the list.
2.  Enter the Custom Organism.
3.  Click on the **Replace** button.

**To enter a Custom Organism on the Sequence screen:**

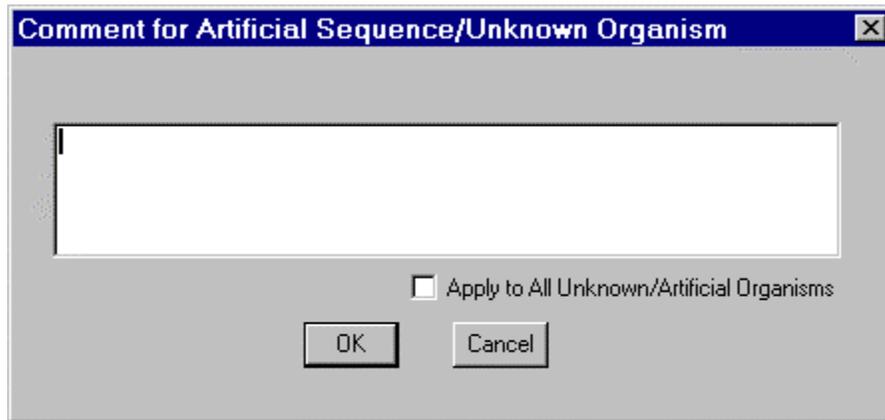
1.  Select the Organism so that the name appears in the “Please enter a custom organism name:” box.
2.  Click on the **OK** button.

**To apply a Custom Organism to all of the Sequences in the Project:**

1. ☞ Click the check box marked: **Apply to all Sequences in the Project**.
2. ☞ Click on the **OK** button.

**5.15 ARTIFICIAL SEQUENCE OR UNKNOWN ORGANISM**

An Artificial Sequence or Unknown organism must have a comment about the organism. After either Artificial Sequence or Unknown is entered into the Organism Name field and the user has moved to another field, an automatic pop-up box will appear.



**Figure 5-18: Artificial Sequence/Unknown Organism Comment**

1. ☞ Select the **Standard** button from the Sequence Screen (Figure 5-1).
2. ☞ Select either Unknown or Artificial from the list of standard organisms.
3. Tab or move to another cell. The comment box, Figure 5-18, will automatically pop-up when a definition of the organism is required.
4. ☞ Enter an appropriate comment for the organism. This information will be placed in a <223> field when the sequence listing is generated.
5. ☞ Check the **Apply to All Unknown/Artificial Organisms** box if the comment is to be applied to all of the Artificial Sequence Organisms and/or Unknown Organisms that have no current comment. The way to use this function is to input all the sequences, label the appropriate ones artificial and/or unknown, then apply the comment to one of the sequences with the box checked. The same comment will then appear for all the artificial sequences in the project.

**i** NOTE: This field can also be accessed by choosing Application Steps | Artificial Sequence/Unknown Comment (Figure 4-1).

**i** NOTE: The content of this box should be as descriptive as possible while remaining as terse as possible.

**i** NOTE: The <223> section of the sequence listing is updated by using either an Artificial Sequence/Unknown Organism or a misc\_feature. The <223> section is the comment or other

information, respectively. PatentIn 3.2 generates “Synthetic Construct” for the <223> section in the protein sequence generated from any coding region in a nucleotide sequence where Artificial or Unknown is used as the organism.

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 6**  
**FEATURE DATA**

## SECTION 6 FEATURE DATA

### 6.1 SEQUENCE FEATURES

The Features Screen (Figure 6-1) enables you to create and modify features pertaining to a sequence. You can access this screen by selecting **Feature Data** from the Application Steps Menu (Figure 4-1), or by selecting the **Fea** button on the PatentIn toolbar. The features that are displayed apply to the sequence that is currently selected on the Sequence Screen.

PatentIn 3.2: Features - Demo Sequence

Sequence Type: DNA      Sequence String Length: 9

Edit Feature

Join All CDSs

Feature Name / Key:  Names

Relevant Residues From:  To:

Other Information:

Clear

(If you have "n" or "X" in the sequence, please define them in the Other Info. field.)

Feature List:

Item	Feature Name	From Column	To Column
------	--------------	-------------	-----------

Insert

Replace

Delete

Validate    Save Project    OK    Cancel    Help

**Figure 6-1: Features Screen**

#### To enter information about a Feature:

1. Check the **Join All CDSs** box if the sequence contains more than one CDS and the CDSs are to be joined.
2. Click on the **Names** button to access the list of Nucleotide Names for **Feature Name/Key**.

3.  Enter the “**Relevant Residue From**” and “**To**” sequence position numbers.
4.  Click in the **Other Information** box to provide other information. This is where you would document X in a protein sequence or n in a base sequence. The Feature Name/Key should be misc\_feature.
5. To clear the **Edit Feature** portion of the screen  Click on the **Clear** button.
6. To insert an entry from the **Feature List**, highlight it and  click on the **Insert** button.
7. To replace an entry from the **Feature List**, highlight it and  click on the **Replace** button.
8. To delete an entry from the **Feature List**, highlight it and  click on the **Delete** button.
9. To validate the information entered,  click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.
10. To save the information,  click on the **Save Project** button.
11. To validate and close,  click on the **OK** button.
12. To cancel the information,  click on the **Cancel** button.
13. To access the help information,  click on the **Help** button.

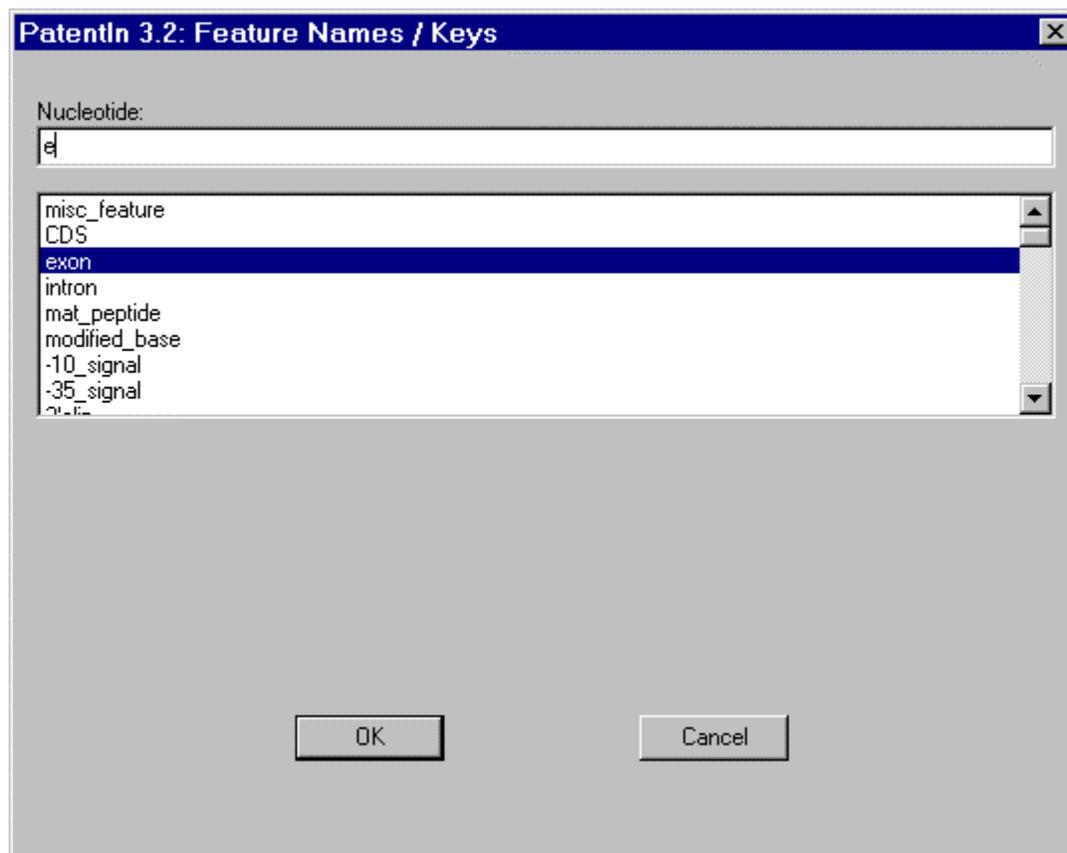
 **NOTE:** The **Join All CDSs** disappears and appears based on whether there is more than one CDS in the Feature List.

 **NOTE:** The <223> section of the sequence listing is updated by using either an Artificial Sequence/Unknown Organism or a misc\_feature. The <223> section is the comment or other information, respectively.

 **NEW TO PATENTIN:** In PatentIn 3.2 there is an automatic expansion of the possible resolutions for “Xaa.”

### 6.1.1 Feature Key Selection

The Feature Names/Key Selection Screen (Figure 6-2) allows the user to select a Nucleotide name.



**Figure 6-2: Feature Names/Key Selection Screen**

**To select a Nucleotide name:**

1.  Select the **Names** button on the Features Screen (Figure 6-1).
2. Begin typing on the **Nucleotide** field, or
3.  Select a feature name from the list by clicking on the feature name or using up/down arrow keys.
4.  Click on **OK** to accept the selection and return to the Features Screen (Figure 6-1).

### 6.1.2 Additional Information Required for Modified\_Base

The Feature Names/Key Selection Screen (Figure 6-3) automatically reveals an additional window when modified\_base is selected.

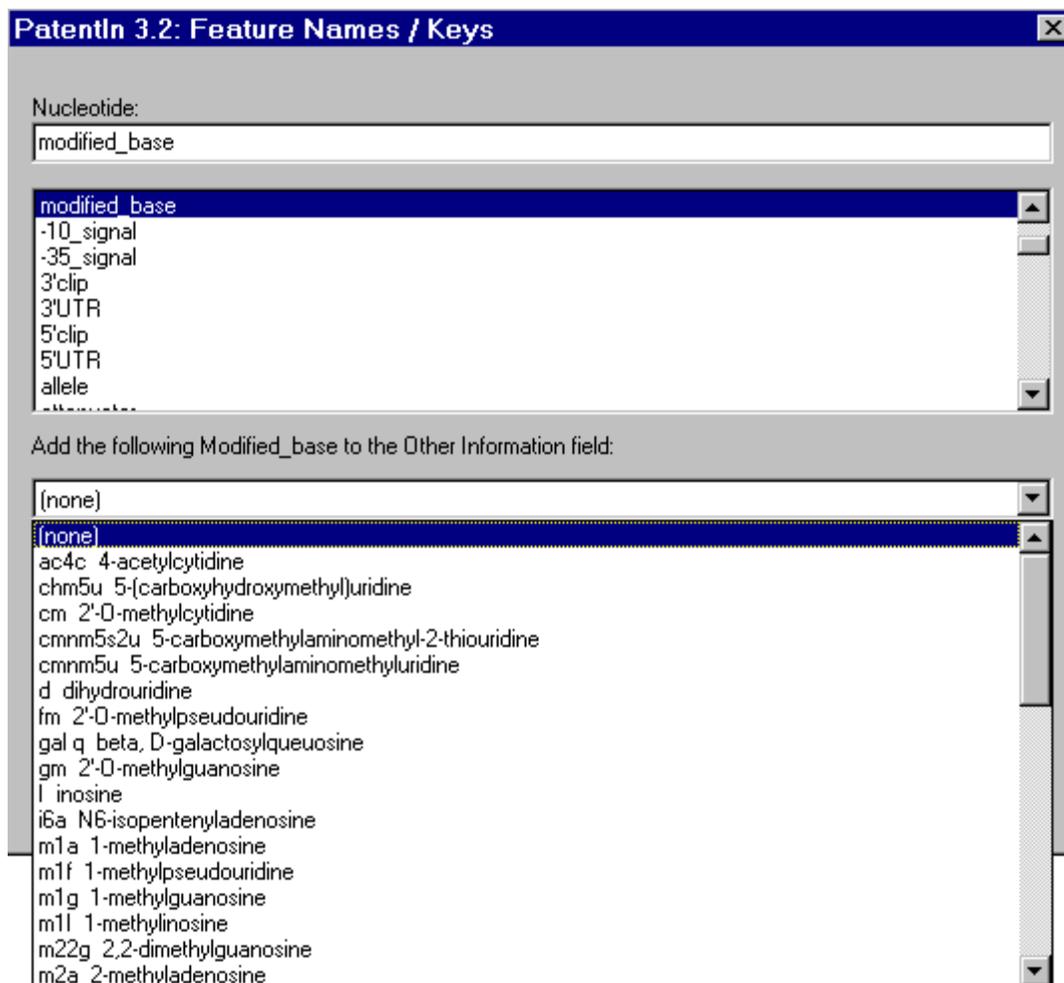


Figure 6-3: Feature Names/Key Selection Screen with Modified Base

#### To add information about the modified\_base

1. Select the pull down arrow on the box marked “Add the following modified\_base to the Other Information.”
2. Select modified\_base from the list (Figure 6-3).
3. Click on **OK** (Figure 6-2).

### 6.1.3 Additional Information on CDS

When a Coding Sequence is specified for a polynucleotide sequence, the DNA sequence will appear in “mixed” format with the DNA split up into codons and the appropriate amino acid beneath each codon. This is exactly what specification of the “exon” feature will do. Selection of

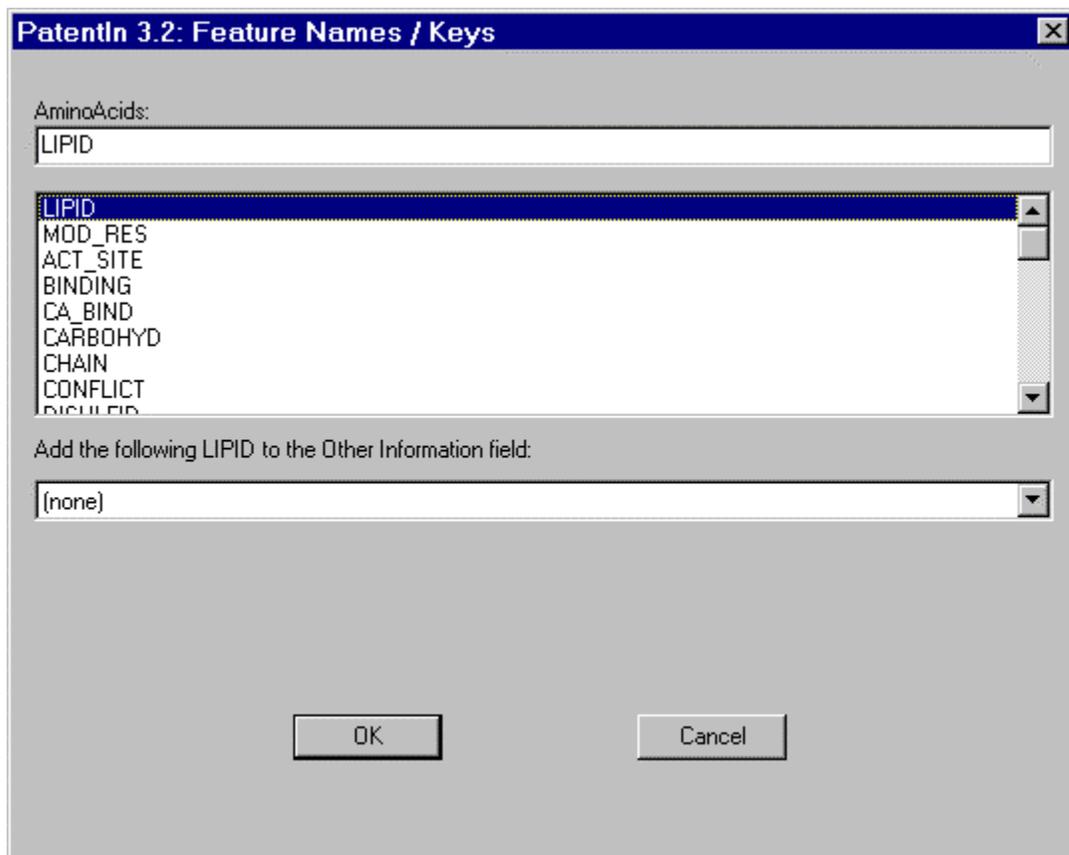
CDS, however, forces PatentIn 3.2 to automatically generate the polypeptide sequence as a “supplemental” sequence.

#### 6.1.4 Further Definition of “n” or “Xaa”

If the variable character “n” appears in a polynucleotide sequence or the variable “Xaa” appears in a polypeptide sequence, ST.25 requires further definition. This is to be provided in the Other Information field using Misc\_Feature. PatentIn will copy the definition of “n” into the supplemental polypeptide sequence and translate it to “Xaa.”

#### 6.1.5 Selecting an Amino Acid

The Feature Names/Key Selection Screen (Figure 6-4) automatically reveals an additional window when LIPID is selected.



**Figure 6-4: Feature Names/Key Selection Screen with LIPID Selected**

**To select an Amino Acid name:**

1.  Select the **Names** button on the Features Screen (Figure 6–1).
2. Begin typing on the **Amino Acid** field, or

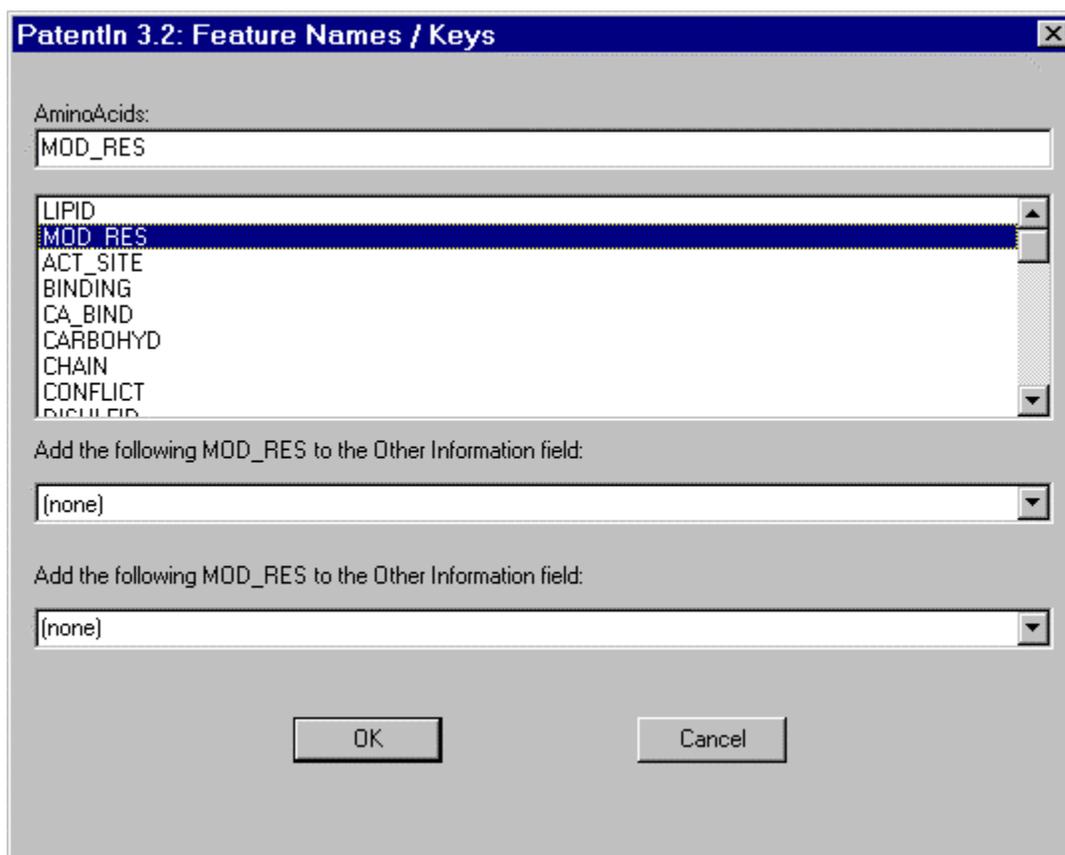
3. Click on the down arrow to open the drop down list (Figure 6-5); select a name from the list.
4. Click on **OK** to accept the selection and return to the Features Screen, (Figure 6-4).

#### To add information about a LIPID:

1. Click on the pull down arrow on the box marked **Add the following LIPID to the Other Information**.
2. Click on the LIPID information from the list.
3. Click on **OK** (Figure 6-4).

#### 6.1.6 Additional Information for MOD\_RES

The Feature Names/Key Selection Screen (Figure 6-5) automatically reveals an additional two windows when MOD\_RES is selected.



**Figure 6-5: Feature Names/Key Selection Screen with MOD\_RES Selected**

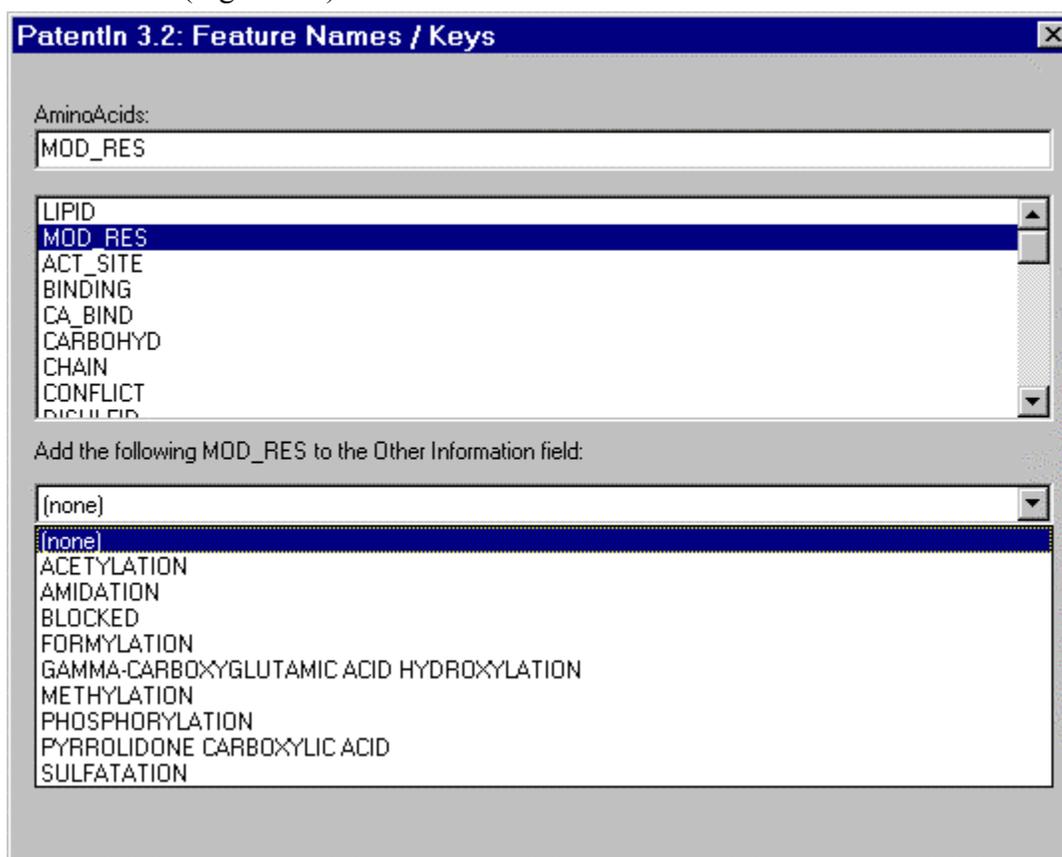
#### To select an Amino Acid name:

1. Click on the **Names** button on the Features Screen (Figure 6-1).
2. Begin typing on the **Amino Acid** field, or

3. Click on the down arrow to open the drop down list (Figure 6-2); select a name from the list.
4. Click on **OK** to accept the selection and return to the Features Screen.

### To add information about the MOD\_RES

1. Select the pull down arrow on the first box in the “Add the following MOD\_RES to the Other Information field.”
2. Select the appropriate information from the list (Figure 6-6).
3. Click on **OK** (Figure 6-5).



**Figure 6-6: First MOD\_RES Pull Down List**

4. Select the pull down arrow on the second box in the “Add the following MOD\_RES to the Other Information field.”
5. Select the appropriate information from the list (Figure 6-7).
6. Click on **OK** (Figure 6-5).

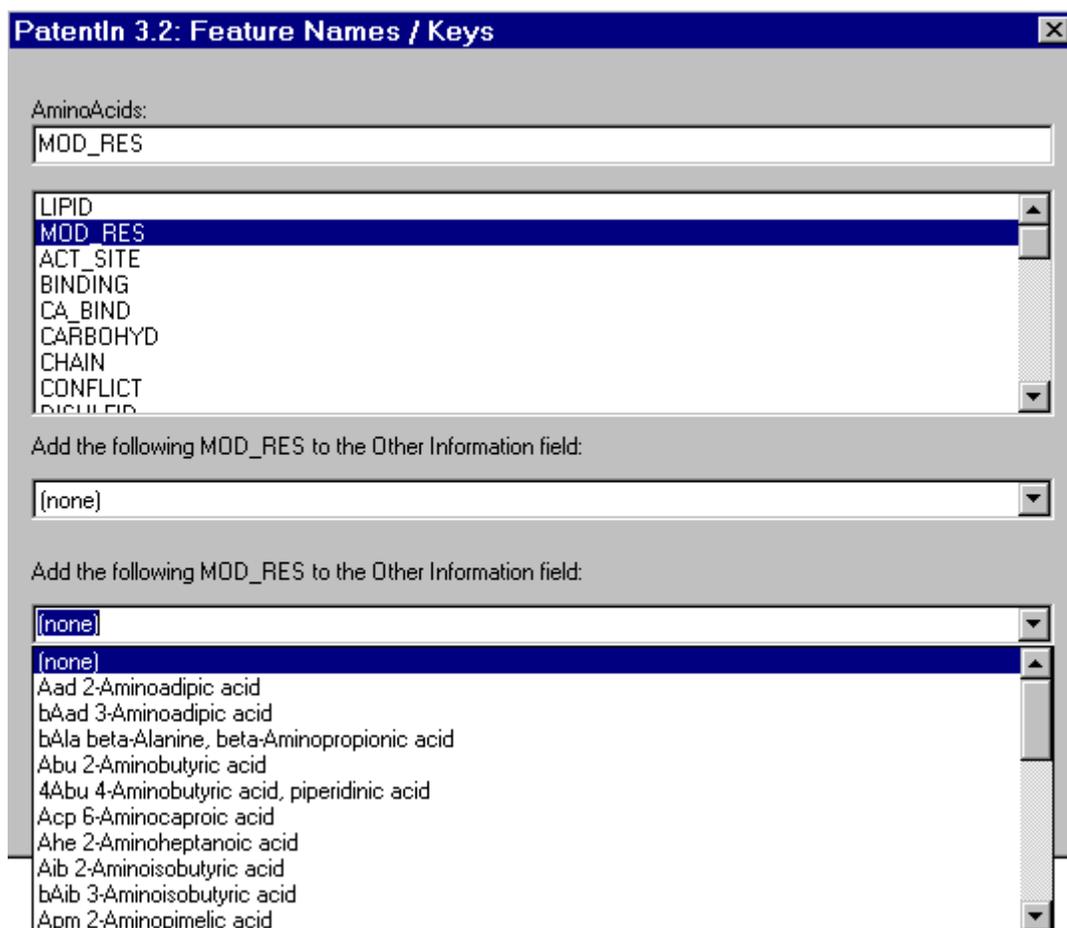


Figure 6-7: Second MOD\_RES Pull Down List

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 7**  
**PUBLICATION DATA**

## SECTION 7 PUBLICATION DATA

### 7.1 PUBLICATION TYPE SCREEN

The Publication Type Screen (Figure 7-1) provides access to four screens for entering publication information. They are Journal, Database, Patent, and Thesis. You can access this screen by selecting Publication Data from the Application Steps menu.

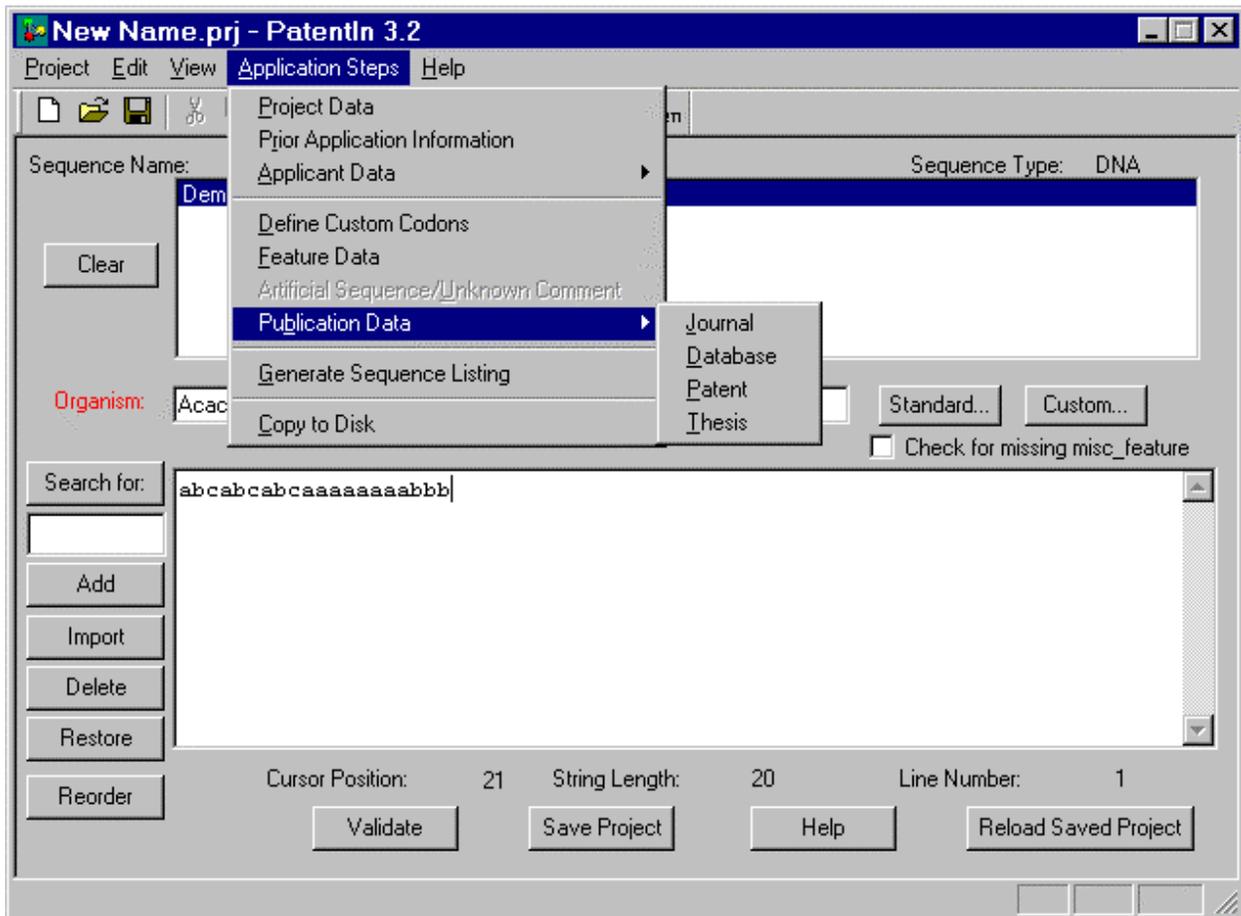


Figure 7-1: Publication Type Screen

To select a Publications Type:

1. From the pull down menu, **Application Steps**, click on **Publication Data**.

## 7.2 JOURNALS PUBLICATION INFORMATION

The Journals Publication Information Screen (Figure 7-2) provides the user with a means to input published supporting scientific literature with the patent application.

PatentIn 3.2: Journals - Demo Sequence

Sequence Type: DNA      Sequence String Length: 20

Journal Publications

Database Name / Accession Number:       Database Entry Date:

Author(s):

Publication Title:

Journal:

Volume:       Issue:

Publication Date:       Page Ranges:

Relevant Residues From:       To:

Journal List:

Item	Publication Title	Accession No.	From	To
------	-------------------	---------------	------	----

**Figure 7-2: Journals Publication Information Screen**

### To enter information about a Journal Publication:

1. From the Application Steps menu, select **Publication Data**, select **Journal**.
2. Enter the **Database Name/Accession Number**.
3. Enter the **Database Entry Date**.
4. Enter the name of the **Author(s)**.
5. Enter the **Publication Title**.
6. Select the **Journal** name from the drop-down pick list. If the name is not on the list you may enter it.
7. Enter the **Volume**.

8.  Enter the **Issue**.
9.  Enter the **Publication Date**.
10.  Enter the **Page Ranges**.
11. In the  Enter the **Relevant Residues From** and **To** sequence position numbers.
12.  Click on the Clear button to clear the **Journal Publications** portion of the screen.
13. To insert an entry from the **Journal List**,  click on the **Insert** button.
14. To replace an entry from the **Journal List**, highlight it, change the entry in the top of the screen and  click on the **Replace** button.
15. To delete an entry from the **Journal List**, highlight it and  click on the **Delete** button.
16. To validate the information entered,  click on **Validate**. Data entered in the table (Insert) is then validated. Information in the **Journal Publications** edit area, that has not yet been inserted, is not validated.
17. To save the information,  click on the **Save Project** button.
18. To validate and close,  click on the **OK** button.
19. To proceed to the next **Publication Data, Database**,  click on the **To Databases**→ button.

### 7.3 DATABASES PUBLICATION INFORMATION

The Databases Publication Information Screen (Figure 7-3) provides the user with a means to input published supporting scientific database with the patent application.

PatentIn 3.2: Databases - Demo Sequence

Sequence Type: DNA      Sequence String Length: 20

Databases

Database Name / Accession Number:       Database Entry Date:

Relevant Residues From:       To:

Clear

Database List:

Item	Accession No.	From	To
------	---------------	------	----

Insert

Replace

Delete

<--To Journals

To Patents-->

Validate      Save Project      OK      Cancel      Help

**Figure 7-3: Databases Publication Information Screen**

#### To enter information about a Database Publication:

1. From the **Application Steps** select **Publications Data**, then select **Database**.
2. Enter the **Database Name/Accession Number**.
3. Enter the **Database Entry Date**.
4. Enter the **Relevant Residues From** and **To** sequence position numbers.
5. Click on the **Clear** button to clear the **Databases** part of the screen.
6. To insert an entry into the **Database List**, and click on the **Insert** button.
7. To replace an entry from the **Database List**, highlight it, change the entry in the top of the screen and click on the **Replace** button.
8. To delete an entry from the **Database List**, highlight it and click on the **Delete** button.

9. To validate the information entered, click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.
10. To save the information, click on the **Save Project** button.
11. To validate and close, click on the **OK** button.
12. To go back to the previous **Publication Data, Journal**, click on the **← To Journals** button.
13. To proceed to the next **Publication Data, Patent**, click on the **To Patents→** button.

#### 7.4 PATENTS PUBLICATIONS INFORMATION

The Patents Publication Information Screen (Figure 7-4) provides the user with a means to input published supporting patent publication information with the patent application.

PatentIn 3.2: Patents - Demo Sequence

Sequence Type: DNA      Sequence String Length: 20

Patents

Database Name / Accession Number:

Database Entry Date:

Document Number:

Filing Date:

Publication Date:

Title:

Relevant Residues From:  To:

Clear

Patent List:

Item	Document No.	Accession No.	From	To
------	--------------	---------------	------	----

Insert

Replace

Delete

<--To Databases

To Theses-->

Validate    Save Project    OK    Cancel    Help

**Figure 7-4: Patents Publication Information Screen**

#### To enter information about a Patent Publication:

1. From the **Application Steps** select **Publications Data**, then select **Patent**.
2. Enter the **Database Name/Accession Number**.

3.  Enter the **Database Entry Date**.
4.  Enter the **Document Number**.
5.  Enter the **Filing Date**.
6.  Enter the **Publication Date**.
7.  Enter the **Title**.
8.  Enter the **Relevant Residues From** and **To** sequence position numbers.
9.  Click on the Clear button to clear the **Patents** portion of the screen.
10. To insert an entry into the **Patent List**,  click on the **Insert** button.
11. To replace an entry from the **Patent List**, highlight it, change the entry in the top of the screen and  click on the **Replace** button.
12. To delete an entry from the **Patent List**, highlight it and  click on the **Delete** button.
13. To validate the information entered,  click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.
14. To save the information,  click on the **Save Project** button.
15. To validate and close,  click on the **OK** button.
16. To go back to the previous **Publication Data, Database**,  click on the **← To Databases** button.
17. To proceed to the next **Publication Data, Thesis**,  click on the **To Theses→** button.

## 7.5 THESES PUBLICATION INFORMATION

The Theses Publications Information Screen (Figure 7-5) provides the user with a means to input published supporting thesis publication information with the patent application.

Sequence Name: DNA      Sequence String Length: 20

Theses

Database Name / Accession Number:       Database Entry Date:

Author(s):

     Title:

Publication Date:       Page Ranges:

Relevant Residues From:       To:

Thesis List:

Item	Title	Accession No.	From	To
------	-------	---------------	------	----

**Figure 7-5: Theses Publications Information Screen**

### To enter information about a Thesis Publication:

1. From the **Application Steps** select **Publications Data**, then select **Thesis**.
2. Enter the **Database Name/Accession Number**.
3. Enter the **Database Entry Date**.
4. Enter the **Author Name**.
5. Enter the **Title**.
6. Enter the **Publication Date**.
7. Enter the **Page Ranges**.
8. Enter the **Relevant Residues From** and **To** sequence position numbers.
9. Click on the **Clear** button to clear the **Theses** portion of the screen.
10. To insert an entry into the **Thesis List**, click on the **Insert** button.

11. To replace an entry from the **Thesis List**, highlight it, change the entry in the top of the screen and  click on the **Replace** button.
12. To delete an entry from the **Thesis List**, highlight it and  click on the **Delete** button.
13. To validate the information entered,  click on **Validate**. Data entered in the table (Insert) is then validated. Information in the edit area, that has not yet been inserted, is not validated.
14. To save the information,  click on the **Save Project** button.
15. To validate and close,  click on the **OK** button.
16. To go back to the previous **Publication Data, Patent**, then  click on the **←To Patents** button.

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**SECTION 8**  
**CREATING A SEQUENCE LISTING PROJECT FILE**

## SECTION 8 CREATING A SEQUENCE LISTING FILE

### 8.1 SEQUENCE LISTING FILE

The sequence listing file includes all the information required by ST.25. PatentIn 3.x will generate a sequence listing with the extension “ST25.txt” appended to the project name. This ST25 file is placed in the directory containing the project.

### 8.2 GENERATING A SEQUENCE LISTING FILE

The Sequence Generation Screen (Figure 8-1) notifies the user that the Generate process is about to occur and gives the option to be notified as errors occur during the process.

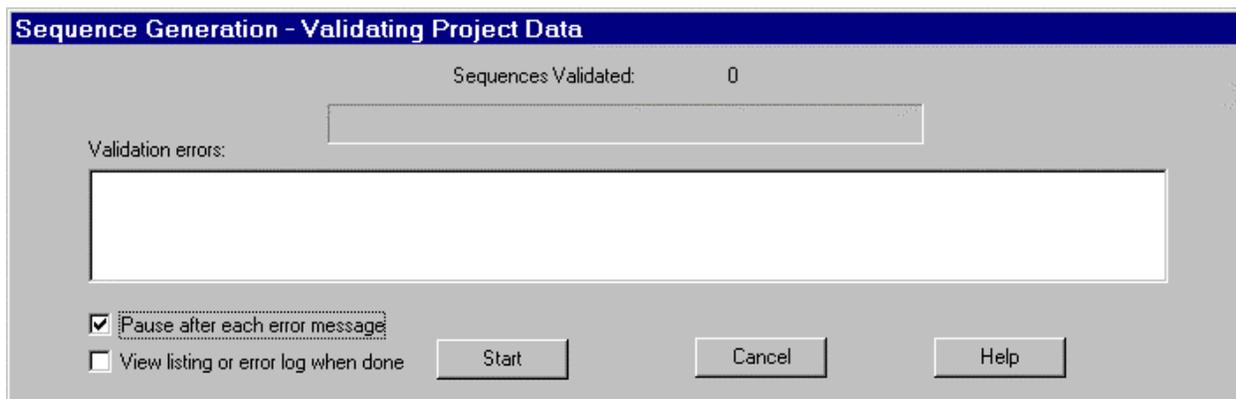
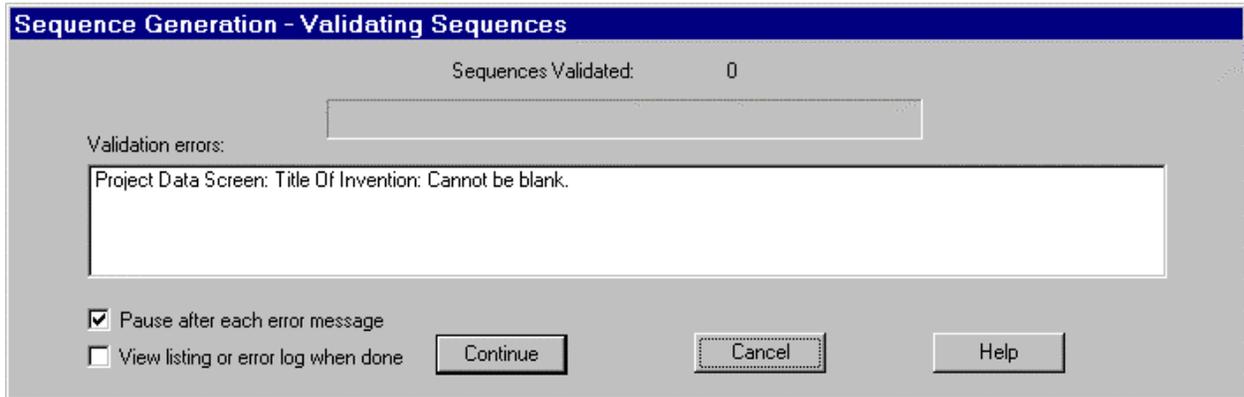


Figure 8-1: Sequence Generation Screen

#### To Generate a Sequence Listing:

1. Select **Generate Sequence Listing** from the **Application Steps** menu (Figure 4-1) or by selecting the **Gen** button on the PatentIn toolbar.
2. Click in the box next to “Pause after each error message” if you wish to be notified of an error in the sequence data when it occurs.
3. Click in the box next to “View listing or error log when done” if you wish to see the listing or error log immediately after the generation. The listing/log will be automatically displayed when the generation has terminated.
4. Click on the **Start** the sequence generation.



**Figure 8-2: Second Sequence Generation Screen**

5.  Click on the **Continue** button to continue validation, Figure 8-2.
6.  Click on the **Cancel** button to cancel validation.
7. If an error message is displayed and “Pause after each error message” was selected, an error message will appear and validation will pause.

### 8.3 VIEWING A SEQUENCE LISTING FILE

#### To view a Sequence Listing Project File:

1. If sequence generation succeeded and **View listing or error log when done** was selected, the generated sequence listing will be shown automatically on the View Results Window (Figure 8-3). If **View listing or error log when done** was not selected, you can view the same results by selecting **View Sequence Listing** from the **Project** menu.

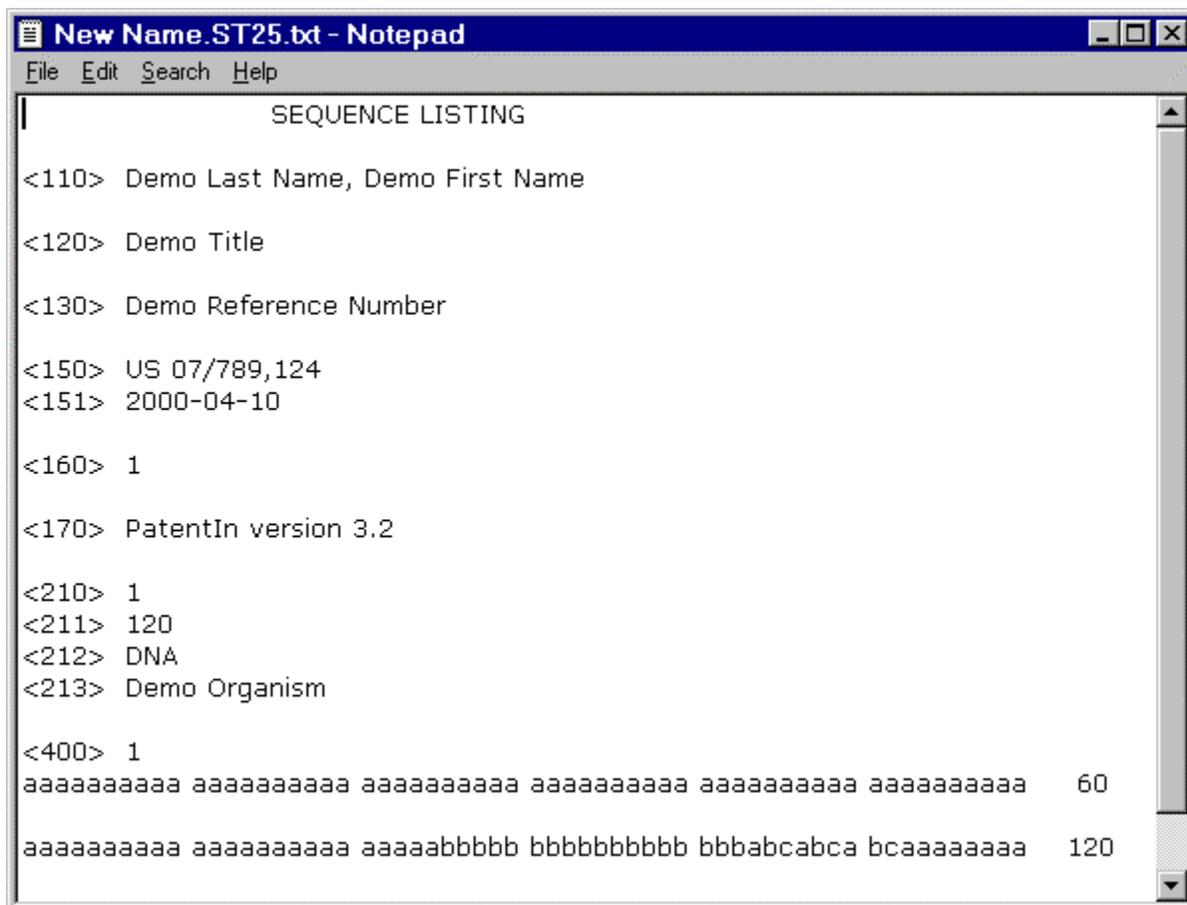
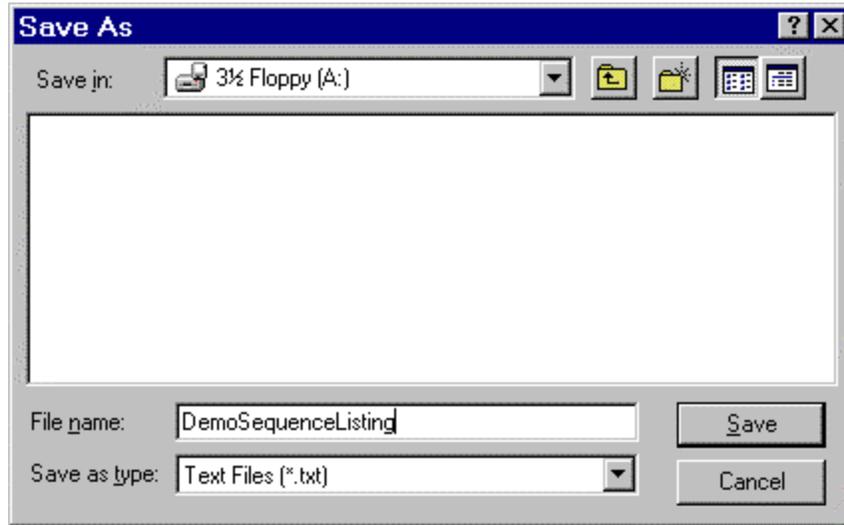


Figure 8-3: View Results Window

**i** SPECIAL NOTE FOR USERS WITH VERY LARGE SEQUENCES AND LARGE NUMBERS OF SEQUENCES: The USPTO has located a viewer that works well for very large text files. A 60-day evaluation version is downloadable at <http://www.fileviewer.com>. The viewer is named “V” and version 2000 SR-1. It was tested with 60MB and 120MB files and “worked great.” It should manage files of almost any size. It was tested on a laptop with Win98 (LocalAdmin may be required for installation). The United States Patent and Trademark Office is not recommending this product but is naming it as an example of the type of product available for this use.

#### 8.4 COPYING THE SEQUENCE LISTING TO A DISK

The Copy to Disk Screen (Figure 8-4) provides the user with the means to name the drive name to where the file is to be copied, the filename for the copied file and the type of copied file.



**Figure 8-4: Copy to Disk Screen**

##### To copy a sequence listing:

1. Select **C**opy to Disk from the **A**pplication Steps menu (Figure 4-1).
2. Select a drive name (Figure 8-4) in the Save in field.
3. Select the type .txt or .zip in the Save as type field.
4. Enter a file name in the File name field.
5. Click on the Save to submit the application.

If **.txt** is selected, PatentIn checks to see if there is enough free disk space to receive the listing file. If so, the file will be copied to the selected location. If not, PatentIn will suggest **.zip**.

NOTE: .zip works only with floppy disk(s) and will format the disk before writing anything to the disk.

NOTE: In general, if a hard drive is selected, the user will be prompted to select a removable medium as the target for this copy.

NOTE: The save to a CD expects the CD to function like any other drive. That is, the user can perform explorer type commands as if the CD were a disk drive (example: F drive). If your CD does not have such a driver, the file can still be located on your hard drive where it was generated. It is the project name with “.ST25.txt” appended to it.

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX A**  
**LIST OF ACRONYMS**

## **APPENDIX A LIST OF ACRONYMS**

ARIPO	African Regional Industrial Property Organization
ASCII	American Standard Code for Information Interchange
CDS	CoDing Sequence
CSC	Computer Sciences Corporation
DLL	Dynamic Link Library
DNA	Deoxyriboneucleic Acid
EPO	European Patent Office (EPO)
FQT	Functional Qualification Test / Functional Quality Testing
GPI	Glycosyl-phosphatidylinositol
LTR	Long Terminal Repeat
MB	Megabytes
MHz	Megahertz
PC	Personal Computer
PCR	Primary Coding Region
PTO	Patent and Trademark Office
RNA	Riboneucleic Acid
scRNA	small cytoplasmic RNA
STS	Sequence Tagged Site
TM02	Task Management Plan
URL	Universal Resource Locator
USPTO	United States Patent and Trademark Office
UTR	Untranslated Region or Untranscribed Region
WIPO	World Intellectual Property Organization
WPI	Web PatentIn
WWW	World Wide Web

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX B**  
**FIELD IDENTIFIERS, LENGTHS AND TYPES**

## APPENDIX B FIELD IDENTIFIERS, LENGTHS AND TYPES

Table B-1, below, includes a complete list of all field names that appear in the data entry screens and their corresponding field identifiers, field lengths, and field types. Field identifiers are used to separate PatentIn data into raw data files and Sequence Listing Project Files.

**Table B-1: Field Name, Identifier, Size, and Type**

<b>Field Identifier</b>	<b>Field Name</b>	<b>Field Length</b>	<b>Field Type</b> A-alpha N- numeric
N/A	Project Name	8	AN
<110>	Applicant Name	1200	AN
<120>	Title of Invention	240	AN
<130>	File Reference	60	AN
<140>	Current Application Number	23	AN
<141>	Current Filing Date	8	N
<150>	Earlier Application Number	23	AN
<151>	Earlier Filing Date	8	N
N/A	Sequence File Name	8	AN
<160>	Number of Sequences	5	N
<170>	Software	60	AN
<210>	Information for SEQ ID No.	5	N
<211>	Length	6	N
<212>	Type	3	A
<213>	Organism	60	AN
<220>	Feature	0	B
<221>	Name/key	20	AN
<222>	Location	12	N
<223>	Other Information	260	AN
<300>	Publication Information	0	B
<301>	Authors	120	AN
<302>	Title (of Publication)	120	AN
<303>	Journal (name)	40	AN

**Table B-1: Field Name, Identifier, Size, and Type (Continued)**

<b>Field Identifier</b>	<b>Field Name</b>	<b>Field Length</b>	<b>Field Type</b> A-alpha N- numeric
<304>	Volume	5	AN
<305>	Issue	5	AN
<306>	Pages	20	AN
<307>	Date	30	AN
<308>	Database Accession Number	45	AN
<309>	Database Entry Date	8	N
<310>	Document Number	18	AN
<311>	Filing Date	8	N
<312>	Publication Date	8	N
<313>	Relevant Residues	20	N
<400>	Sequence Description	100,000	AN

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX C**  
**COUNTRY CODES**

## APPENDIX C COUNTRY CODES

Table C-1 includes a complete list of country codes that are used when completing the **Current Application Number** field in the **Project Data Screen** (Figure 4-2) and the **Prior Application Number** field in the **Prior Application Information Screen** (Figure 4-3).

**Table C-1: Country Codes**

Code	Country
AF	Afghanistan
OA	African Intellectual Property Organization (OAPI)
AP	African Regional Industrial Property Organization (ARIPO)
AL	Albania
DZ	Algeria
AO	Angola
AI	Anguilla
AG	Antigua & Barbuda
AR	Argentina
AU	Australia
AT	Austria
BS	Bahamas
BH	Bahrain
BD	Bangladesh
BB	Barbados
BE	Belgium
BZ	Belize
BX	Benelux Trademark Office and Benelux Designs Office
BJ	Benin
BM	Bermuda
BT	Bhutan
BO	Bolivia

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
BW	Botswana
BR	Brazil
VG	British Virgin Islands
BN	Brunei Darussalam
BG	Bulgaria
BF	Burkina Faso
BU	Burma
BI	Burundi
CM	Cameroon
CA	Canada
CV	Cape Verde
KY	Cayman Islands
CF	Central African Republic
TD	Chad
CL	Chile
CN	China
CO	Columbia
KM	Comoros
CG	Congo
CR	Costa Rica
CI	Cote d'Ivoire
CU	Cuba
CY	Cyprus
CS	Czechoslovakia
KH	Democratic Kampuchea
KP	Democratic People's Republic of Korea
YD	Democratic Yemen
DK	Denmark

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
DJ	Djibouti
DM	Dominica
DO	Dominican Republic
EC	Ecuador
EG	Egypt
SV	El Salvador
GQ	Equatorial Guinea
ET	Ethiopia
EP	European Patent Office (EPO)
FK	Falkland Islands (Malvinas)
FJ	Fiji
FI	Finland
FR	France
GA	Gabon
GM	Gambia
DD	German Democratic Republic
DE	Germany, Federal Republic of
GH	Ghana
GI	Gibraltar
GR	Greece
GD	Grenada
GT	Guatemala
GN	Guinea
GW	Guinea-Bissau
GY	Guyana
HT	Haiti
VA	Holy See
HN	Honduras

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
HK	Hong Kong
HU	Hungary
IS	Iceland
IN	India
ID	Indonesia
IR	Iran (Islamic Republic of)
IQ	Iraq
IE	Ireland
IL	Israel
IT	Italy
JM	Jamaica
JP	Japan
JO	Jordan
KE	Kenya
KI	Kiribati
KW	Kuwait
LA	Laos
LB	Lebanon
LS	Lesotho
LR	Liberia
LY	Libya
LI	Liechtenstein
LU	Luxembourg
MG	Madagascar
MW	Malawi
MY	Malaysia
MV	Maldives
ML	Mali

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
MT	Malta
MR	Mauritania
MU	Mauritius
MX	Mexico
MC	Monaco
MN	Mongolia
MS	Montserrat
MA	Morocco
MZ	Mozambique
NR	Nauru
NP	Nepal
NL	Netherlands
AN	Netherlands Antilles
NZ	New Zealand
NI	Nicaragua
NE	Niger
NG	Nigeria
NO	Norway
OM	Oman
PK	Pakistan
PA	Panama
PG	Papua New Guinea
PY	Paraguay
PE	Peru
PH	Philippines
PL	Poland
PT	Portugal
QA	Qatar

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
KR	Republic of Korea
RO	Romania
RW	Rwanda
KN	Saint Christopher & Nevis
SH	Saint Helena
LC	Saint Lucia
VC	Saint Vincent & the Grenadines
WS	Samoa
SM	San Marino
ST	Sao Tome & Principe
SA	Saudi Arabia
SN	Senegal
SC	Seychelles
SL	Sierra Leone
SG	Singapore
SB	Solomon Islands
SO	Somalia
ZA	South Africa
SU	Soviet Union
ES	Spain
LK	Sri Lanka
SD	Sudan
SR	Suriname
SZ	Swaziland
SE	Sweden
CH	Switzerland
SY	Syria
TW	Taiwan, Province of China

**Table C-1: Country Codes (Continued)**

<b>Code</b>	<b>Country</b>
TH	Thailand
TG	Togo
TO	Tonga
TT	Trinidad & Tobago
TN	Tunisia
TR	Turkey
TV	Tuvalu
UG	Uganda
AE	United Arab Emirates
GB	United Kingdom
TZ	United Republic of Tanzania
US	United States of America
UY	Uruguay
VU	Vanuatu
VE	Venezuela
VN	Viet Nam
WO	World Intellectual Property Organization (WIPO)
YE	Yemen
YU	Yugoslavia
ZR	Zaire
ZM	Zambia
ZW	Zimbabwe

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX D**  
**CONVERSION TABLE BETWEEN NUCLEOTIDE TRIPLETS (CODONS)**  
**AND ONE- AND THREE-LETTER AMINO ACID CODES**

**APPENDIX D**  
**CONVERSION TABLE BETWEEN NUCLEOTIDE TRIPLETS (CODONS)**  
**AND ONE- AND THREE-LETTER AMINO ACID CODES**

Table D-1 provides a list of acceptable characters to be used for hand keying PRT/1 data into the sequence description field, importing PRT/3 data, and converting PRT/1 data in the sequence description field. The nucleotide triplet equivalent data is utilized during sequence listing project file generation when all CDS-featured codons (nucleotide triplets) are translated into an amino abbreviated name (PRT/3).

**Table D-1: Exchange Nucleotide Characters and Amino Characters**

<b>PRT/1</b>	<b>PRT/3</b>	<b>Nucleotide Equivalent</b>
A	Ala	gcu, gcc, gca, gcg, gct
R	Arg	cgu, cgc, cga, cgg, cgt, aga, agg
N	Asn	aau, aac, aat
D	Asp	gau, gac, gat
B	Asx	
C	Cys	ugu, ugc, tgt, tgc
Q	Gln	caa, cag
E	Glu	gaa, gag
Z	Glx	
G	Gly	ggu, ggc, gga, ggg, ggt
H	His	cau, cac, cat
I	Ile	auu, auc, aua, att, atc, ata
L	Leu	uua, uug, cuu, cuc, cua, cug, tta, ttg, ctt, ctc, cta, ctg
K	Lys	aaa, aag
M	Met	aug, atg
F	Phe	uuu, uuc, ttt, ttc
P	Pro	ccu, ccc, cca, ccg, cct
S	Ser	ucu, ucc, uca, ucg, tct, tcc, tca, tcg, agu, agc, agt
T	Thr	acu, acc, aca, acg, act

**Table D-1: Exchange Nucleotide Characters and Amino Characters (Continued)**

<b>PRT/1</b>	<b>PRT/3</b>	<b>Nucleotide Equivalent</b>
W	Trp	ugg, tgg
Y	Tyr	uau, uac, tat, tac
V	Val	guu, guc, gua, gug, gtt, gtc, gta, gtg
X	Xaa	any set containing "n"

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX E**  
**NUCLEOTIDE SEQUENCE FEATURES**

## APPENDIX E NUCLEOTIDE SEQUENCE FEATURES

Table E-1 provides a list (in alphanumeric order) of **Nucleotide** sequence features that are displayed in the Features Screen when you have previously selected DNA or RNA as the **Sequence Type** and  clicked on the **Names** button. A list of nucleotides appears in a pick list after  clicking on the down arrow at the end of the **Nucleotide** box. After  clicking on a sequence feature in the pick list, the sequence feature name appears in the **Feature Name/Key** field (<221>).

**Table E-1: Nucleotide Sequence Features**

Key	Description
allele	a related individual or strain contains stable, alternative forms of the same gene which differs from the presented sequence at this location (and perhaps others)
attenuator	(1) region of DNA at which regulation of termination of transcription occurs, which controls the expression of some bacterial operons; (2) sequence segment located between the promoter and the first structural gene that causes partial termination of transcription
C_region	constant region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; includes one or more exons depending on the particular chain
CAAT_signal	CAAT box; part of a conserved sequence located about 75 bp up-stream of the start point of eukaryotic transcription units which may be involved in RNA polymerase binding; consensus=gg (c or t) caatct
CDS	coding sequence; sequence of nucleotides that corresponds with the sequence of amino acids in a protein (location includes stop codon); feature includes amino acid conceptual translation
conflict	independent determinations of the "same" sequence differ at this site or region
D-loop	displacement loop; a region within mitochondrial DNA in which a short stretch of RNA is paired with one strand of DNA, displacing the original partner DNA strand in this region; also used to describe the displacement of a region of one strand of duplex DNA by a single stranded invader in the reaction catalyzed by RecA protein
D-segment	diversity segment of immunoglobulin heavy chain, and T-cell receptor beta chain

**Table E-1: Nucleotide Sequence Features (Continued)**

Key	Description
enhancer	a cis-acting sequence that increases the utilization of (some) eukaryotic promoters, and can function in either orientation and in any location (upstream or downstream) relative to the promoter
exon	region of genome that codes for portion of spliced mRNA; may contain 5'UTR, all CDSs, and 3'UTR
GC_signal	GC box; a conserved GC-rich region located upstream of the start point of eukaryotic transcription units which may occur in multiple copies or in either orientation; consensus=gggagg
gene	region of biological interest identified as a gene and for which a name has been assigned
iDNA	intervening DNA; DNA which is eliminated through any of several kinds of recombination
intron	a segment of DNA that is transcribed, but removed from within the transcript by splicing together the sequences (exons) on either side of it
J_segment	joining segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains
LTR	long terminal repeat, a sequence directly repeated at both ends of a defined sequence, of the sort typically found in retroviruses
mat_peptide	mature peptide or protein coding sequence; coding sequence for the mature or final peptide or protein product following post-translational modification; the location does not include the stop codon (unlike the corresponding CDS)
misc_binding	site in nucleic acid which covalently or non-covalently binds another moiety that cannot be described by any other Binding key (primer_bind or protein_bind)
misc_difference	feature sequence is different from that presented in the entry and cannot be described by any other Difference key (conflict, unsure, old_sequence, mutation, variation, allele, or modified_base)
misc_feature	region of biological interest which cannot be described by any other feature key; a new or rare feature
misc_recomb	site of any generalized, site-specific or replicative recombination event where there is a breakage and reunion of duplex DNA that cannot be described by other recombination keys (iDNA and virion) or qualifiers of source key (/insertion_seq, /transposon, /proviral)

**Table E-1: Nucleotide Sequence Features (Continued)**

Key	Description
misc_RNA	any transcript or RNA product that cannot be defined by other RNA keys (prim_transcript, precursor_RNA, mRNA, 5'clip, 3'clip, 5'UTR, 3'UTR, exon, CDS, sig_peptide, transit_peptide, mat_peptide, intron, polyA_site, rRNA, tRNA, scRNA, and snRNA)
misc_signal	any region containing a signal controlling or altering gene function or expression that cannot be described by other Signal keys (promoter, CAAT_signal, TATA_signal, -35_signal, -10_signal, GC_signal, RBS, polyA_signal, enhancer, attenuator, terminator, and rep_origin)
misc_structure	any secondary or tertiary structure or conformation that cannot be described by other Structure keys (stem_loop and D-loop)
modified_base	the indicated nucleotide is a modified nucleotide and should be substituted for by the indicated molecule (given in the mod_base qualifier value)
mRNA	messenger RNA; includes 5' untranslated region (5'UTR), coding sequences (CDS, exon) and 3' untranslated region (3'UTR)
mutation	a related strain has an abrupt, inheritable change in the sequence at this location
N_region	extra nucleotides inserted between rearranged immunoglobulin segments
old_sequence	the presented sequence revises a previous version of the sequence at this location
polyA_signal	recognition region necessary for endonuclease cleavage of an RNA transcript that is followed by polyadenylation; consensus=aataaa
polyA_site	site on an RNA transcript to which will be added adenine residues by post-transcriptional polyadenylation
precursor_RNA	any RNA species that is not yet the mature RNA product; may include 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)
prim_transcript	primary (initial, unprocessed) transcript; includes 5' clipped region (5'clip), 5' untranslated region (5'UTR), coding sequences (CDS, exon), intervening sequences (intron), 3' untranslated region (3'UTR), and 3' clipped region (3'clip)

**Table E-1: Nucleotide Sequence Features (Continued)**

Key	Description
primer_bind	non-covalent primer binding site for initiation of replication, transcription, or reverse transcription; includes site(s) for synthetic, for example, PCR primer elements
promoter	region on a DNA molecule involved in RNA polymerase binding to initiate transcription
protein_bind	non-covalent protein binding site on nucleic acid
RBS	ribosome binding site
repeat_region	region of genome containing repeating units
repeat_unit	single repeat element
rep_origin	origin of replication; starting site for duplication of nucleic acid to give two identical copies
rRNA	mature ribosomal RNA; the RNA component of the ribonucleoprotein particle (ribosome) which assembles amino acids into proteins
S_region	switch region of immunoglobulin heavy chains; involved in the rearrangement of heavy chain DNA leading to the expression of a different immunoglobulin class from the same B-cell
satellite	many tandem repeats (identical or related) of a short basic repeating unit; many have a base composition or other property different from the genome average that allows them to be separated from the bulk (main band) genomic DNA
scRNA	small cytoplasmic RNA; any one of several small cytoplasmic RNA molecules present in the cytoplasm and (sometimes) nucleus of a eukaryote
sig_peptide	signal peptide coding sequence; coding sequence for an N-terminal domain of a secreted protein; this domain is involved in attaching nascent polypeptide to the membrane; leader sequence
snRNA	small nuclear RNA; any one of many small RNA species confined to the nucleus; several of the snRNAs are involved in splicing or other RNA processing reactions
source	identifies the biological source of the specified span of the sequence; this key is mandatory; every entry will have, as a minimum, a single source key spanning the entire sequence; more than one source key per sequence is permissible
stem_loop	hairpin; a double-helical region formed by base-pairing between adjacent (inverted) complementary sequences in a single strand of RNA or DNA

**Table E-1: Nucleotide Sequence Features (Continued)**

Key	Description
	(inverted) complementary sequences in a single strand of RNA or DNA
STS	Sequence Tagged Site; short, single-copy DNA sequence that characterizes a mapping landmark on the genome and can be detected by PCR; a region of the genome can be mapped by determining the order of a series of STSs
TATA_signal	TATA box; Goldberg-Hogness box; a conserved AT-rich septamer found about 25 bp before the start point of each eukaryotic RNA polymerase II transcript unit which may be involved in positioning the enzyme for correct initiation; consensus=tata(a or t)a(a or t)
terminator	sequence of DNA located either at the end of the transcript or adjacent to a promoter region that causes RNA polymerase to terminate transcription; may also be site of binding of repressor protein
transit_peptide	transit peptide coding sequence; coding sequence for an N-terminal domain of a nuclear-encoded organellar protein; this domain is involved in post-translational import of the protein into the organelle
tRNA	mature transfer RNA, a small RNA molecule (75-85 bases long) that mediates the translation of a nucleic acid sequence into an amino acid sequence
unsure	author is unsure of exact sequence in this region
V_region	variable region of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for the variable amino terminal portion; can be made up from V_segments, D_segments, N_regions, and J_segments
V_segment	variable segment of immunoglobulin light and heavy chains, and T-cell receptor alpha, beta, and gamma chains; codes for most of the variable region (V_region) and the last few amino acids of the leader peptide
variation	a related strain contains stable mutations from the same gene (for example, RFLPs, polymorphisms, etc.) which differ from the presented sequence at this location (and possibly others)
3'clip	3'-most region of a precursor transcript that is clipped off during processing
3'UTR	region at the 3' end of a mature transcript (following the stop codon) that is not translated into a protein

**Table E-1: Nucleotide Sequence Features (Continued)**

<b>Key</b>	<b>Description</b>
5'UTR	region at the 5' end of a mature transcript (preceding the initiation codon) that is not translated into a protein
-10_signal	pribnow box; a conserved region about 10 bp upstream of the start point of bacterial transcription units which may be involved in binding RNA polymerase; consensus=tataat
-35_signal	a conserved hexamer about 35 bp upstream of the start point of bacterial transcription units; consensus=ttgaca [ ] or tgttgaca [ ]

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX F**  
**AMINO ACID SEQUENCE FEATURES**

## APPENDIX F AMINO ACID SEQUENCE FEATURES

Table F-1 provides a list (in alphabetical order) of **AminoAcids** sequence features that are displayed in the Features Screen when you have previously selected PRT as the **Sequence Type** and  clicked on the **Names** button. The listed sequence features appear in a pick list after  clicking on the down arrow at the end of the **AminoAcids** box. After  clicking on a sequence feature in the pick list, the sequence feature name appears in the **Feature Name/Key** field.

**Table F-1: Amino Acid Sequence Features**

Key	Description
ACT_SITE	amino acid(s) involved in the activity of an enzyme
BINDING	binding site for any chemical group (co-enzyme, prosthetic group, etc.); the chemical nature of the group is given in the description field
CA_BIND	extent of a calcium-binding region
CARBOHYD	glycosylation site; the nature of the carbohydrate (if known) is given in the description field
CHAIN	extent of a polypeptide chain in the mature protein
CONFLICT	different papers report differing sequences
DISULFID	disulfide bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by an intra-chain disulfide bond; if the 'FROM' and 'TO' endpoints are identical, the disulfide bond is an interchain one and the description field indicates the nature of the cross-link
DNA_BIND	extent of a DNA-binding region
DOMAIN	extent of a domain of interest on the sequence; the nature of that domain is given in the description field
HELIX	secondary structure: Helices, for example, Alpha-helix, 3(10) helix, or Pi-helix
INIT_MET	the sequence is known to start with an initiator methionine
LIPID	covalent binding of a lipidic moiety
MAT_PEPTIDE	mature peptide; sequence of the mature or final peptide or protein product following post-translational modification
METAL	binding site for a metal ion; the description field indicates the nature of the metal

**Table F-1: Amino Acid Sequence Features (Continued)**

Key	Description
MISC_FEATURE	region of biological interest which cannot be described by any other feature key; a new or rare feature
MOD_RES	post-translational modification of a residue
MUTAGEN	site which has been experimentally altered
NON_CONS	non consecutive residues; indicates that two residues in a sequence are not consecutive and that there are a number of unsequenced residues between them
NON_TER	the residue at an extremity of the sequence is not the terminal residue; if applied to position 1, this signifies that the first position is not the N-terminus of the complete molecule; if applied to the last position, it signifies that this position is not the C-terminus of the complete molecule; there is no description field for this key
NP_BIND	extent of a nucleotide phosphate binding region; the nature of the nucleotide phosphate is indicated in the description field
PEPTIDE	extent of a released active peptide
PROPEP	extent of a propeptide
REPEAT	extent of an internal sequence repetition
SIGNAL	extent of a signal sequence (prepeptide)
SIMILAR	extent of a similarity with another protein sequence; precise information, relative to that sequence is given in the description field
SITE	any other interesting site on the sequence
STRAND	secondary structure: Beta-strand, for example, Hydrogen bonded beta-strand, or Residue in an isolated beta-bridge
THIOETH	thioether bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thioether bond
THIOLEST	thiolester bond; the 'FROM' and 'TO' endpoints represent the two residues which are linked by the thiolester bond
TRANSIT	extent of a transit peptide (mitochondrial, chloroplastic, or for a microbody)
TRANSMEM	extent of a transmembrane region
TURN	secondary structure Turns, for example, H-bonded turn (3-turn, 4-turn or 5-turn)

**Table F-1: Amino Acid Sequence Features (Continued)**

<b>Key</b>	<b>Description</b>
UNSURE	uncertainties in the sequence; used to describe region(s) of a sequence or which the authors are unsure about the sequence assignment
VARIANT	authors report that sequence variants exist
VARSPLIC	description of sequence variants produced by alternative splicing
ZN_FING	extent of a zinc finger region

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX G**  
**DATA TABLES FOR MOD\_RES SEQUENCE FEATURES**

## APPENDIX G

### DATA TABLES FOR MOD\_RES SEQUENCE FEATURES

Appendix G provides a list (in alphabetical order) of additional modified residue (MOD\_RES) **Sequence Features** that are displayed in the Features Screen when you have previously selected PRT as the **Sequence Type**, and MOD\_RES from the listed sequence features in the pick list. After clicking on a sequence feature in the pick list, **MOD\_RES** appears in the **Feature Name/Key** field (<221>), and the first **Add the following MOD-RES to the Other Information** field for MOD\_RES (Table G-1) and second **Add the following MOD-RES to the Other Information** field for MOD\_RES (Table G-2) sequence features appears. You can select from either one or both of the **Add the following MOD-RES to the Other Information** fields, and the data will appear in the **Other Information** field (<223>).

**Table G-1: First Data Table for MOD\_RES Sequence Features**

Key	Description
(none)	blank space (default option)
ACETYLATION	N-terminal or other
AMIDATION	Generally at the c-terminal of a mature active peptide
BLOCKED	Undetermined n- or c-terminal blocking group
FORMYLATION	Of the n-terminal methionine
GAMMA-CARBOXYGLUTAMIC ACID HYDROXYLATION	Of asparagine, aspartic acid, proline or lysine
METHYLATION	Generally of lysine or arginine
PHOSPHORYLATION	Of serine, threonine, tyrosine, aspartic acid or histidine
PYRROLIDONE CARBOXYLIC ACID	N-terminal glutamate which has formed an internal cyclic lactam
SULFATATION	Generally of tyrosine

**Table G-2: Second Data Table for MOD\_RES Sequences**

Symbol	Meaning
(none)	Blank space (default option)
<b>Aad</b>	<b>2-Amino adipic acid</b>

Symbol	Meaning
<b>bAad</b>	<b>3-Aminoadipic acid</b>
<b>bAla</b>	<b>beta-Alanine, beta-Aminopropionic acid</b>
<b>Abu</b>	<b>2-Aminobutyric acid</b>
<b>4Abu</b>	<b>4-Aminobutyric acid, piperidinic acid</b>
<b>Acp</b>	<b>6-Aminocaproic acid</b>
<b>Ahe</b>	<b>2-Aminoheptanoic acid</b>
<b>Aib</b>	<b>2-Aminoisobutyric acid</b>
<b>bAib</b>	<b>3-Aminoisobutyric acid</b>
<b>Apm</b>	<b>2-Aminopimelic acid</b>
<b>Dbu</b>	<b>2,4 Diaminobutyric acid</b>
<b>Des</b>	<b>Desmosine</b>
<b>Dpm</b>	<b>2,2'-Diaminopimelic acid</b>
<b>Dpr</b>	<b>2,3-Diaminopropionic acid</b>
<b>EtGly</b>	<b>N-Ethylglycine</b>
<b>EtAsn</b>	<b>N-Ethylasparagine</b>
<b>Hyl</b>	<b>Hydroxylysine</b>
<b>aHyl</b>	<b>allo-Hydroxylysine</b>
<b>3Hyp</b>	<b>3-Hydroxyproline</b>
<b>4Hyp</b>	<b>4-Hydroxyproline</b>
<b>Ide</b>	<b>Isodesmosine</b>
<b>alle</b>	<b>allo-Isoleucine</b>
<b>MeGly</b>	<b>N-Methylglycine, sarcosine</b>
<b>MeIle</b>	<b>N-Methylisoleucine</b>
<b>MeLys</b>	<b>6-N-Methyllysine</b>
<b>MeVal</b>	<b>N-Methylvaline</b>
<b>Nva</b>	<b>Norvaline</b>
<b>Nle</b>	<b>Norleucine</b>
<b>Orn</b>	<b>Ornithine</b>

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-0

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX H**  
**ADDITIONAL LIPID SEQUENCE FEATURES**

## APPENDIX H ADDITIONAL LIPID SEQUENCE FEATURES

Table H-1 provides a list (in alphabetical order) of additional lipid sequence features that are displayed in the Features Screen when you have previously selected PRT as the **Sequence Type**, and **LIPID** from the listed sequence features in the pick list. After clicking on a **Sequence Feature** in the **Sequence Feature Pick List**, lipid appears in the **Feature Name/Key** field (<221>) and the **Add the following LIPID to the Other Information** field appears. When you select from the **Add the following LIPID to the Other Information** field, the data appears in the **Other Information** field (<223>).

**Table H-1: Additional Lipid Sequence Features**

Key	Description
(none)	blank space (default option)
MYRISTATE	Myristate group attached through an amide bond to the N-terminal glycine residue of the mature form of a protein or to an internal lysine residue
PALMITATE	Palmitate group attached through a thioether bond to a cysteine residue or through an ester bond to a serine or threonine residue
FARNESYL	Farnesyl group attached through a thioether bond to a cysteine residue
GERANYL-GERANYL	Geranyl-geranyl group attached through a thioether bond to a cysteine residue
GPI-ANCHOR	Glycosyl-phosphatidylinositol (GPI) group linked to the alpha-carboxyl group of the C-terminal residue of the mature form of a protein
N-ACYL DIGLYCERIDE	N-terminal cysteine of the mature form of a prokaryotic lipoprotein with an amide-linked fatty acid and a glyceryl group to which two fatty acids are linked by ester linkages

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX I**  
**ACCEPTABLE CHARACTERS IN THE SEQUENCE DESCRIPTION**  
**FIELD**

**APPENDIX I**  
**ACCEPTABLE CHARACTERS IN THE SEQUENCE DESCRIPTION**  
**FIELD**

Table I-1 provides an acceptable list of characters used as a filter for hand keying or importing DNA or RNA data into the sequence description field. PRT/1 and PRT/3 data lists are used during sequence listing project file generation when all single letter protein data is translated into amino abbreviated name (PRT/3) data.

**Table I-1: Acceptable Characters in the Sequence Description Field**

DNA	RNA	DNA/RNA	Protein/1	Protein/3
a	a	a	A	Ala
g	g	g	C	Cys
c	c	c	D	Asp
t		t	E	Glu
	u	u	F	Phe
r	r	r	G	Gly
y	y	y	H	His
m	m	m	I	Ile
k	k	k	K	Lys
s	s	s	L	Leu
w	w	w	M	Met
b	b	b	N	Asn
d	d	d	P	Pro
h	h	h	Q	Gln
v	v	v	R	Arg
n	n	n	S	Ser
			T	Thr
			V	Val
			W	Trp
			Y	Tyr
			B	Asx
			Z	Glx
			X	Xaa

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX J**  
**ADDITIONAL MODIFIED\_BASE SEQUENCE FEATURES**

## APPENDIX J ADDITIONAL MODIFIED\_BASE SEQUENCE FEATURES

Table J-1 provides a list (in alphabetical order) of additional **modified\_base** sequence features that are displayed in the Features Screen when you have previously selected DNA or RNA as the **Sequence Type**, and **modified\_base** from the listed sequence features in the pick list. After clicking on a sequence feature in the list, **modified\_base** appears in the **Feature Name/Key** field (<221>), and the **Add the following modified\_base to the Other Information** field appears in the Features Screen. The user selects from the pick lists, and the data appears in the **Other Information** field (<223>).

**Table J-1: Modified\_base Sequence Features**

Symbol	Meaning
ac4c	4-acetylcytidine
chm5u	5-(carboxyhydroxymethyl)uridine
cm	2'-O-methylcytidine
cmnm5s2u	5-carboxymethylaminomethyl-2-thiouridine
cmnm5u	5-carboxymethylaminomethyluridine
d	dihydrouridine
fm	2'-O-methylpseudouridine
gal q	beta, D-galactosylqueuosine
gm	2'-O-methylguanosine
I	inosine
i6a	N6-isopentenyladenosine
m1a	1-methyladenosine
m1f	1-methylpseudouridine
m1g	1-methylguanosine
m1I	1-methylinosine
m22g	2,2-dimethylguanosine
m2a	2-methyladenosine
m2g	2-methylguanosine
m3c	3-methylcytidine
m5c	5-methylcytidine

**Table J-1: Modified\_base Sequence Features (Continued)**

Symbol	Meaning
m6a	N6-methyladenosine
m7g	7-methylguanosine
mam5u	5-methylaminomethyluridine
mam5s2u	5-methoxyaminomethyl-2-thiouridine
man q	beta, D-mannosylqueuosine
mcm5s2u	5-methoxycarbonylmethyl-2-thiouridine
mcm5u	5-methoxycarbonylmethyluridine
mo5u	5-methoxyuridine
ms2i6a	2-methylthio-N6-isopentenyladenosine
ms2t6a	N-((9-beta-D-ribofuranosyl-2-methylthiopurine-6-yl)carbamoyl)threonine
mt6a	N-((9-beta-D-ribofuranosylpurine-6-yl)N-methylcarbamoyl)threonine
mv	uridine-5-oxyacetic acid-methylester
o5u	uridine-5-oxyacetic acid
osyw	wybutoxosine
p	pseudouridine
q	queuosine
s2c	2-thiocytidine
s2t	5-methyl-2-thiouridine
s2u	2-thiouridine
s4u	4-thiouridine
t	5-methyluridine
t6a	N-((9-beta-D-ribofuranosylpurine-6-yl)-carbamoyl)threonine
tm	2'-O-methyl-5-methyluridine
um	2'-O-methyluridine
yw	wybutosine
x	3-(3-amino-3-carboxy-propyl)uridine, (acp3)u

October 16, 2003  
Version 2.2  
Task Order Number CSC-04-04

**APPENDIX K**  
**TECHNICAL NOTES**

## **APPENDIX K TECHNICAL NOTES**

### **K.1 MICROSOFT® ACCESS NOTES**

PatentIn 3.2 is packaged with a Microsoft® Access program, patin2xconvert.mdb, that contains links to the standard database files that PatentIn 2.1 uses. In order to enable the user to view the directory, the Link Table Manager must be installed. Similarly, the Data Access files must include dBASE 5 in order to read the PatentIn 2.1 files.

#### **K.1.1 Installing Microsoft® Access 97**

These general instructions are for Office 97.

1. Run Setup from the CD.
2. Select Custom installation. The typical install does not install all of the necessary components.
3. Select Microsoft® Access.
4. Click on the Change Options button.
5. Select Advanced Wizards.
6. Return to main screen.
7. Select Data Access.
8. Click on the Change Options button.
9. Select Database Drivers.
10. Click on the Change Options button.
11. Select dBase & Microsoft® FoxPro Drivers.
12. Return to main screen.
13. Finish installation (or Add).

#### **K.1.2 Hints and Tips**

1. Patin2xconvert.mdb must be in the same directory as the CWPI.exe and must have the help directory in it.
2. Patin2xconvert.mdb refreshes (updates) its links each time an import is done.
3. Patin2xconvert.mdb can be opened independently to verify that all of the Microsoft® Access components are available.
  - a. The Link Table Manager is found in the Tools | Add-Ins in Office 97 and in Database Utilities in Office 2000.

- b. The drivers can be verified by selecting File | Get External Data | Link Tables... . The From File Types should include dBASE 5 (\*.dbf).
4. No relationships are maintained in patin2xconvert.mdb.
5. No code/queries are run from patin2xconvert.mdb.
6. Exploring data from patin2xconvert.mdb, though sometimes necessary, is not recommended since it is possible to update the linked data itself. (Be sure the data is backed-up if this is new to you.)
7. MDAC, a distributable component from Microsoft®, may not be sufficient to run this import.

## **K.2 GENERAL HINTS AND TIPS**

1. The installation requires some Dynamic Link Library (DLL) registration for each machine.
2. PatentIn 3.x uses long-names.
3. In order to reduce the amount of processing of very, very long sequences, paste does not pre-scan (validate) the data.
4. The file submission command, Copy to Disk, copies a file already on your hard drive to an external medium. In general, it cannot be used to copy to your hard drive.
5. Help files are ASCII files than can be locally updated, including translated to a native language.
6. Screens that will not appear or only flicker are usually a symptom of an installation that was not specifically installed on this machine.
7. Windows 2000 defaults to not showing the underlines on the toolbar without pressing the Alt-key.

Although no future enhancement releases are envisioned, maintenance releases may be necessary; however, unreported problems cannot be fixed. Any future enhancements are expected to exclude the “retail” version of Windows 95. Windows 95SP2 will still be supported.

## **K.3 INSTALLATION AND TESTING NOTES**

With the advent of PatentIn 3.2 3 new DLLs are installed in the PatentIn 3.2 directory, normally the user's personal folder for operating systems subsequent to Windows XP and C:\Program Files\USPTO\PatentIn 3.2 otherwise.

DDAO35.DLL

MFC42.DLL

MSVCRT.DLL

Each of these are self-registering. As with any new installation it is recommended that a backup of the system be made first. Should the system network staff wish to unregister these DLLs because of perceived conflicts the commands:

```
regsvr32 -u "C:\Program Files\USPTO\PatentIn 3.2\ddao35.dll"
```

```
regsvr32 -u "C:\Program Files\USPTO\PatentIn 3.2\mfc42.dll"
```

```
regsvr32 -u "C:\Program Files\USPTO\PatentIn 3.2\msvcr.dll"
```

may be executed from Start|Run. (It is the “-u” that causes these statements to be unregister statements.) For the user, PatentIn is designed to be totally uninstalled from the Control Panel Add/Remove processes.

### **K.3.1 Testing configurations:**

During testing the following configurations were used:

Microsoft® Windows 95

Version 4.00.950 B

Internet Explorer: 3.0(4.70.1158)

Microsoft® Windows 98

Version 4.10.2222 A

Internet Explorer 5.00.2614.3500

Microsoft® Windows NT

Version 4.00,1381

Internet Explorer 5.50.4134.0600

Microsoft® Windows 2000 Professional

Version 5.00.2195

Internet Explorer 5.00.3103.1000

Microsoft® Windows ME

This was not formally tested. However, one of our beta testers has informed us that no problems were found with this version.

Microsoft® Windows XP Professional

Version 2002

Internet Explorer 6.0.2600.0000

Although there is no way to test all configurations inclusively, it is believed that PatentIn will not negatively impact any of the above.

### **K.3.2 Internet Explorer Considerations**

Internet Explorer ships with several of the DLLs that either Checker and/or PatentIn use. Although PatentIn has tried to isolate these DLLs for its own use, USPTO has been notified by a

Checker/PatentIn 3.0 user that one or both of these products had a problem with Internet Explorer 4.0 on Microsoft® Windows 98. The user quickly isolated the problem and informed USPTO that an upgrade to Internet Explorer 5.0 eliminated the problem.