User's Manual (UM)
for
PatentIn 3.5.1

Approved by:

Program Manager	Date
Robert Vanni	12/3/2010

Prepared for:
United States Patent and Trademark Office

Prepared by:
Computer Sciences Corporation
2051 Jamieson Avenue - Suite 500
Alexandria, Virginia 22314

Document Identification	
Document Title	User's Manual (UM) for PatentIn 3.5.1
CDRL	PTO-SP-03
USPTO Contract	DOC50PAPT0501005
Task Order Title	PatentIn 3.5.1
Task Order Type	Discrete
Task Order Number	CSCS-07-61

Version Control Log

Version Control Number	Date Change Effective	Nature of Change	Pages Affected	Signature Authority
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 Functional Qualification Test (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
Version 2.3	03/18/04	Updates to PatentIn 3.3 release	All	J. Paccassi
Version 2.4	07/30/04	Updates to PatentIn 3.3 release	5-16	S. Babil
Version 2.5	06/14/06	Updates to PatentIn 3.4 release	All	S. Babil
Version 2.6	10/09/2007	Updates to PatentIn 3.5 release	All	S. Babil
Version 2.7	11/15/2010	Updates to PatentIn 3.5.1	5-2, J-3, J-4	M. Thomas

Executive Summary

This User's Manual (UM) does not conform to the Data Item Description (DID) for the UM, PTO-SP-03. Instead, as required by Section 2.4.4.6 of the Web PatentIn (WPI) Task Management Plan (TM02) version 1.11, it is being updated using the format in which the document was originally prepared.

This UM is being updated at United States Patent and Trademark Office's (USPTO's) direction under task order number CSCS-07-61 to include changes for PatentIn 3.5.1.

Table of Contents

Section 1 Introduction 1-2
1.1 Purpose 1-2
1.2 Conventions 1-2
1.3 Overview 1-2

Section 2 System Requirements and PatentIn 3.5.1 Access 2-1
2.1 System Requirements 2-1
2.2 PatentIn Access 2-1

Section 3 Getting Started 3-1
3.1 Sequence Screen 3-1
3.2 Project Menu 3-1
3.3 Creating and Saving a New Project 3-2
3.4 Opening a Project 3-3
3.5 Saving a Project 3-4
3.6 Viewing a Work File 3-5
3.7 Viewing Work in Progress 3-6
3.8 View a Sequence Listing 3-6
3.9 Viewing Error Reports 3-8
3.10 Rename Sequence 3-8
3.11 Exit PatentIn 3-9
3.12 How to Use Online Help 3-9
3.13 Message Dialog 3-10

Section 4 Project and Application Data 4-1
4.1 Application Steps Menu 4-1
4.2 Project Data 4-1
4.3 Prior Application Information 4-2
4.4 Applicant Data 4-4
4.4.1 Individual Applicants 4-4
4.4.2 Organization Applicants 4-6

Section 5 Sequence Data 5-1
5.1 Sequence 5-1
5.1.1 Selecting a Standard Organism 5-2
5.1.2 Creating Default Explanations for "n”s and "Xaa”s 5-2
5.1.3 Searching for a Sequence 5-3
5.1.4 Clearing the Screen 5-3
5.2 Adding a Sequence 5-3
5.3 Importing a Sequence 5-4
5.3.1 Format for Multi-Sequence Data Files (Non-ST.25 Sequence Listing File) to be Imported by PatentIn 3.5.1 5-5
5.3.2 Format for Single Sequence Data Files to be Imported by PatentIn 3.5.1 5-6
5.3.3 Importing Sequences from a Project 5-7
5.3.4 Importing Sequences from a PatentIn Generated ST.25 Sequence Listing File 5-8

5.4 Copying a Sequence 5-14
5.5 Pasting a Sequence 5-14
5.6 Deleting a Sequence 5-15
5.7 Skip a Sequence 5-15
5.8 Restoring a Sequence 5-15
5.9 Reordering Sequences 5-16
5.10 Changing a Sequence Type 5-18
5.11 Validating Sequences 5-18
5.12 Saving Sequences 5-19
5.13 Reloading a Saved Project 5-19
5.14 Adding Custom Codons 5-19
5.15 Adding a Custom Organism 5-20
5.16 Artificial Sequence or Unknown Organism 5-22

Section 6 Feature Data 6-1

6.1 Sequence Features 6-1
6.1.1 Feature Key Selection 6-2
6.1.2 Additional Information Required for Modified_Base 6-3
6.1.3 Additional Information on CDS (Coding Sequence) 6-4
6.1.4 Further Definition of “N” or “Xaa” 6-4
6.1.5 Selecting an Amino Acid 6-5
6.1.6 Additional Information for MOD_RES 6-6

Section 7 Publication Data 7-1

7.1 Publication Type Screen 7-1
7.2 Journals Publication Information 7-2
7.3 Databases Publication Information 7-3
7.4 Patents Publications Information 7-4
7.5 Theses Publication Information 7-5

Section 8 Creating a Sequence Listing Project File 8-1

8.1 Sequence Listing File 8-1
8.2 Generating a Sequence Listing File 8-1
8.3 Viewing a Sequence Listing File 8-2
8.4 Copying the Sequence Listing to a Disk 8-3

Section 9 List of Acronyms 9-1

Appendix A Field Identifiers, Lengths and Types A-1

Appendix B Country Codes B-1

Appendix C Conversion Table Between Nucleotide Triplets (Codons) and One- and Three Letter Amino Acid Codes C-1

Appendix D Nucleotide Sequence Features D-1

Appendix E Amino Acid Sequence Features E-1

Appendix F Data Tables for Mod_Res Sequence Features F-1

Appendix G Additional Lipid Sequence Features G-1

Appendix H Acceptable Characters In The Sequence Description Field H-1

Appendix I Additional Modified_Base Sequence Features I-1

Appendix J Technical Notes J-1

J.1 Microsoft® Access Notes J-1
J.1.1 Installing Microsoft® Access 97 ..... J-1
J.1.2 Hints and Tips J-1
J.2 General Hints and Tips J-2
J.3 Installation and Testing Notes J-2
J.3.1 PatentIn Installation and Repair J-2
J.3.2 Testing Configurations J-3
J.3.3 Internet Explorer Considerations J-4

List of Figures

Figure 3-1: Sequence Screen 3-1
Figure 3-2: Project Menu 3-2
Figure 3-3: Save As Screen 3-3
Figure 3-4: Open Screen 3-4
Figure 3-5: Save As Screen 3-5
Figure 3-6: View Work in Progress Window 3-6
Figure 3-7: View Sequence Listing Window 3-7
Figure 3-8: View Error Report Window 3-8
Figure 3-9: Rename Sequence Screen 3-8
Figure 3-10: Exit PatentIn Screen 3-9
Figure 3-11: Help Screen 3-9
Figure 3-12: Message Dialog Screen 3-10
Figure 4-1: Application Steps Menu 4-1
Figure 4-2: Project Data Screen 4-2
Figure 4-3: Prior Application Information Screen 4-3
Figure 4-4: Application Data Screen 4-4
Figure 4-5: Individual Applicants Screen 4-5
Figure 4-6: Organization Applicants Screen 4-6
Figure 5-1: Sequence Screen 5-1

Figure 5-2: Selecting an Organism Screen 5-2
Figure 5-3: Add a Sequence Screen 5-3
Figure 5-4: Import Sequence(s) Screen 5-4
Figure 5-5: Sample ASCII Sequence Data 5-6
Figure 5-6: Sequence Type Selection Screen 5-6
Figure 5-7: Sequences Being Imported Screen 5-7
Figure 5-8: Validation Errors Screen 5-7
Figure 5-9: Select Sequences From Project Screen 5-8
Figure 5-10: Import to New Project Message Dialog Screen 5-9
Figure 5-11: Save Project Screen 5-9
Figure 5-12: Parsing Applicant Names Screen 5-10
Figure 5-13: Added All Names Message Dialog Screen 5-11
Figure 5-14: The Skipped Sequence Type Selection Screen 5-12
Figure 5-15: Invalid Sequence Type Encountered Screen 5-12
Figure 5-16: Select Sequences to Import Screen 5-13
Figure 5-17: Edit Menu 5-14
Figure 5-18: Sequence Recovery Screen 5-16
Figure 5-19: Reorder Sequences Screen 5-17
Figure 5-20: Change Sequence Type Screen 5-18
Figure 5-21: Custom Codons Input Screen 5-19
Figure 5-22: Amino Acid Drop Down List Screen 5-20
Figure 5-23: Custom Organism Input Screen 5-21
Figure 5-24: Artificial Sequence/Unknown Organism Comment 5-22
Figure 6-1: Features Screen 6-1
Figure 6-2: Feature Names/Key Selection Screen 6-3
Figure 6-3: Feature Names/Key Selection Screen with Modified Base 6-4
Figure 6-4: Feature Names/Key Selection Screen with LIPID Selected 6-5
Figure 6-5: Feature Names/Key Selection Screen with MOD_RES Selected 6-6
Figure 6-6: First MOD_RES Pull Down List 6-7
Figure 6-7: Second MOD_RES Pull Down List 6-8
Figure 7-1: Publication Type Screen 7-1
Figure 7-2: Journals Publication Information Screen 7-2
Figure 7-3: Databases Publication Information Screen 7-3
Figure 7-4: Patents Publication Information Screen 7-4
Figure 7-5: Theses Publications Information Screen 7-5
Figure 8-1: Sequence Generation Screen 8-1
Figure 8-2: Second Sequence Generation Screen 8-2
Figure 8-3: View Results Window 8-2
Figure 8-4: Copy to Disk Screen 8-3

Figure J-1: PatentIn Repair and Uninstall J-3

List of Tables

Table 1-1: Document Conventions 1-2
Table 5-1: Header for Sequence 5-5
Table 9-1: Acronyms 9-1

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

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.

Table 1-1: Document Conventions

Notation	Represents	Example
Bold type	Name of a function, file, menu item, or programming construct.	Click on **Exit**.

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 United States (US) final rule, "Requirements for Patent Applications Containing Nucleotide Sequence or Amino Acid Disclosures." The application runs on Windows XP/Windows Vista/Windows 7 and the screen displays are in English. Since the sequence listings generated by PatentIn comply with the ST.25, this program has worldwide applicability.

For ease of use, the design follows the 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 PatentIn generated ST.25 sequence listing files or sequence data files created by another editor or word processor (provided they are stored as American Standard Code for Information Interchange (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.

Section 2 System Requirements and PatentIn 3.5.1 Access

2.1 System Requirements

PatentIn is a self-contained application that can be downloaded from the USPTO website. It operates in a Windows XP, Windows Vista and Windows 7 environment. A minimum of 512 Megabytes (MBs) of memory is required. Additional memory may be required for large patent applications. For the best performance 1 Gigabytes (GBs) of memory is recommended for very large projects, projects with 100,000 sequences or a sequence approaching 12 MB. The disk space required to install PatentIn 3.5.1 is about 6.5 MBs. Additional disk space is required to store project files and sequence listing files.

For PatentIn to work correctly, the “TMP” environment variable must point to a valid directory, and the “PATH” environment variable must include the Disc Operating System (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: USPTO has located a viewer that works for very large text files. A 60 day evaluation version is downloadable at www.fileviewer.com/. The viewer is named "V" and the version is 2000 SR-1. USPTO was able to successfully open files of 60 MB and 120 MB using an earlier version of the V software and the viewer was tested on a laptop with Windows 98. (LocalAdmin may be required for installation.)

2.2 PatentIn Access

PatentIn was designed for installation on individual computers. It can be downloaded onto your Personal Computer (PC) and, if desired, the project files can be stored remotely. The program can be downloaded from the USPTO web page, http://www.uspto.gov/web/offices/pac/patin/patentinrel.htm. Before the installation of PatentIn 3.5.1, be sure to close all running applications on your computer and then follow the instructions found on the Web page to download PatentIn 3.5.1 and install it on your PC. Upon completing the installation, an icon will be placed on your desktop. Access to the PatentIn 3.5.1 application program occurs when you double-click on the PatentIn 3.5.1 icon.

Section 3 Getting Started

3.1 Sequence Screen

When you first access PatentIn by double-clicking on the PatentIn 3.5.1 icon on your desktop, you have immediate access to the Sequence Screen.

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

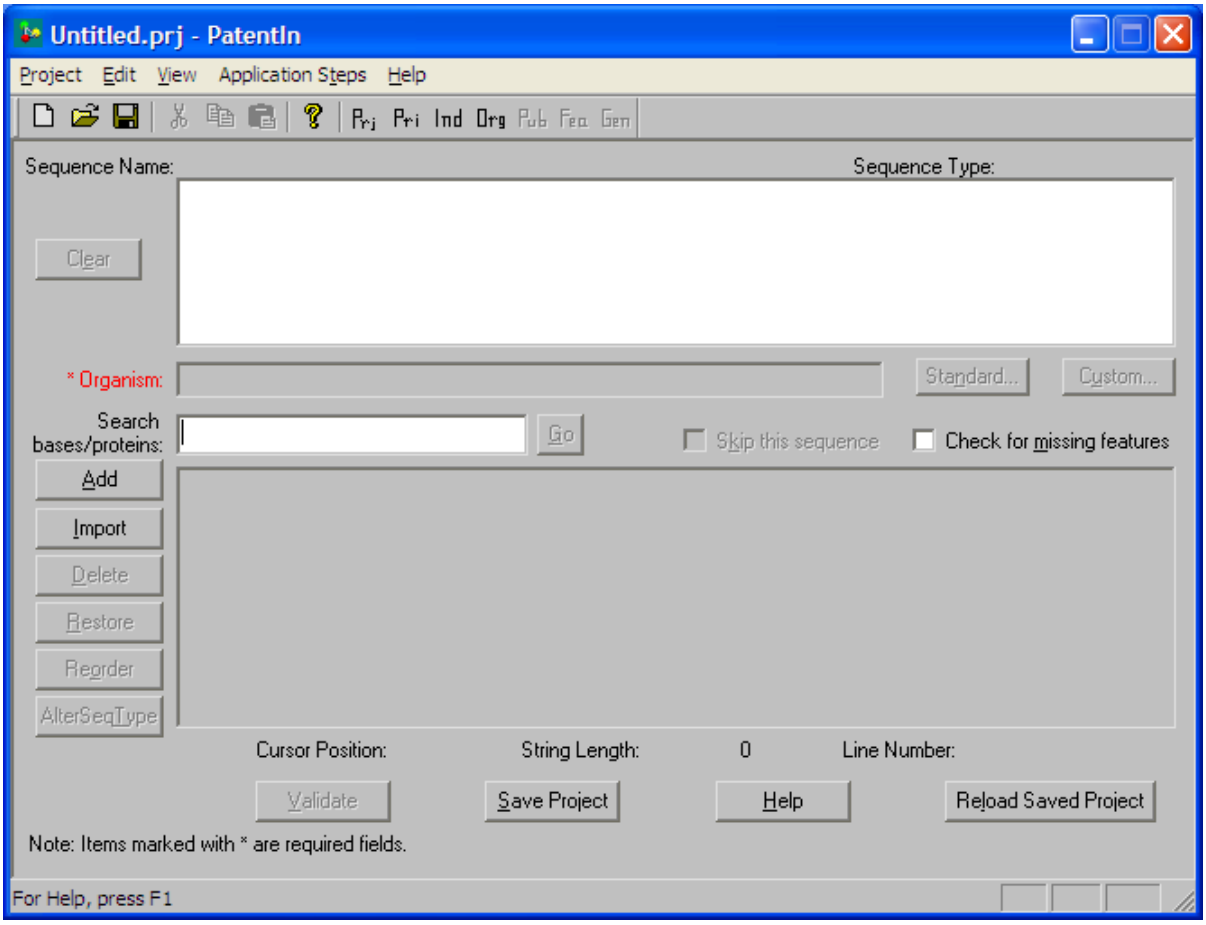

Figure 3-1: Sequence Screen

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 (MS) Windows-type menus. The user may select any one of the three drop-down menus when a project is started. 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 3-5) where a new project file is created and saved. Selecting “**Open**” displays the Open Screen (Figure 3-4) where the user can select a previously saved

project to open. The “Exit PatentIn” 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 selections are shown in Figure 3-2. 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.

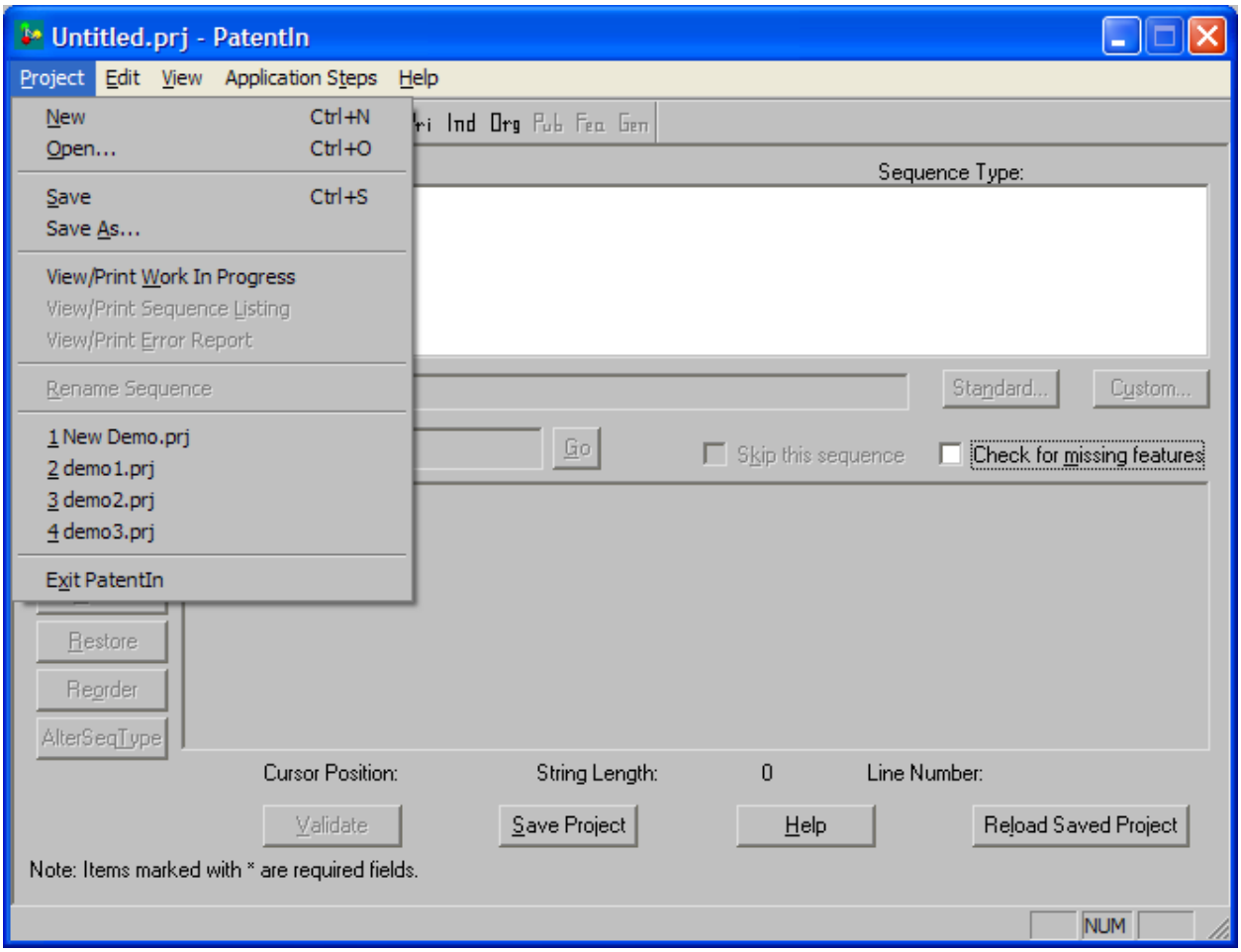

Figure 3-2: Project Menu

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 perform the following steps:

1. Select **New** from the Project Menu. This clears the 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.

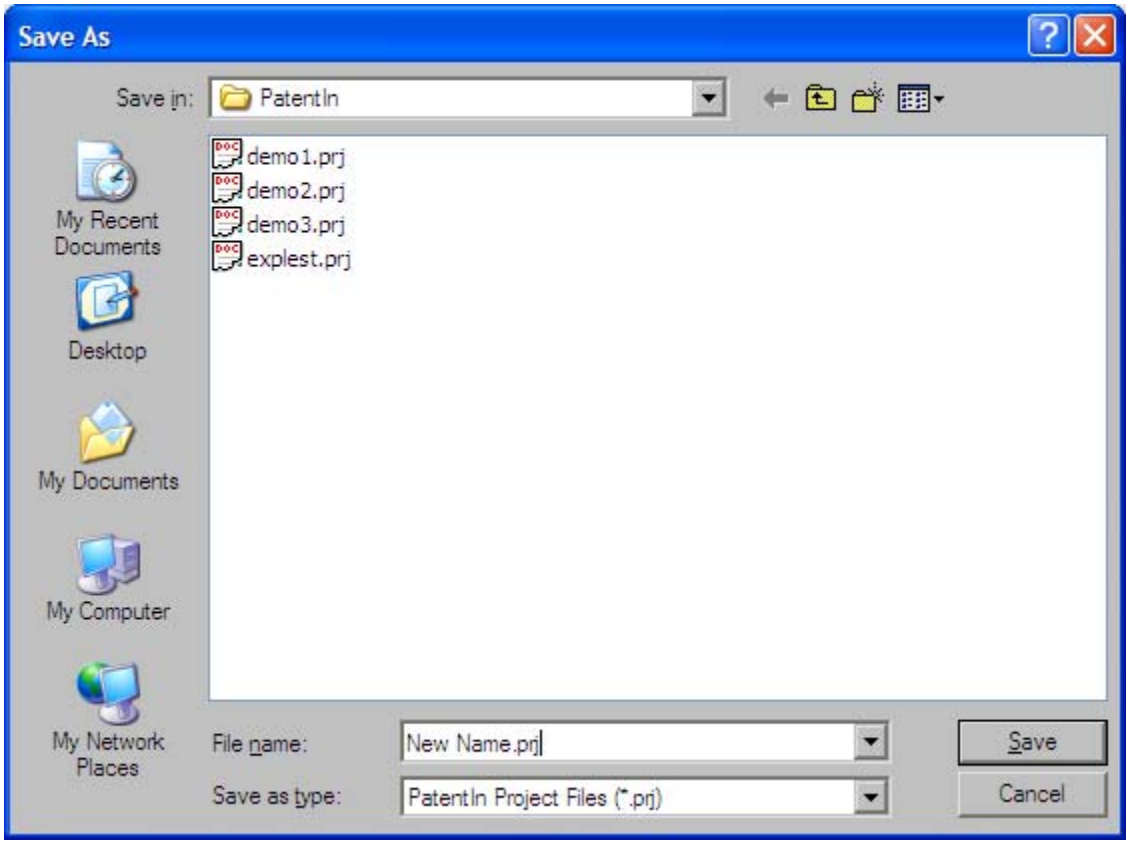

Figure 3-3: Save As Screen

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.

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.

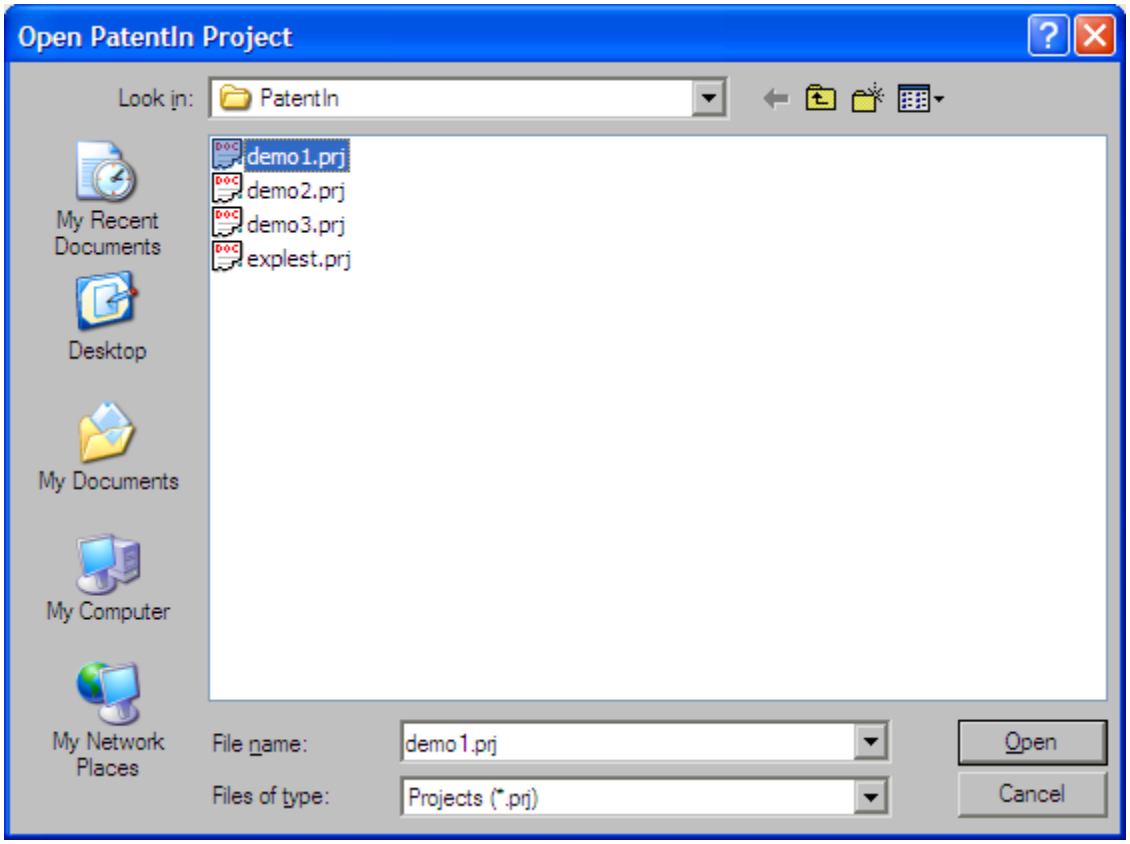

Figure 3-4: Open Screen

Special Note for users with very large sequences 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.

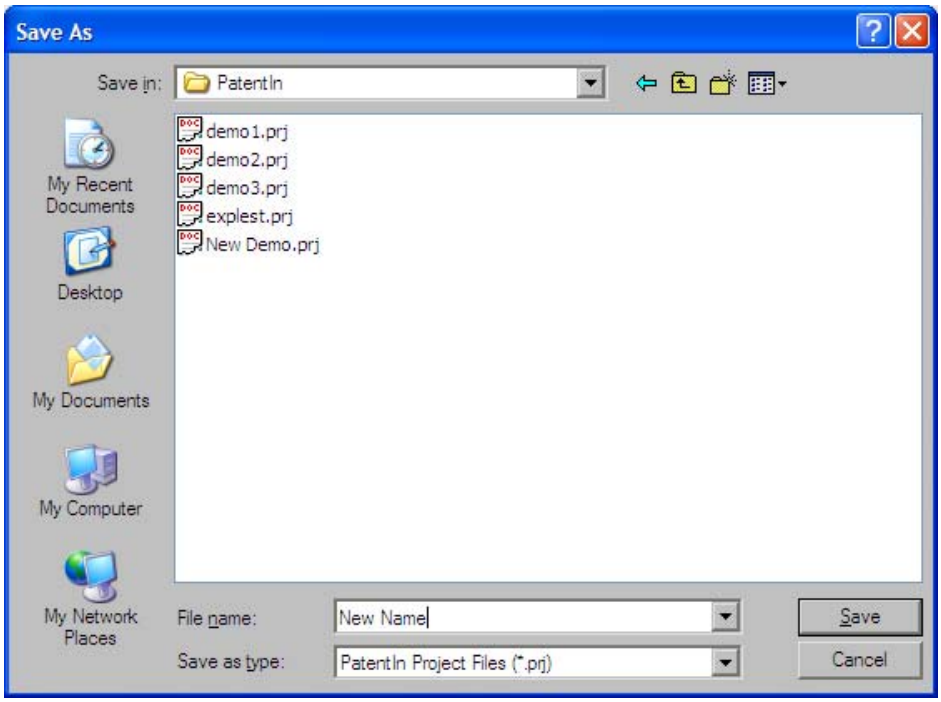

Figure 3-5: Save As Screen

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**.

3.7 Viewing Work in Progress

PatentIn provides the user with an on-screen display of the Patent Application with the, as shown in Figure 3-6. The name of the work file is the name of the current PatentIn project followed by "_WorkFile.txt."

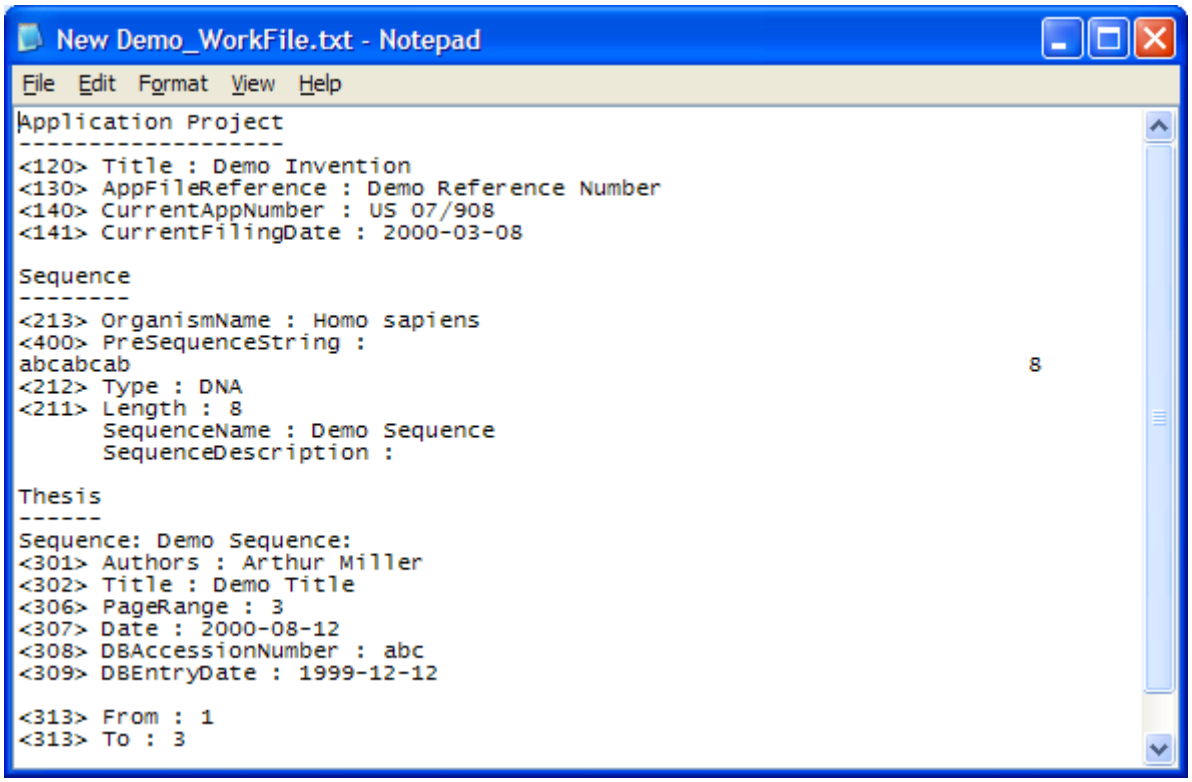

Figure 3-6: View Work in Progress Window

To view the current patent application:

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

3.8 View a Sequence Listing

PatentIn provides the user with an on-screen view of the Sequence Listing with the, as shown in Figure 3-7. The name of the generated listing is the name of the current PatentIn project followed by "_ST25.txt."

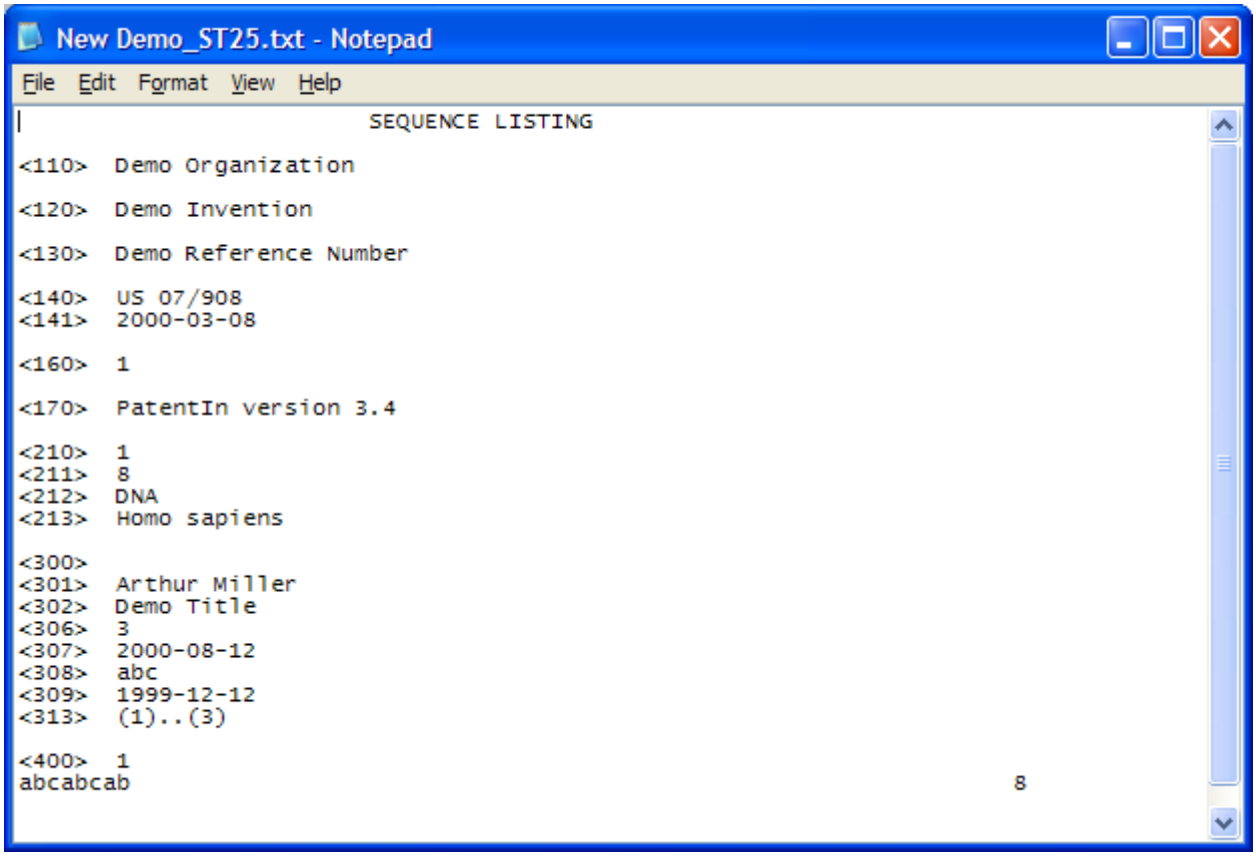

Figure 3-7: View Sequence Listing Window

To view the sequence listing:

1. From the Project menu, select **View/Print Sequence Listing**.
2. To print the report, click on **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 Figure 3-8, if one exists, for the opened project. The name of the error log file is the name of the current PatentIn project followed by "_ErrorLog.txt."

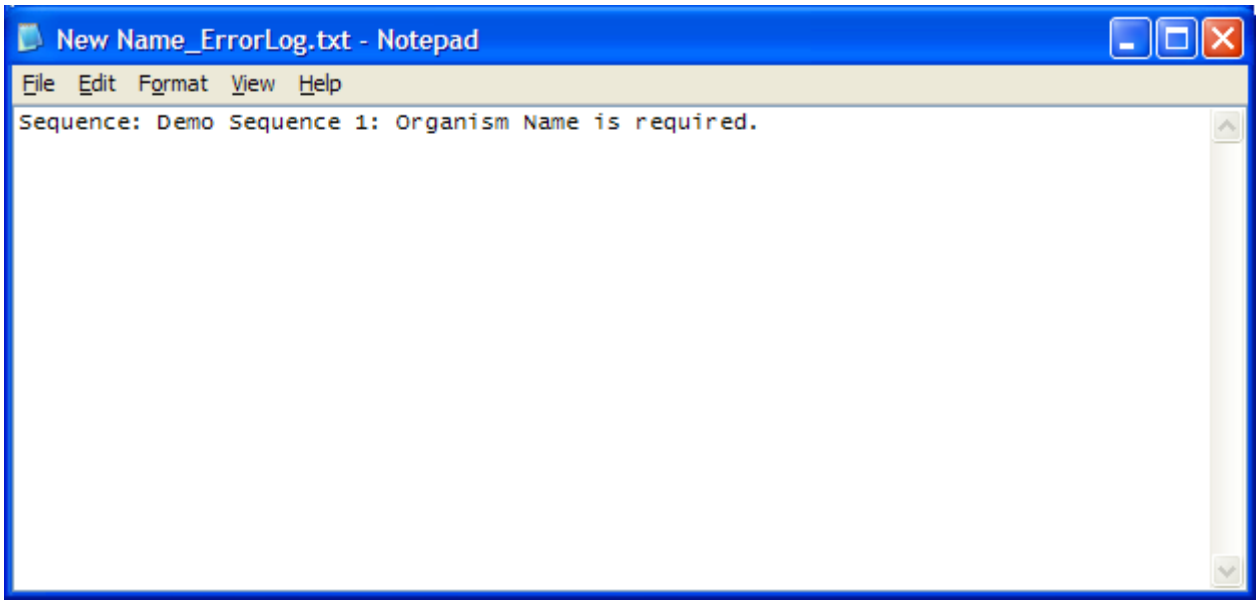

Figure 3-8: View Error Report Window

To view an error report:

1. From the Project menu, select View/Print 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, as shown in Figure 3-9.

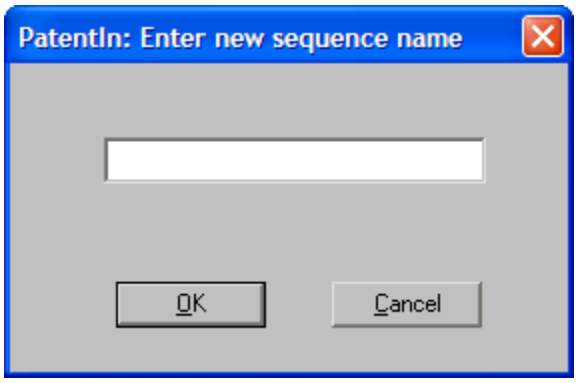

Figure 3-9: Rename Sequence Screen

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 about whether the project should be saved, as shown in Figure 3-10.

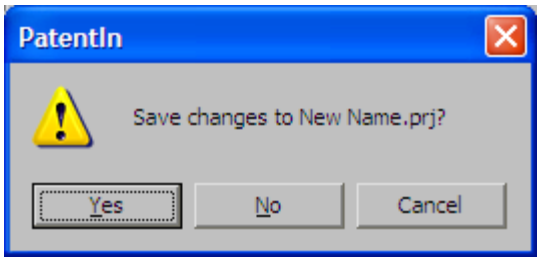

Figure 3-10: Exit PatentIn Screen

3.12 How to Use Online Help

Online help is available for most of the PatentIn screens. 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. To exit the **Help** screen, click on the **OK** button to close caption.

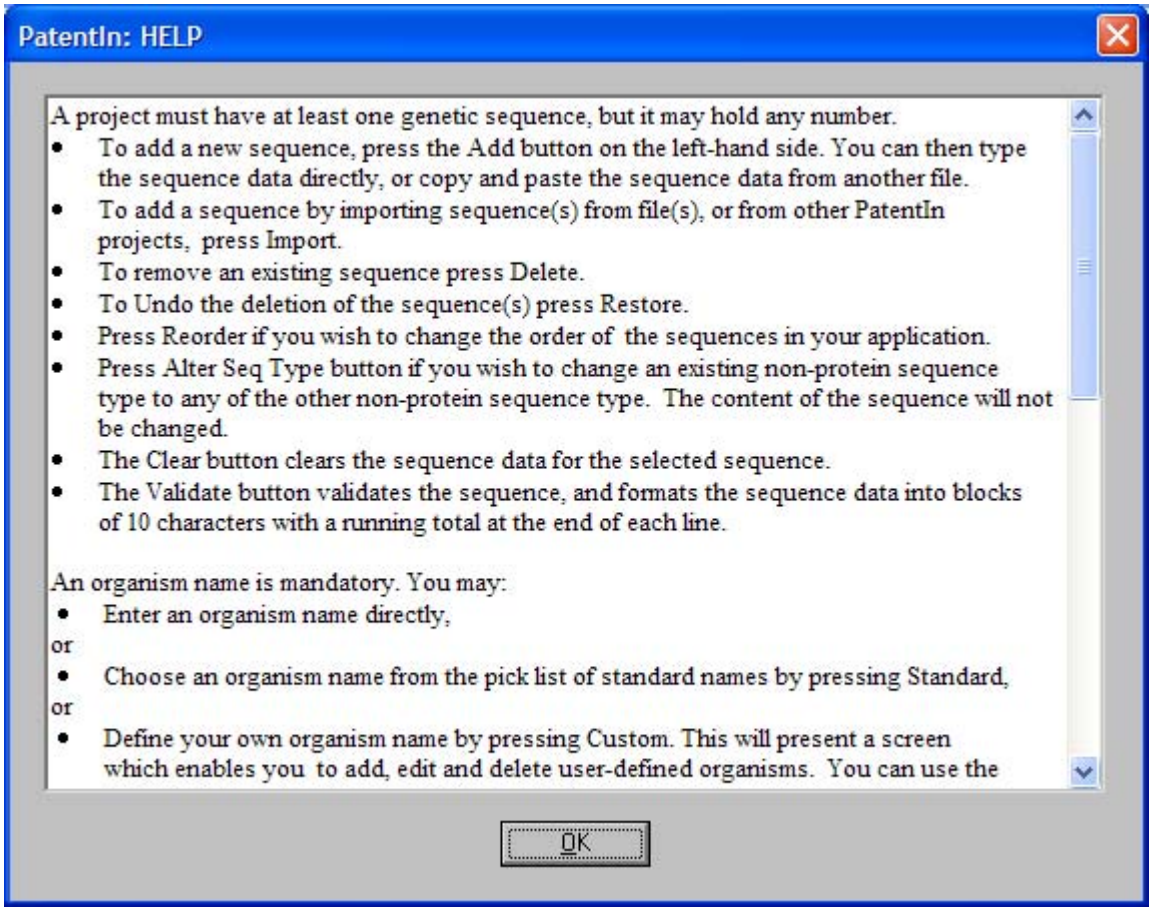

Figure 3-11: Help Screen

3.13 Message Dialog

The 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.

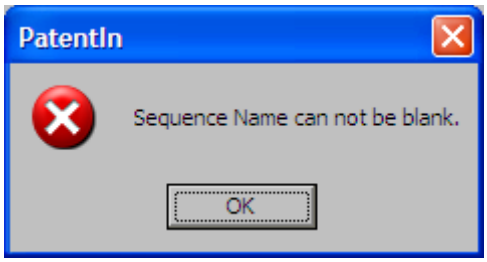

Figure 3-12: Message Dialog Screen

Section 4 Project and Application 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 started. The project name is visible at the upper left corner of the screen. In this example the user has opened an existing project called New Name.

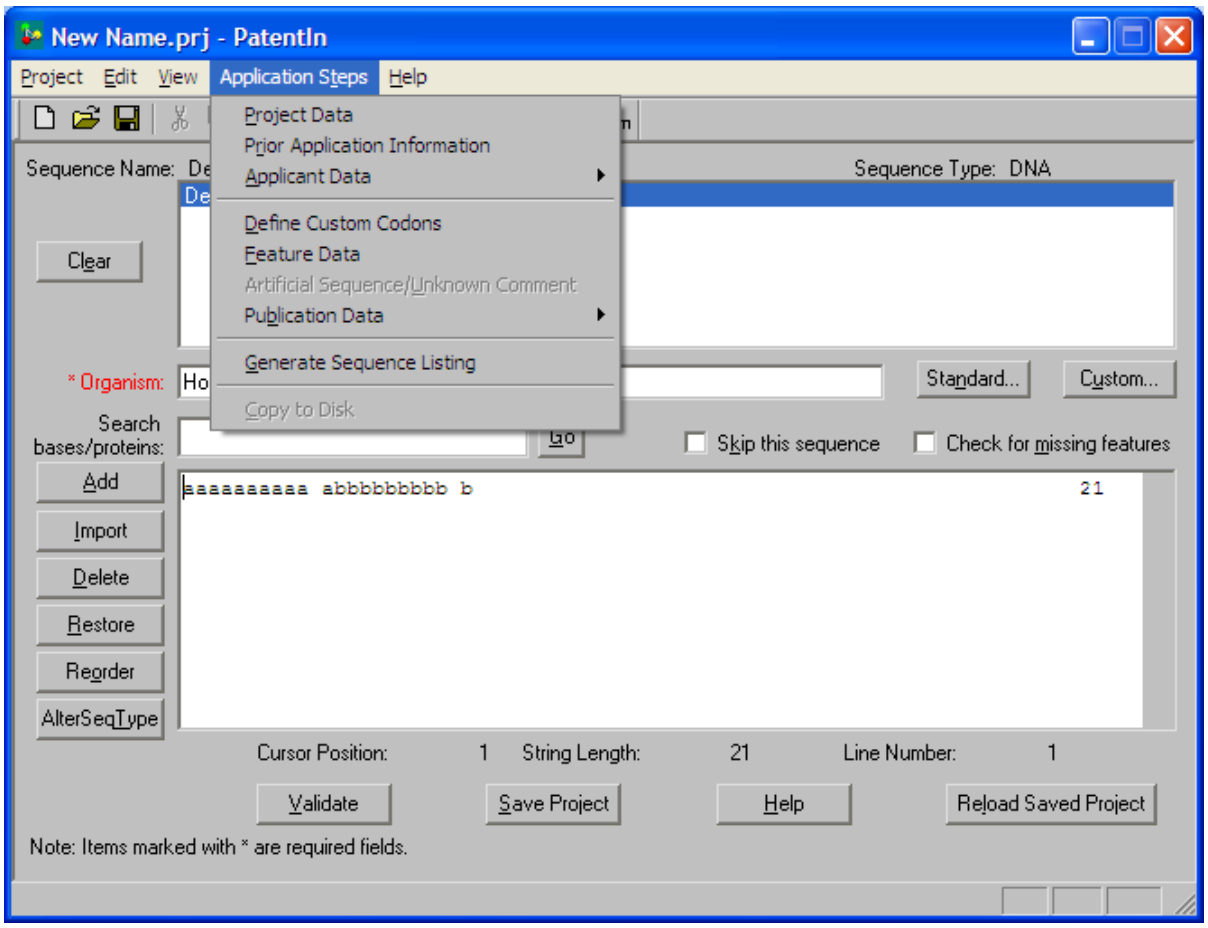

Figure 4-1: Application Steps Menu

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.

Note: The mandatory information fields (Title of Invention and Application File Reference) are in red and marked with asterisks.

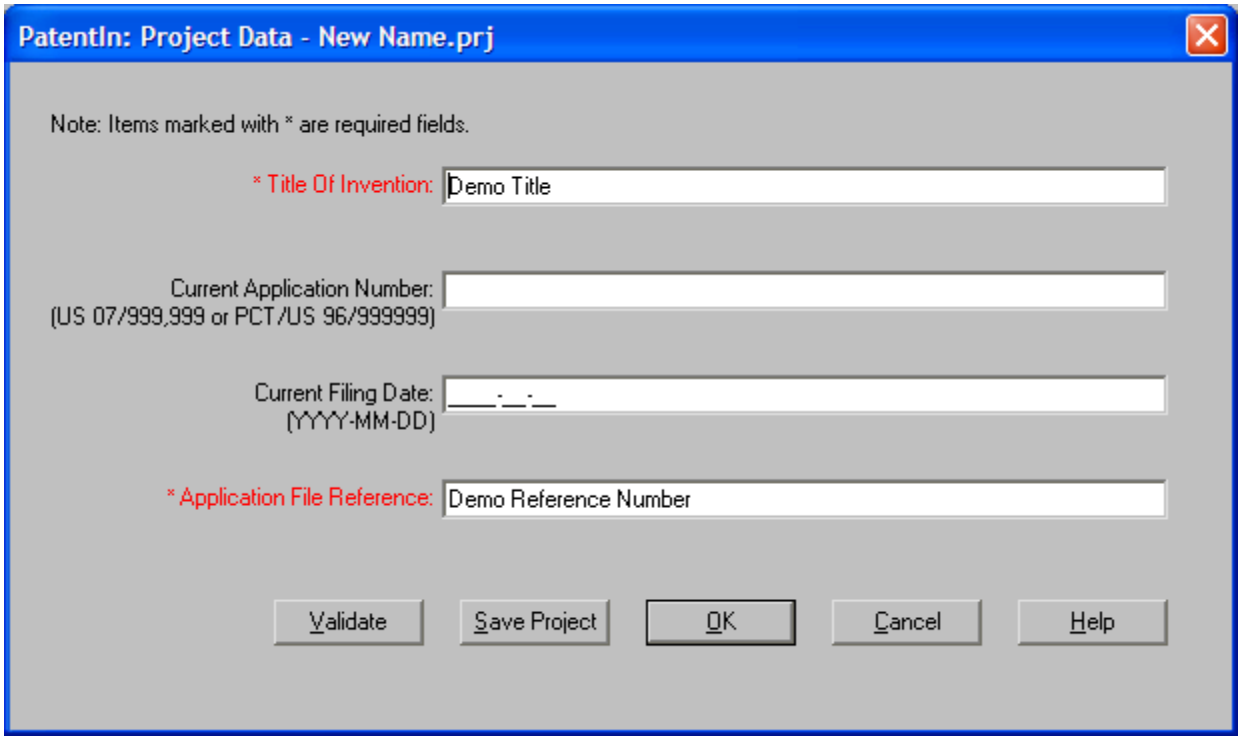

Figure 4-2: Project Data Screen

To enter project data:

1. Enter the **Title Of 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**. This information is mandatory.
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.

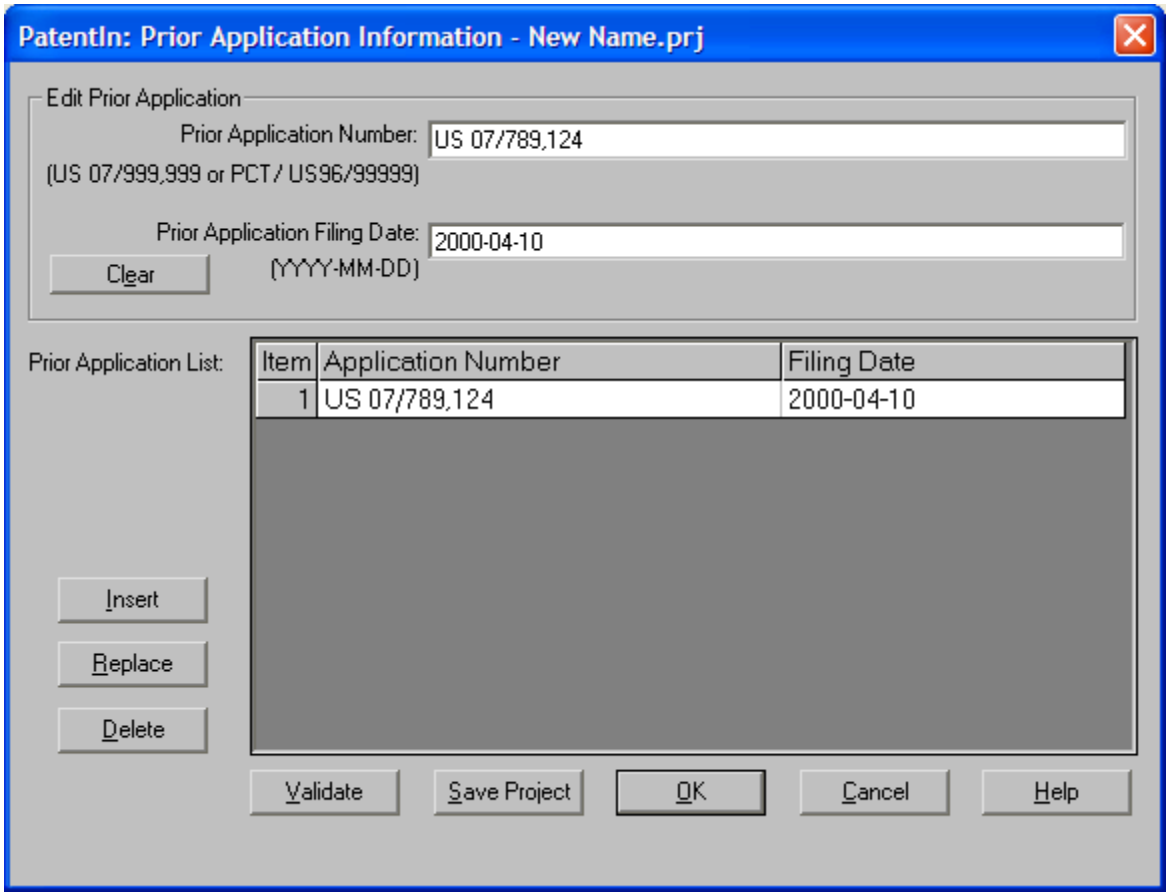

Figure 4-3: Prior Application Information Screen

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 Numbe**r 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.

Note: 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 Application 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.

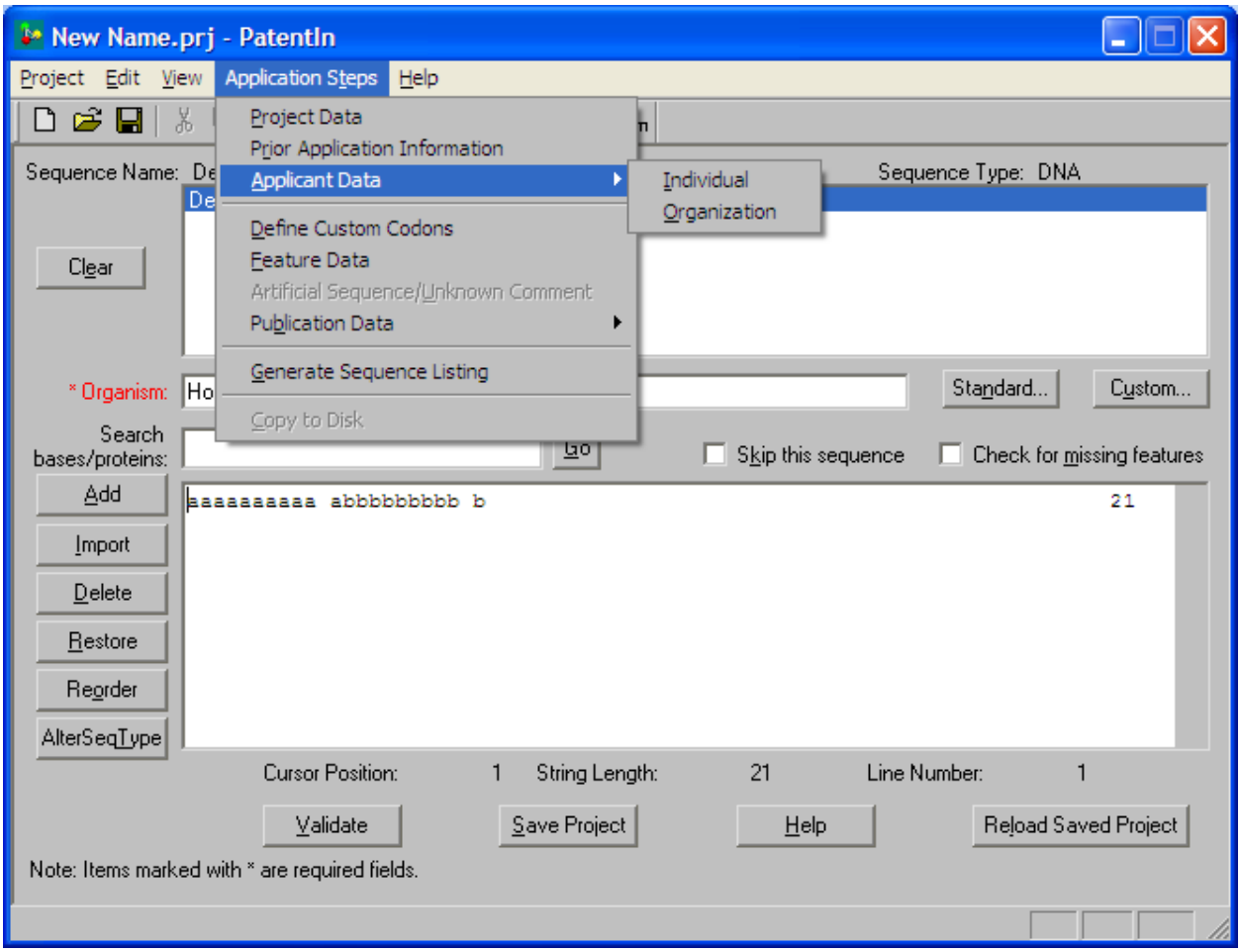

Figure 4-4: Application Data Screen

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.

Note: The field names in red and marked with asterisks (Last Name and First Name) are mandatory information.

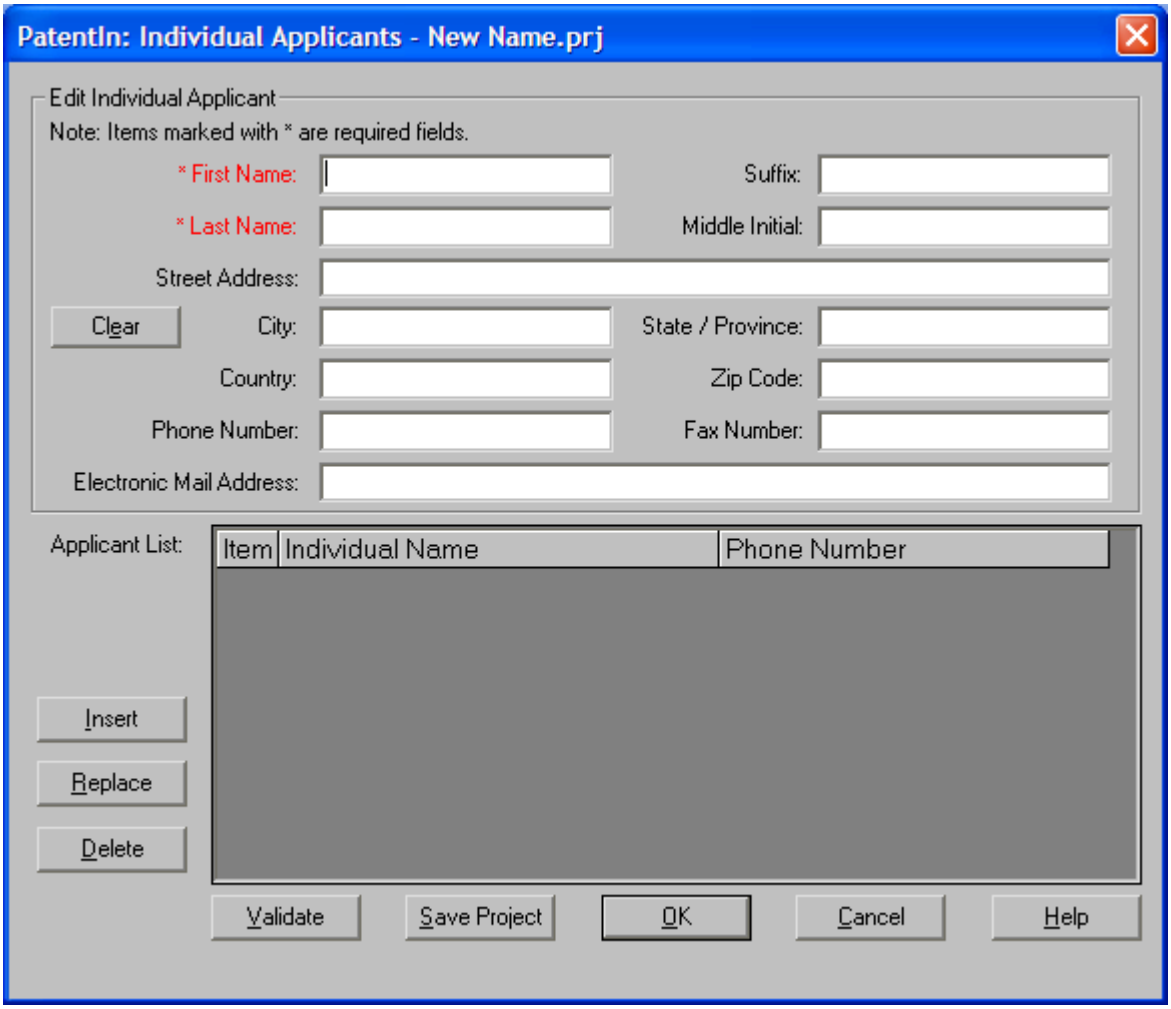

Figure 4-5: Individual Applicants Screen

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, and 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 not 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. Also 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.

Note: The field name in red and marked with asterisk (Organization) is mandatory information.

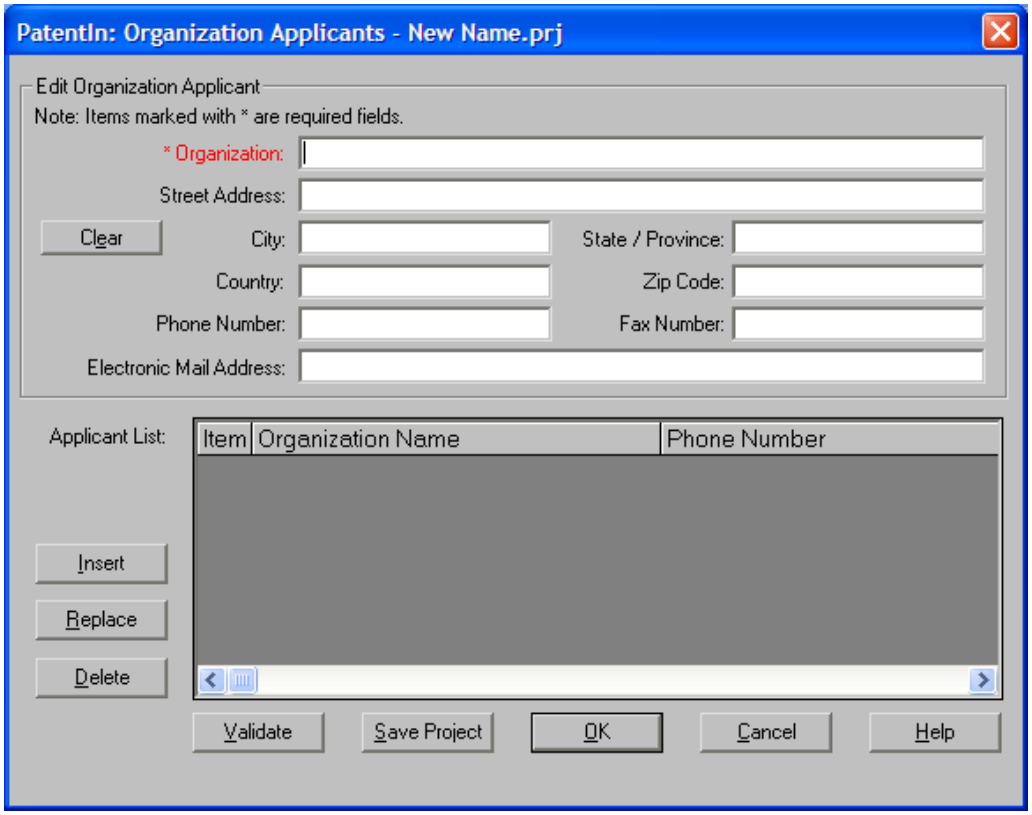

Figure 4-6: Organization Applicants Screen

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, and 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 8 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. Also 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.

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.

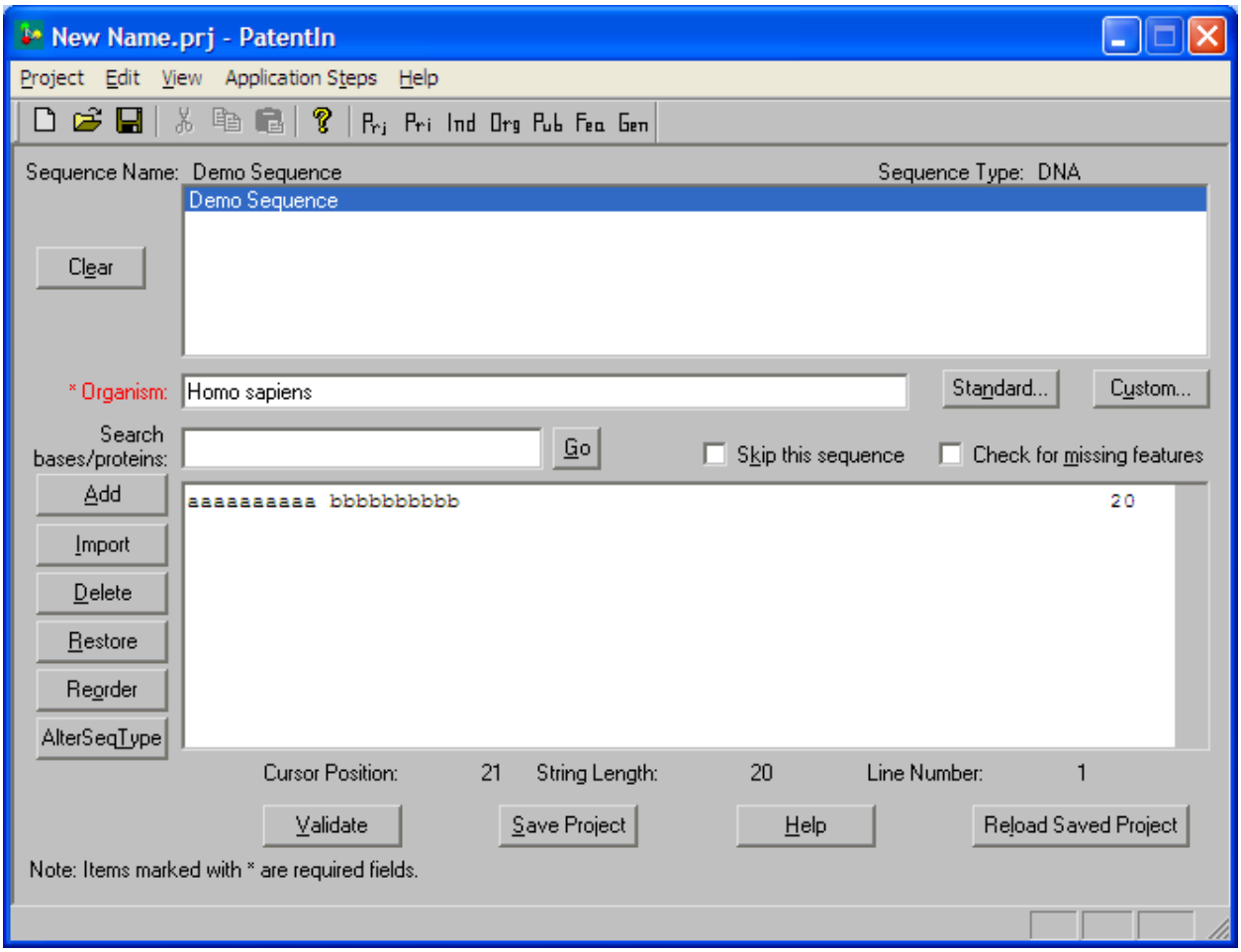

Figure 5-1: Sequence Screen

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

To select a sequence for editing:

1. Select a sequence name from the list of sequences.

The following sequence characteristics are displayed:

- **Cursor Position -** 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

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.

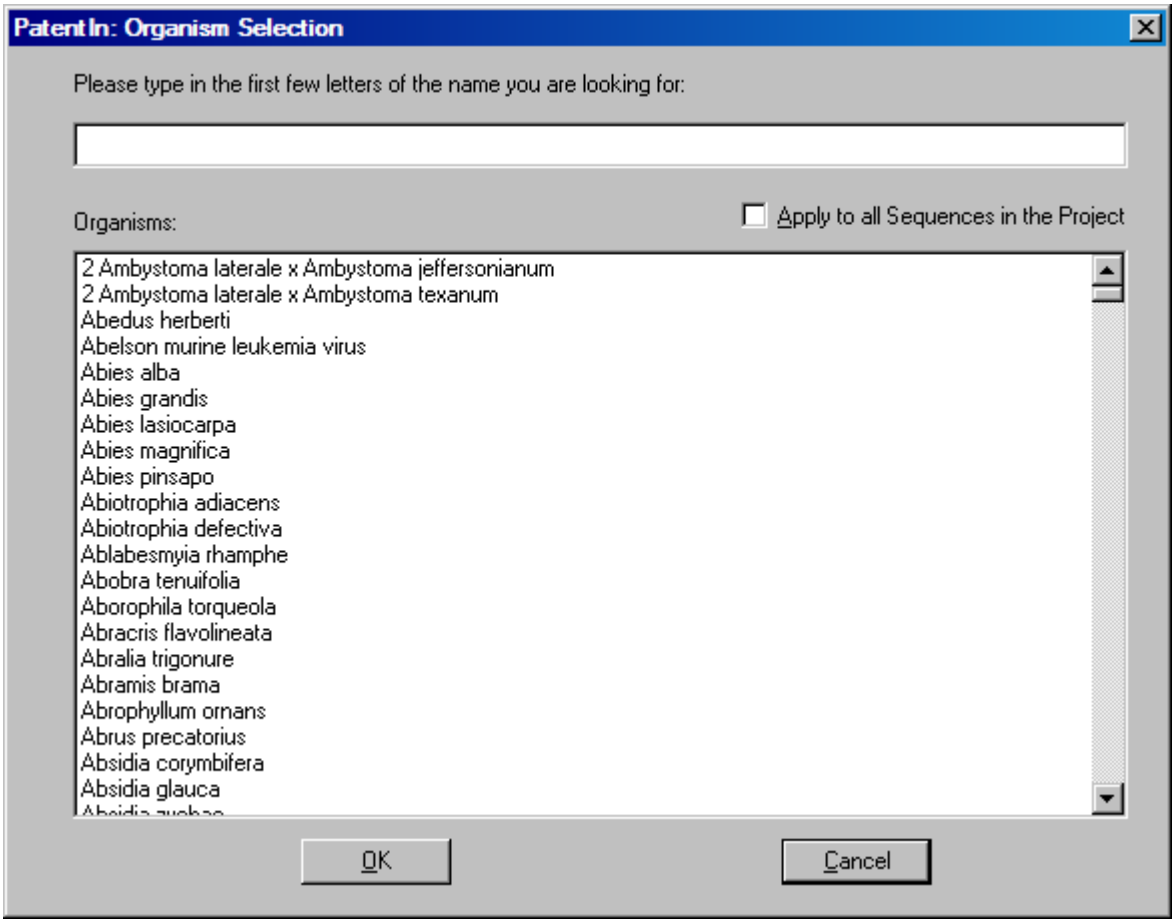

Figure 5-2: Selecting an Organism Screen

To select an organism name:

1. Click on the **Standard** button (Sequence Screen).
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.

Note: The Organism in red and marked with asterisks is mandatory information.

5.1.2 Creating Default Explanations for "n"s and "Xaa"s

PatentIn 3.5.1 can create default explanations for the entries "n" and "Xaa" 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 features** box on the Sequence Screen 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 3.5.1 will create any missing misc_features. Users may create selected definitions by hand and PatentIn 3.5.1 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 **Search bases/proteins** edit field (Figure 5-1).
2. Click on the **Go** button. The cursor will move to the first instance of that sub-sequence beginning with its current position.

Note: Each search is limited or 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, as shown in Figure 5-1.

5.2 Adding a Sequence

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

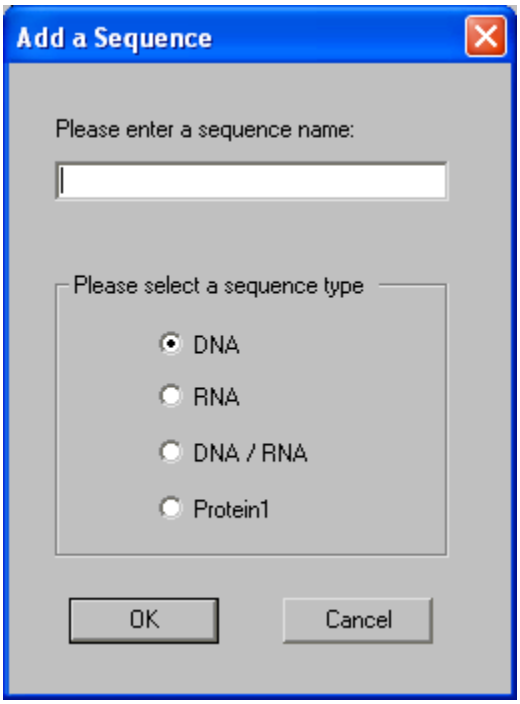

Figure 5-3: Add a Sequence Screen

To add a sequence:

1. From the Sequence Screen, select the **Add** button. Figure 5-3 is displayed.
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 XP or Windows Vista, the upper limit is over 12 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 provides a means to import sequences from a file, a project or a PatentIn generated ST.25 sequence Listing file by selecting one of the radio buttons in Figure 5-4.

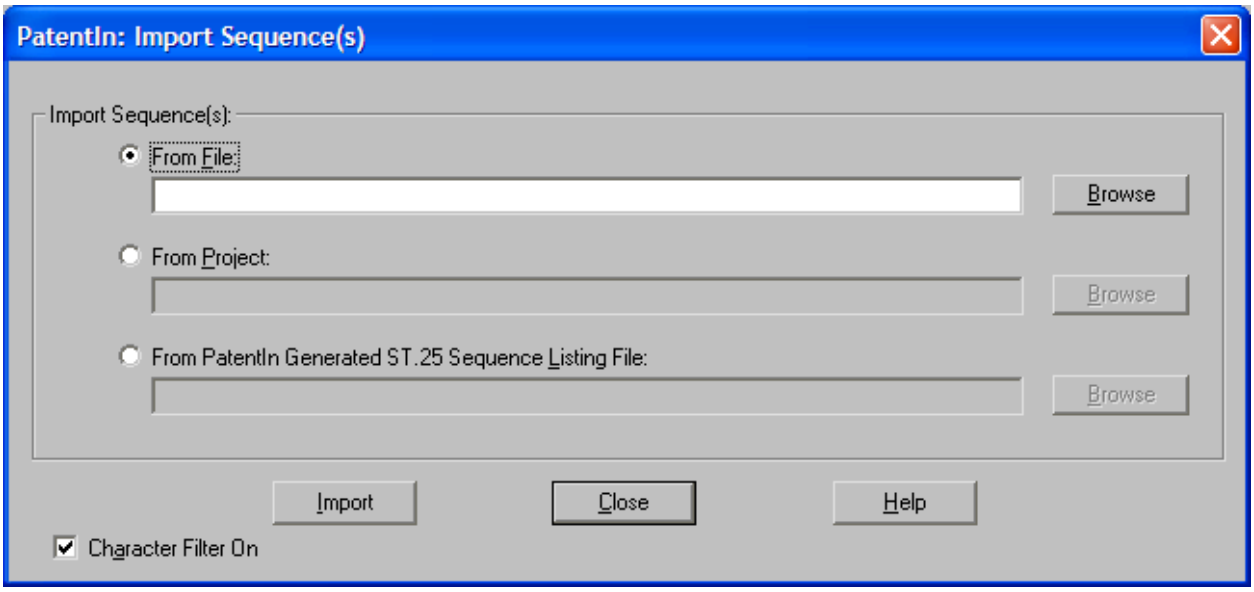

Figure 5-4: Import Sequence(s) Screen

1. To import sequences from a non ST.25 text file, click on the **Import** button on the Sequence Screen, and then click on the **From File** radio button (See section 5.3.1).
2. To import sequences from another project, click on the **Import** button on the Sequence Screen, and then click on the **From Project** radio button (See section 5.3.3).
3. To import sequences from PatentIn generated ST.25 sequence listing file, click on the **Import** button on the Sequence Screen, and then click on the **From PatentIn Generated ST.25 Sequence Listing File** radio button (See section 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. Deselecting 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 3.5.1 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 when importing sequences from a non ST.25 text file. For best results it is recommended that, the non ST.25 sequence files contain only the sequence with spaces and numbering, and that any title or other text is deleted prior to importing. For example, if the title of the sequence is Genomic Deoxyribonucleic Acid (DNA), PatentIn 3.5.1 would filter out the characters "e," "o" and "i" and so the first seven characters of your sequence would be "gnmcdna."

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 C: 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 (Non-ST.25 Sequence Listing File) to be Imported by PatentIn 3.5.1

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

Sequence Name	The Name of the Sequence
SequenceType	One of the following: • DNA • RNA • DNA/RNA • Protein/1 • PRT • PRT/1 • PRT1 • Protein/3 • PRT/3 • PRT3
OrganismName	The name of the organism is optional. If it is omitted, the header looks like: *<SequenceName;SequenceType;>*

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 in Figure 5-5.

```
< First Sequence;DNA;Abies alba>
ttttcttattgtttctcctactgcttatcataatgattgtcgtagtggcttcctcatcgt
ctcccccaccgcctaccacaacgactgccgcagcggattactaatagtatcaccaacagc
ataacaaaaagaatgacgaagagggttgctgatggtgtcgccgacggcgtagcagaagga
gtggcggagggg

< Second Sequence;RNA; >
uuuucuuauuguuucuccuacugcuuaucauaaugauugucguaguggcuuccucaucgu
cucccccaccgccuaccacaacgacugccgcagcggauuacuaauaguaucaccaacagc
auaacaaaaagaaugacgaagaggguugcugaugguguucgccgacggcguagcagaagga
guggcggagggg
```

Figure 5-5: Sample ASCII Sequence Data

5.3.2 Format for Single Sequence Data Files to be Imported by PatentIn 3.5.1

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 in Figure 5-6 and the file is assumed a single sequence data file.

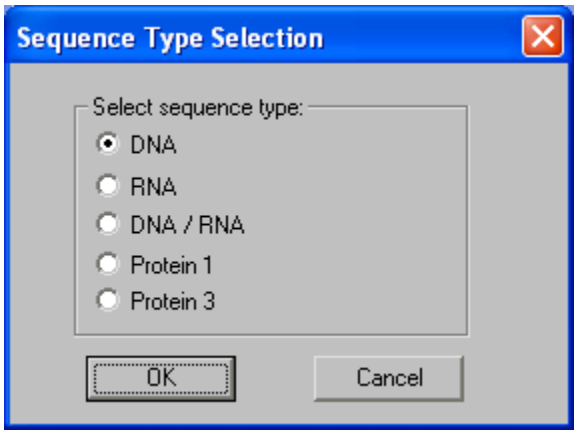

Figure 5-6: Sequence Type Selection Screen

While the import feature is running, a screen displays the total number of sequences that have thus far been analyzed. Figure 5-7 shows that a multi-sequence file named SEQ10000.txt has analyzed 696 sequences at the time this screen was captured.

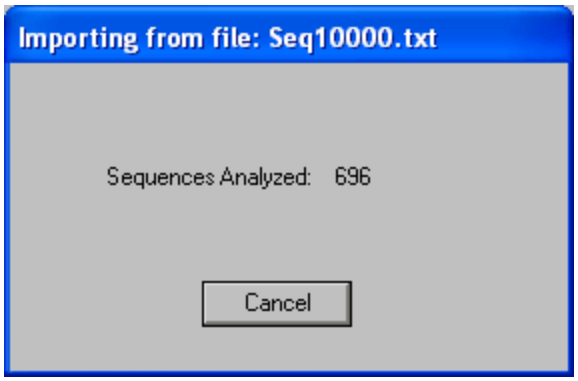

Figure 5-7: Sequences Being Imported Screen

A Validation Errors screen is displayed if validation errors occur. In the next example (Figure 5-8), “e” is an invalid character for Sequence 1 of type DNA and Sequence 2 contains an invalid sequence type.

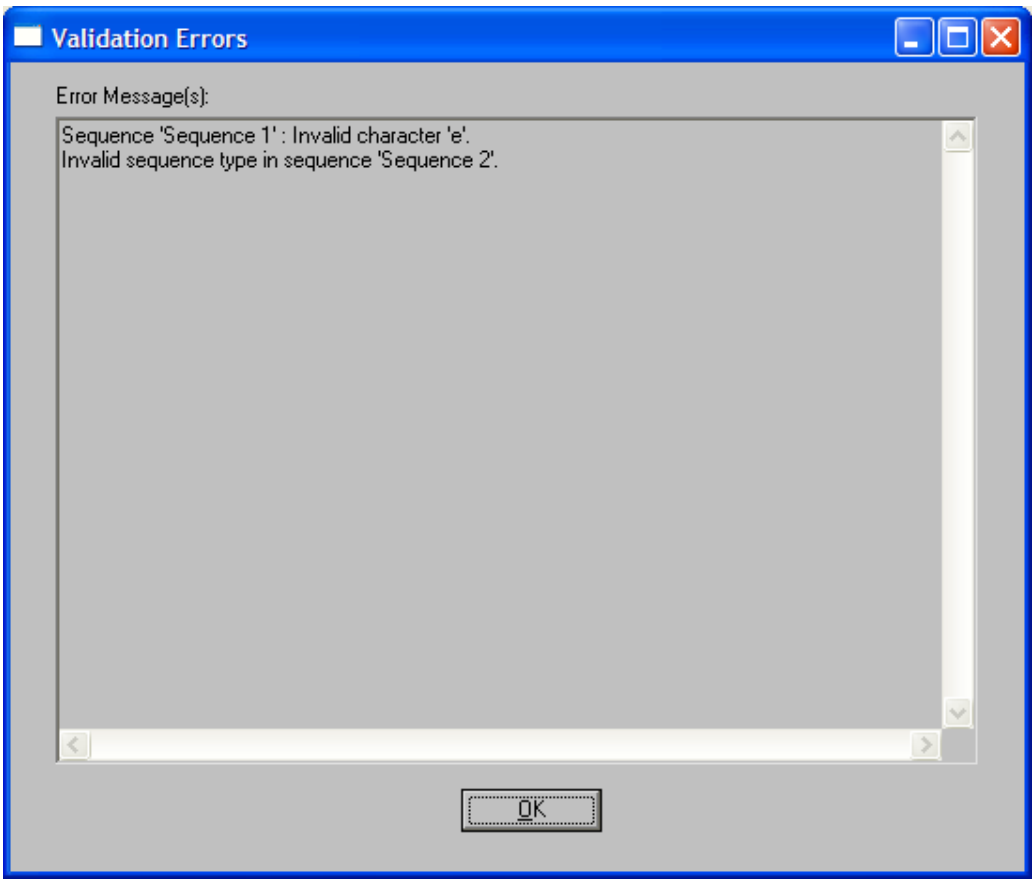

Figure 5-8: Validation Errors Screen

5.3.3 Importing Sequences from a Project

PatentIn provides a mechanism to import sequences from a PatentIn 3.5.1 project (Figure 5-4).

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).

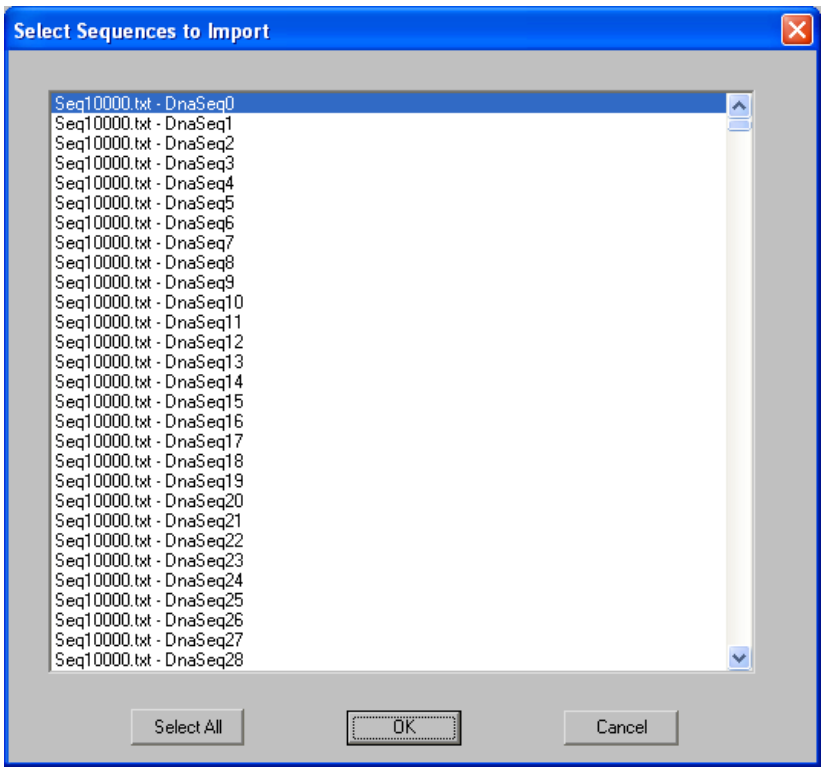

Figure 5-9: Select Sequences From Project Screen

5.3.4 Importing Sequences from a PatentIn Generated ST.25 Sequence Listing File

PatentIn provides a mechanism to import sequence information from a PatentIn generated ST.25 sequence listing file. A PatentIn generated ST.25 sequence listing file is a sequence listing file generated by PatentIn and is complied with ST.25. The name of the file is composed of the PatentIn project name followed by "_ST25.txt". An example of such a file would be "PatentInProject_ST25.txt".

1. To import a PatentIn generated ST.25 sequence listing file, press the **Import** button on the sequence screen (Figure 5-1) and then select the **From PatentIn Generated ST.25 Sequence Listing File** radio button on the Import Sequence(s) screen (Figure 5-4).
2. Select a PatentIn generated ST.25 sequence listing file that has "_ST25.txt" at the end of the file name. The **Browse** button is provided to assist the user in selecting the file folder and file name.

3. Click on the **Import** button. PatentIn will acknowledge the user with a message box (Figure 5-10) with "The sequence listing will be imported to a new PatentIn project" message.

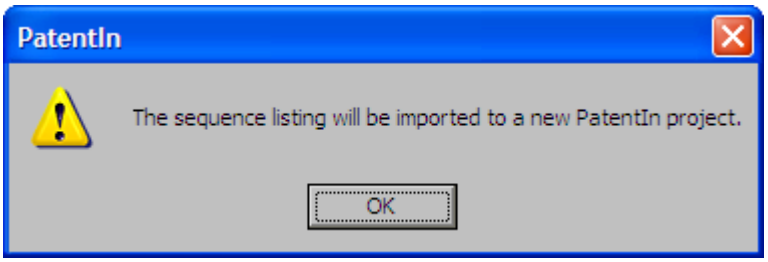

Figure 5-10: Import to New Project Message Dialog Screen

4. Press the **OK** button to close the message box. If the user has already a PatentIn project open and the project has not been saved, PatentIn will display a dialog box (Figure 5-11) and ask whether to save changes to the currently working project. Press the **Yes** button to save the changes or the **No** button to discard the changes. To cancel the import process, select the **Cancel** button.

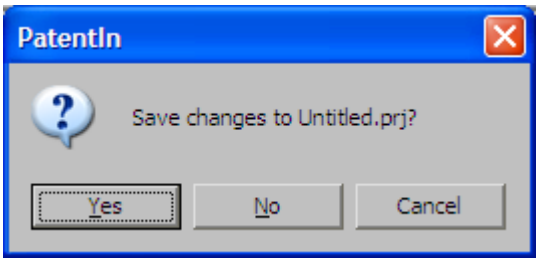

Figure 5-11: Save Project Screen

5. Once the import process starts, the **Parsing the Applicant Names** screen (Figure 5-12) will be displayed.

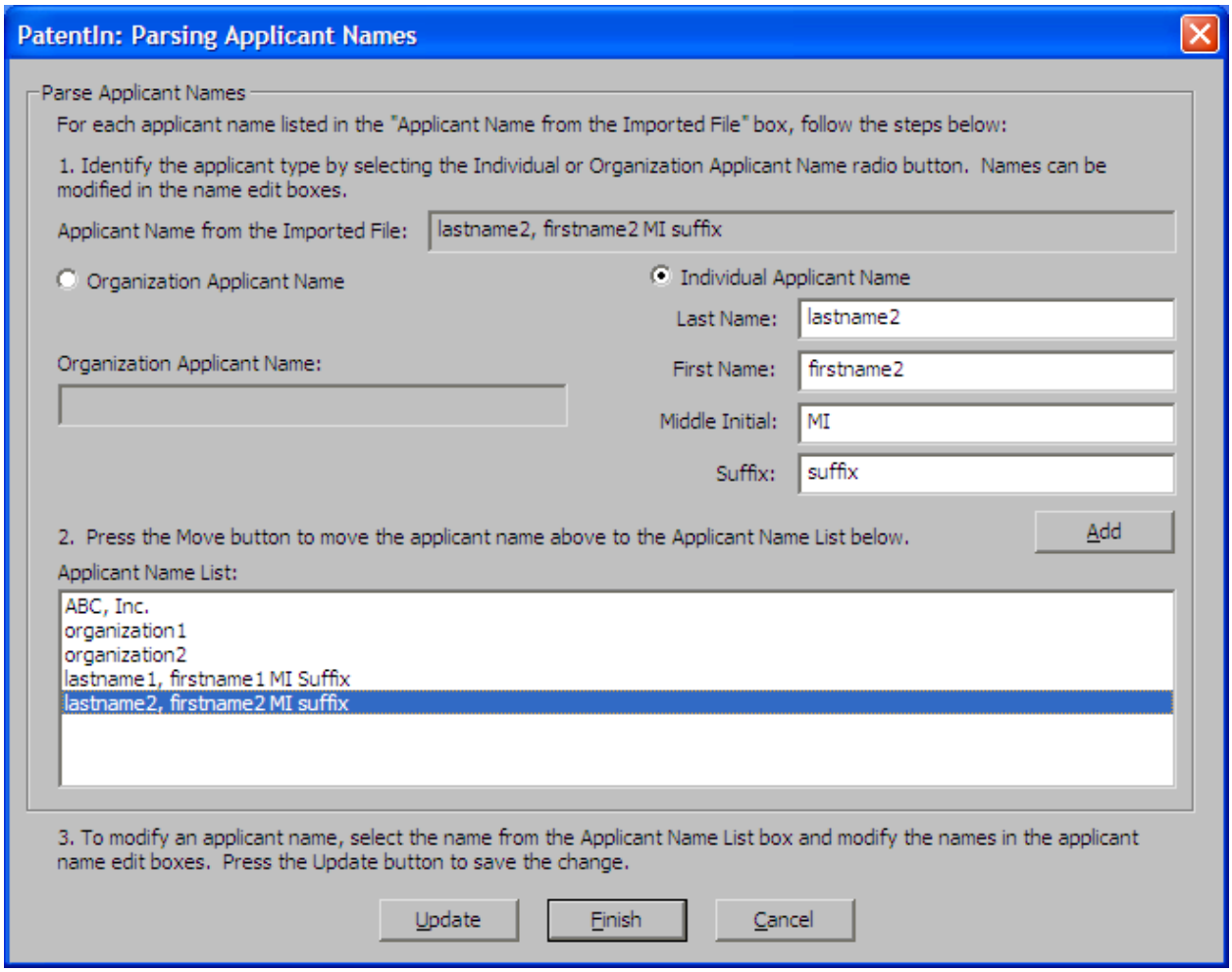

Figure 5-12: Parsing Applicant Names Screen

For each applicant name listed in the ST.25 sequence listing file, PatentIn displays it in the non-editable **Applicant Name from the Imported File** box on the **Parsing the Applicant Names** screen. The user should first identify the type of this applicant name by selecting the **Individual Applicant Name** or **Organization Applicant Name** radio button (Organization Applicant Name radio button is selected by default). The applicant names can be modified in the **Organization Applicant Name** box or in the individual applicant name text boxes composed by the **Last Name**, **First Name**, **Middle Initial** and **Suffix** boxes. Once the type of the applicant name is selected and/or the applicant name is modified, press the **Add** button to move the current applicant name to the **Applicant Name List** box as well as to get and display the next applicant name from the imported file to the **Applicant Name from the Imported File** box. Repeat this process until all names are added to the **Applicant Name List** box. Then PatentIn will acknowledge the user with a message "All applicant names have been added". Press the **OK** button to close the message box (Figure 5-13).

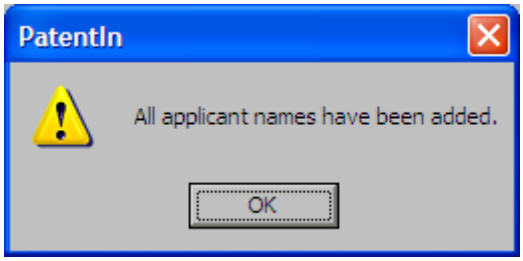

Figure 5-13: Added All Names Message Dialog Screen

To modify any applicant name again after all names have been added to the **Applicant Name List** box, select the name from the Applicant Name List box and modify the names in the applicant name text boxes. Press the **Update** button to save the change. If no more modifications are needed, press the **Finish** button to close the **Parsing the Applicant Names** screen. The import process continues.

Note: It is necessary for PatentIn users to verify if the imported sequence information is interpreted and stored in the PatentIn system as anticipated after the import process since the data may or may not be converted to be 100% compliant with the imported ST.25 sequence listing.

Additional Information on Importing PatentIn Generated ST.25 Sequence Listing:

- **Individual Applicant Name Parsing**: While parsing an individual applicant name, PatentIn will identify everything before the leftmost comma of an applicant name as the last name. When no comma is present in the individual applicant name, PatentIn will identify the last word of an imported line of applicant name as the last name.
- **Primary Sequences and Supplemental sequences**: Only primary sequences will be imported to PatentIn system. All supplemental sequences (coding sequence) immediately following the primary sequence will not be imported because they will be generated during the sequence listing generation. However, all supplemental sequences immediately following a suppressed supplemental sequence (a supplemental sequences containing less than four amino acids and is not present in the imported file) will be imported to PatentIn as new protein sequences.
- **Skipped Sequences**: All skipped sequences will be imported to PatentIn. Because a skipped sequence does not have a sequence type, PatentIn will ask the user to select a type by displaying the **Skipped Sequence Type Selection** screen (Figure 5-14). Press the **OK** button after the selection is made.

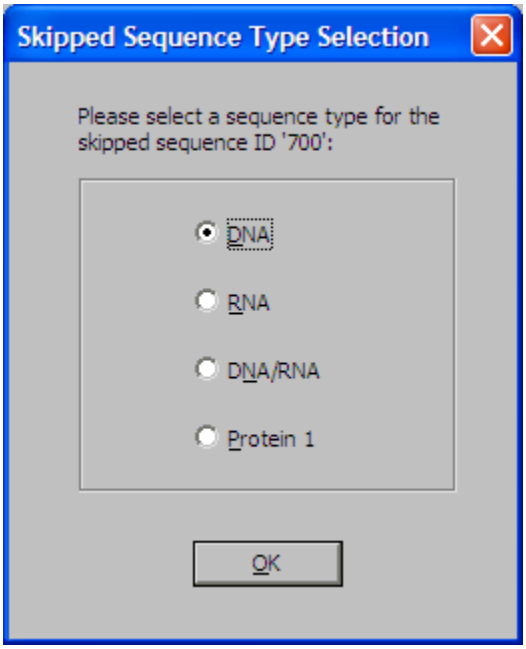

Figure 5-14: The Skipped Sequence Type Selection Screen

- **Sequences with Invalid Type**: If PatentIn encountered a sequence with an invalid sequence type while importing, it will display the **Invalid Sequence Type Encountered** screen (Figure 5-15) with information of the sequence ID, the invalid type and the first 60 characters of the sequence. Press the **OK** button after the correct type is selected.

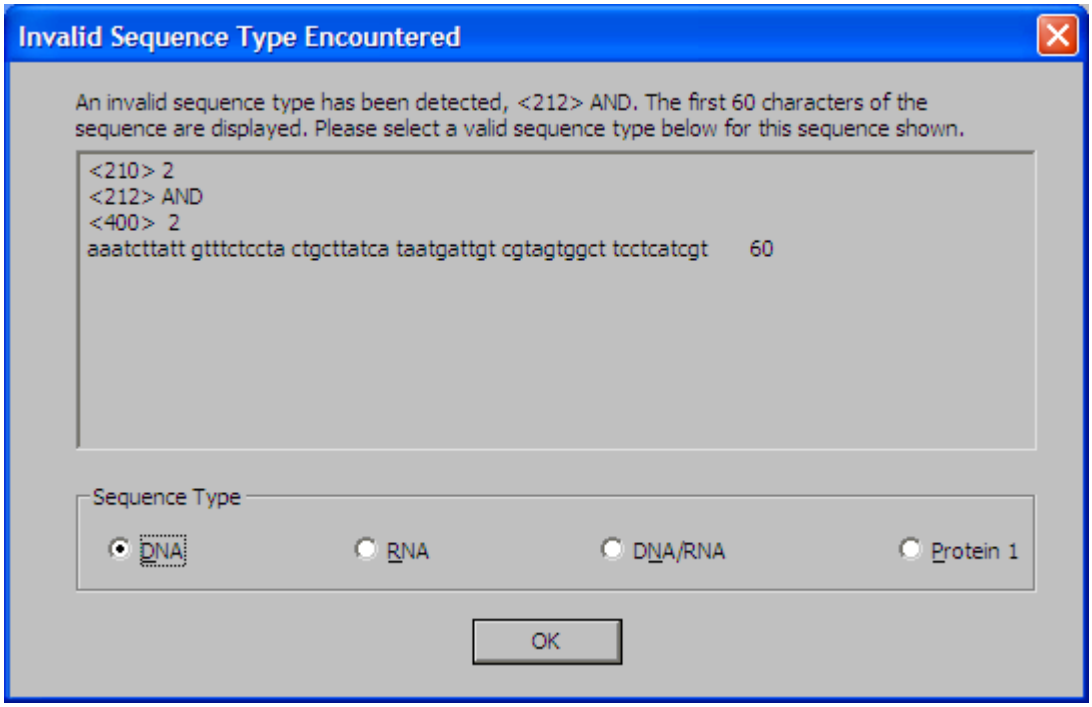

Figure 5-15: Invalid Sequence Type Encountered Screen

- **A List of Sequences to Be Imported**: After all sequences have been parsed, PatentIn displays a list of sequences ready to be brought in to PatentIn on the **Select Sequences to Import** screen (Figure 5-16). Select the sequence(s) to be imported (**Note**: By holding the Ctrl key down multiple sequences may be selected) or press the **Select All** button to select all of the sequences (Figure 5-16). Press the **OK** button to bring the selected sequences to the PatentIn system.

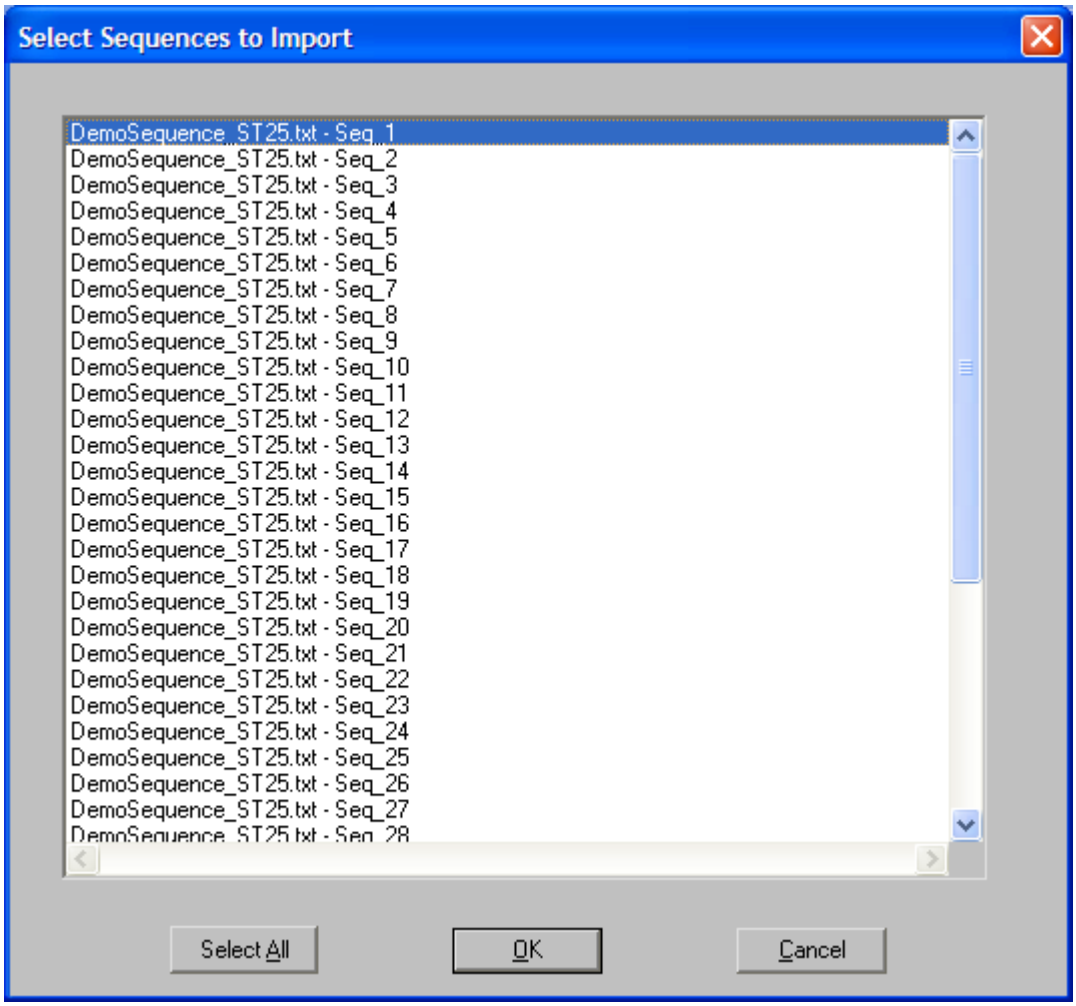

Figure 5-16: Select Sequences to Import Screen

- **Log File**: The parsing messages generated (if there is any) during the import process will be recorded in a log file with the name <PatentInProjectName>_ ParseFileErrorLog.txt.

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 5-17).

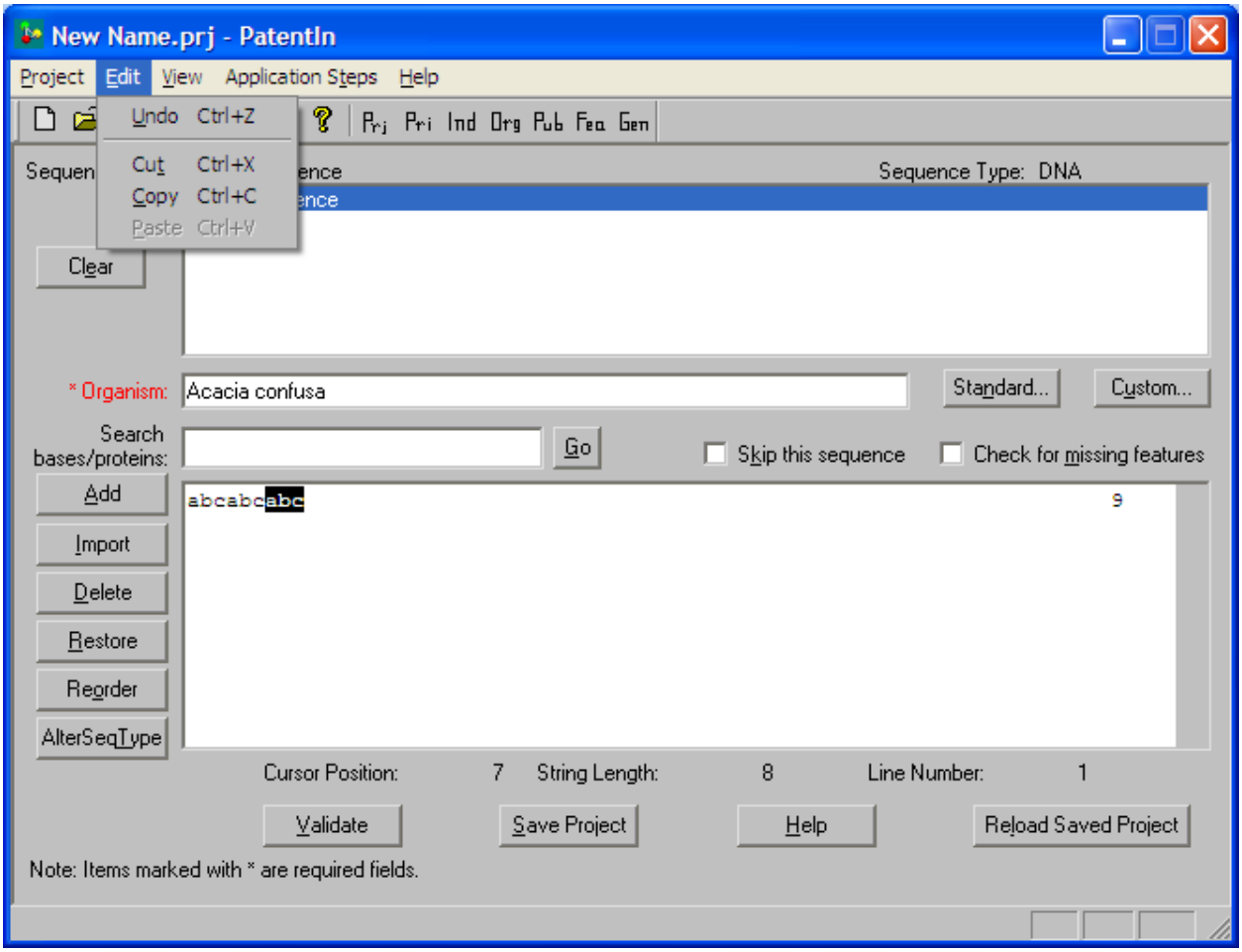

Figure 5-17: Edit Menu

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**, and 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 3.5.1 will not generate anything between the <210> and the <400> for this sequence.

To skip a sequence:

1. Select the sequence to be skipped.
2. Click on the **Skip this sequence** checkbox to turn on the skip sequence flag on the Sequence Screen (Figure 5-1). PatentIn 3.5.1 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:

1. Click on the **Restore** button on the main screen. The Sequence Recovery Screen (Figure 5-18) appears.
2. Select the sequence(s) to be restored.
3. Click on the **Restore** button.

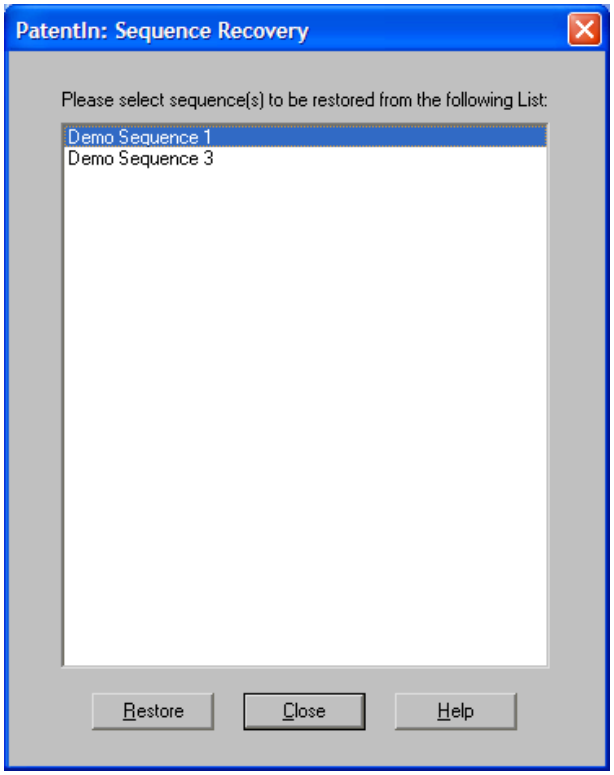

Figure 5-18: Sequence Recovery Screen

5.9 Reordering Sequences

The Recorder Sequence Screen (Figure 5-19) 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.

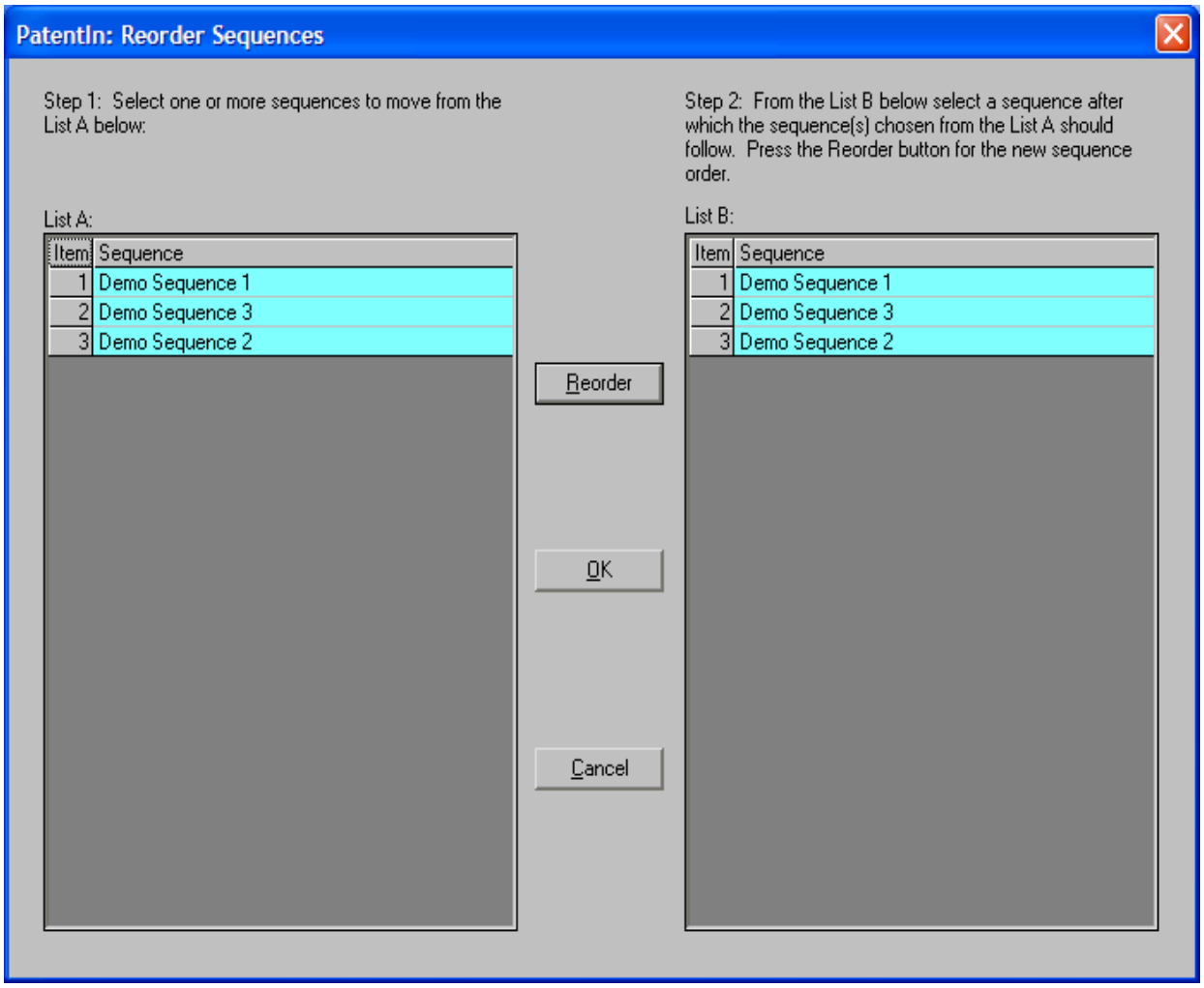

Figure 5-19: Reorder Sequences Screen

To reorder sequences:

1. Select the sequence row(s) to move from the sequence List A.
2. From the sequence List B select a sequence after which the sequence(s) chosen from the List A should follow. Press the Reorder button for the new sequence order.
3. Repeat steps 1-2 until the sequences are in the desired order.

5.10 Changing a Sequence Type

The **AlterSeqType** button on the Sequence Screen provides a means to change the type of a non-protein sequence to one of the sequence types listed on the Change Sequence Type Screen, as shown in Figure 5-20.

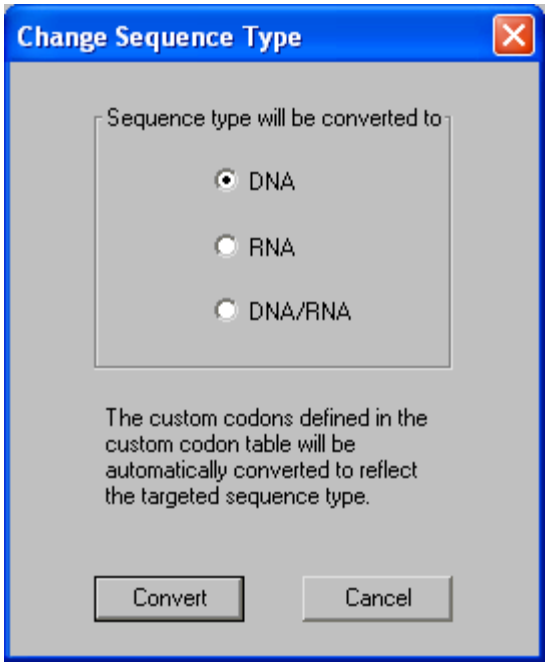

Figure 5-20: Change Sequence Type Screen

To change the type of an existing non-protein sequence:

1. From the Sequence Screen, select the sequence to be converted.
2. Click on the **AlterSeqType** button. The Change Sequence Type screen appears.
3. Select a different sequence type by clicking on the radio button next to the appropriate sequence type.
4. Click **Convert** button.

After the sequence type conversion, the custom codons defined in the custom codon table will be automatically converted to reflect the targeted sequence type. The resulting duplicates of custom codons (if they exist) will be deleted from the custom codon table.

5.11 Validating Sequences

To validate the sequences:

1. On the Sequence Screen, click on the **Validate** button. A message screen will inform you if there was an error, otherwise “Validation is successful” will appear on the status bar.

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

5.12 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 prevent hours of rework, especially when there is a system failure.

5.13 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).

5.14 Adding Custom Codons

The Custom Codons Input Screen (Figure 5-21) 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.

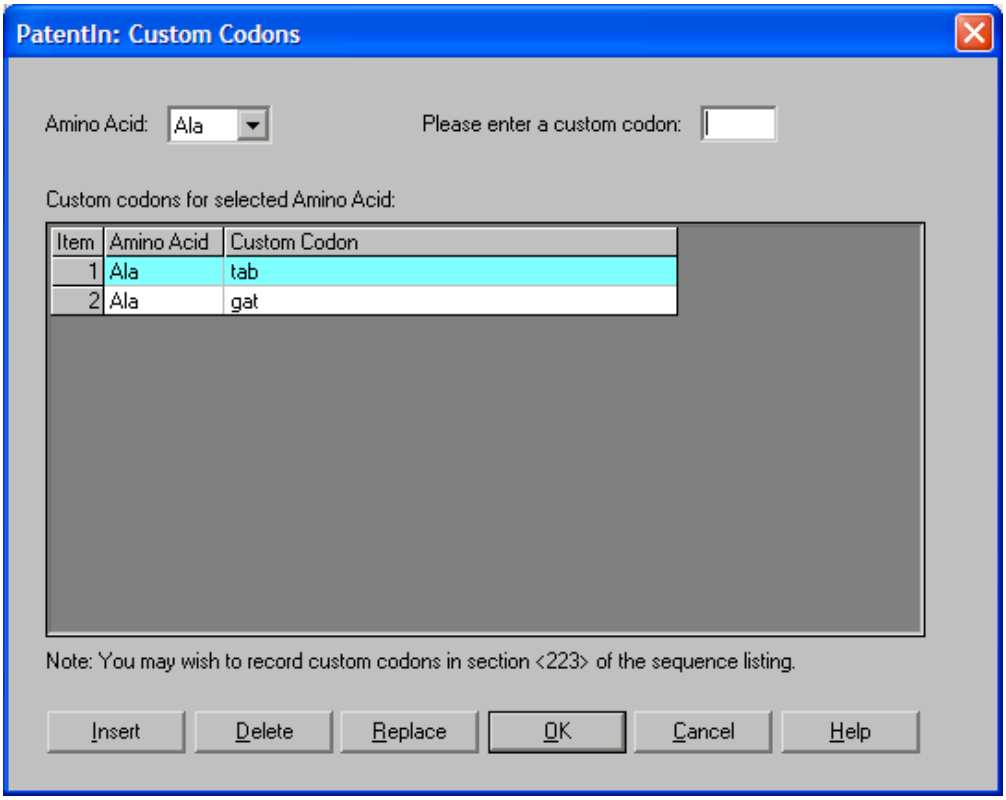

Figure 5-21: Custom Codons Input Screen

To add a custom codon:

1. Select the Define Custom Codons item from the Application Steps menu in Figure 4-1.
2. Select an **Amino Acid** from the drop-down list (Figure 5-22).
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.

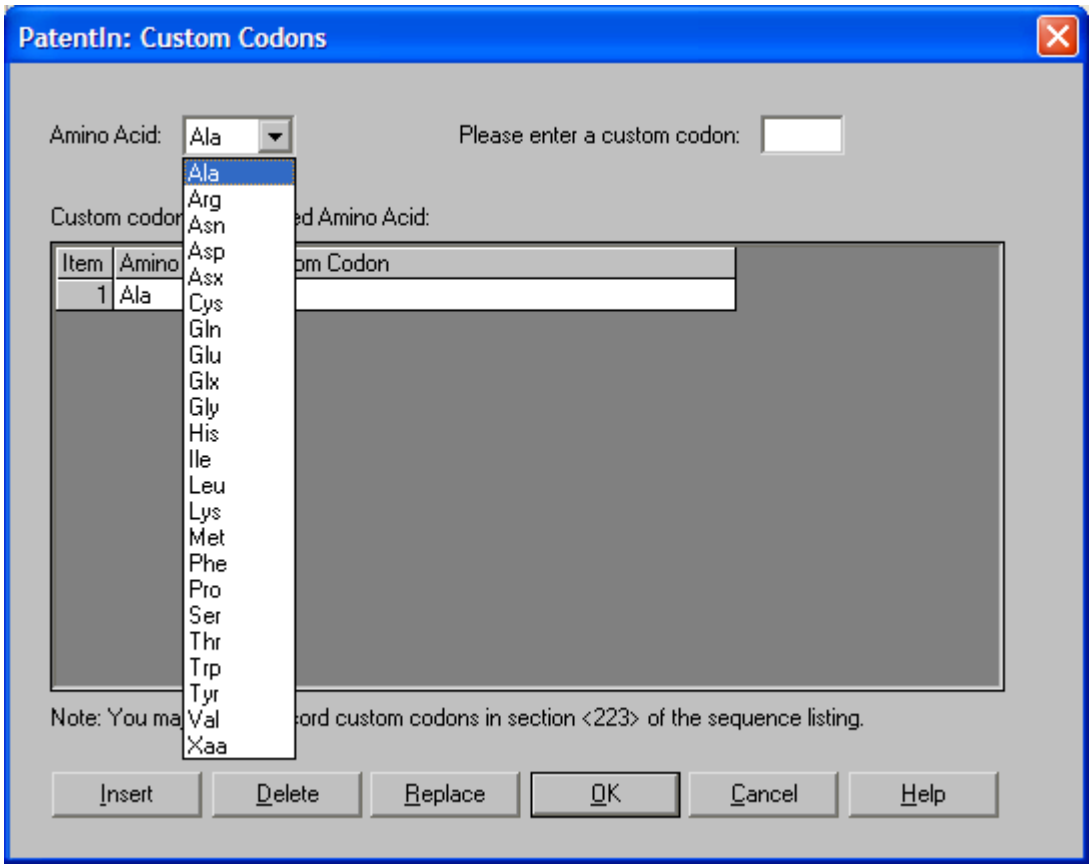

Figure 5-22: Amino Acid Drop Down List Screen

5.15 Adding a Custom Organism

The Custom Organism Input Screen (Figure 5-23) 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.

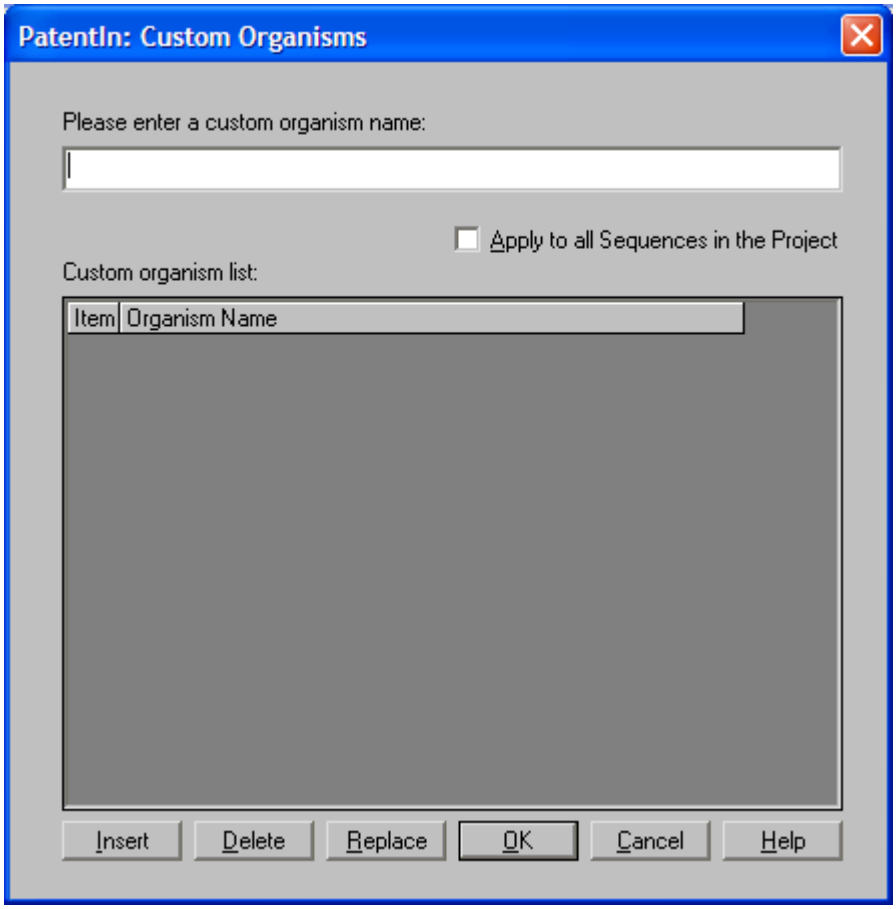

Figure 5-23: Custom Organism Input Screen

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.

5.16 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 as shown in Figure 5-24.

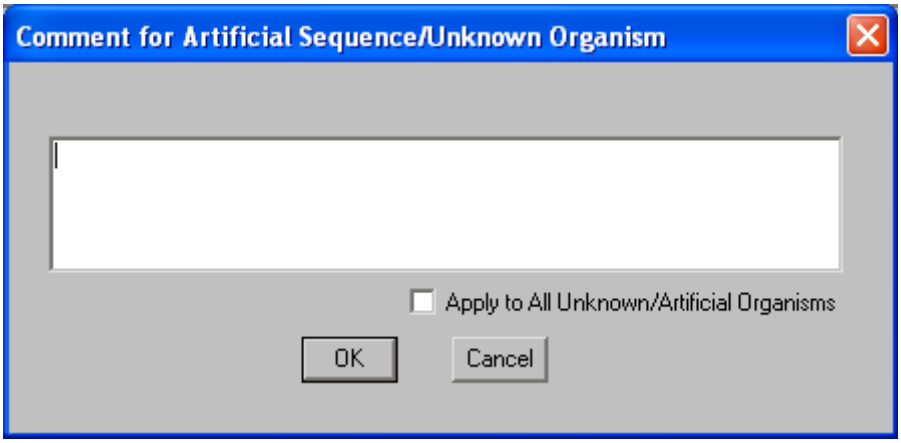

Figure 5-24: Artificial Sequence/Unknown Organism Comment

1. Select the **Standard** button from the Sequence Screen.
2. Select either Unknown or Artificial from the list of standard organisms.
3. Tab or move to another cell, the comment box, Figure 5-24, 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 ways 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.

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

Note: The content of this box should be as descriptive as possible while remaining as terse as possible.

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 3.5.1 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.

Section 6 Feature Data

6.1 Sequence Features

The Feature 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.

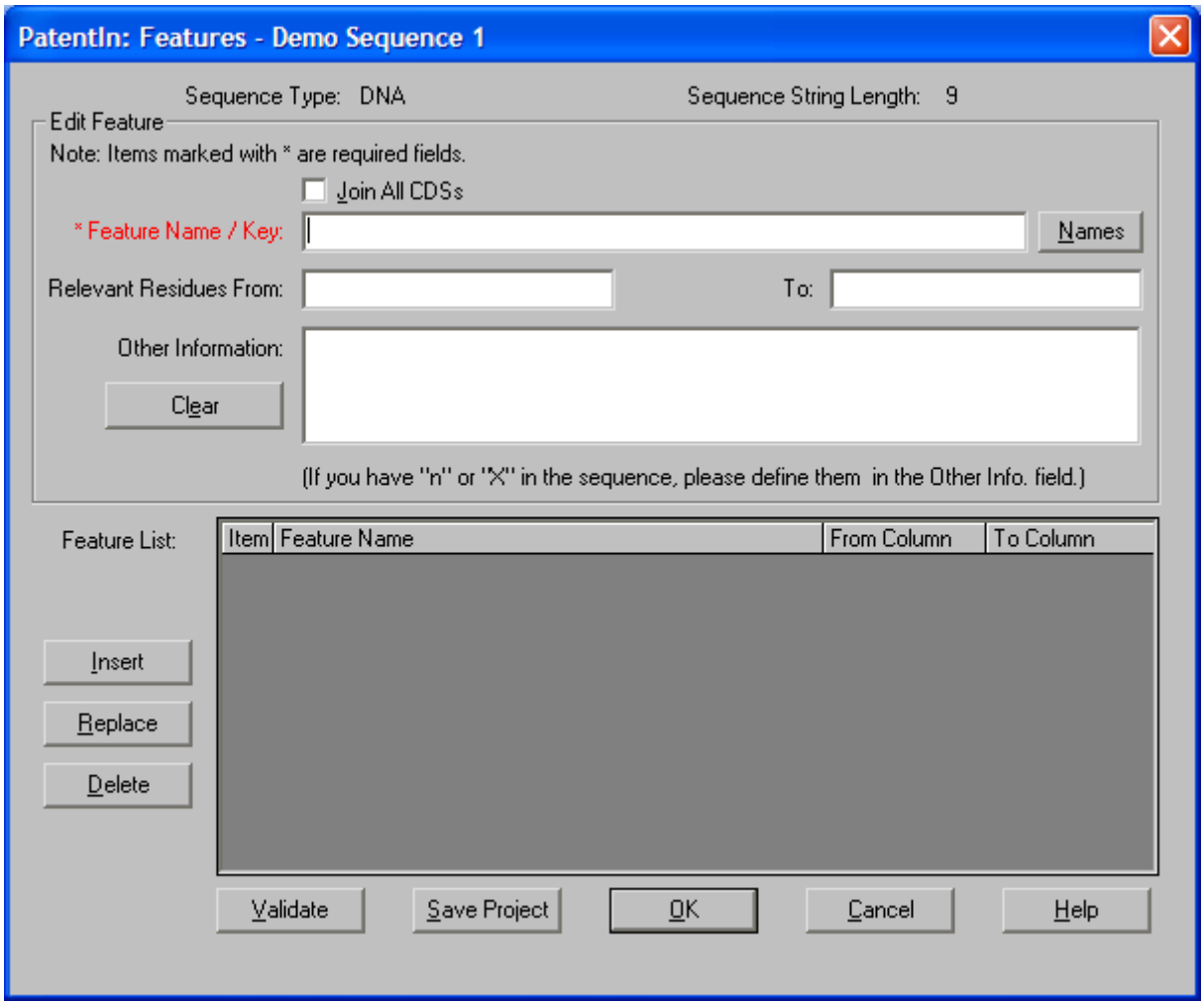

Figure 6-1: Features Screen

To enter information about a Feature:

1. Check the **Join All Code Sequence (CDS)** 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**, 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. Also 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.

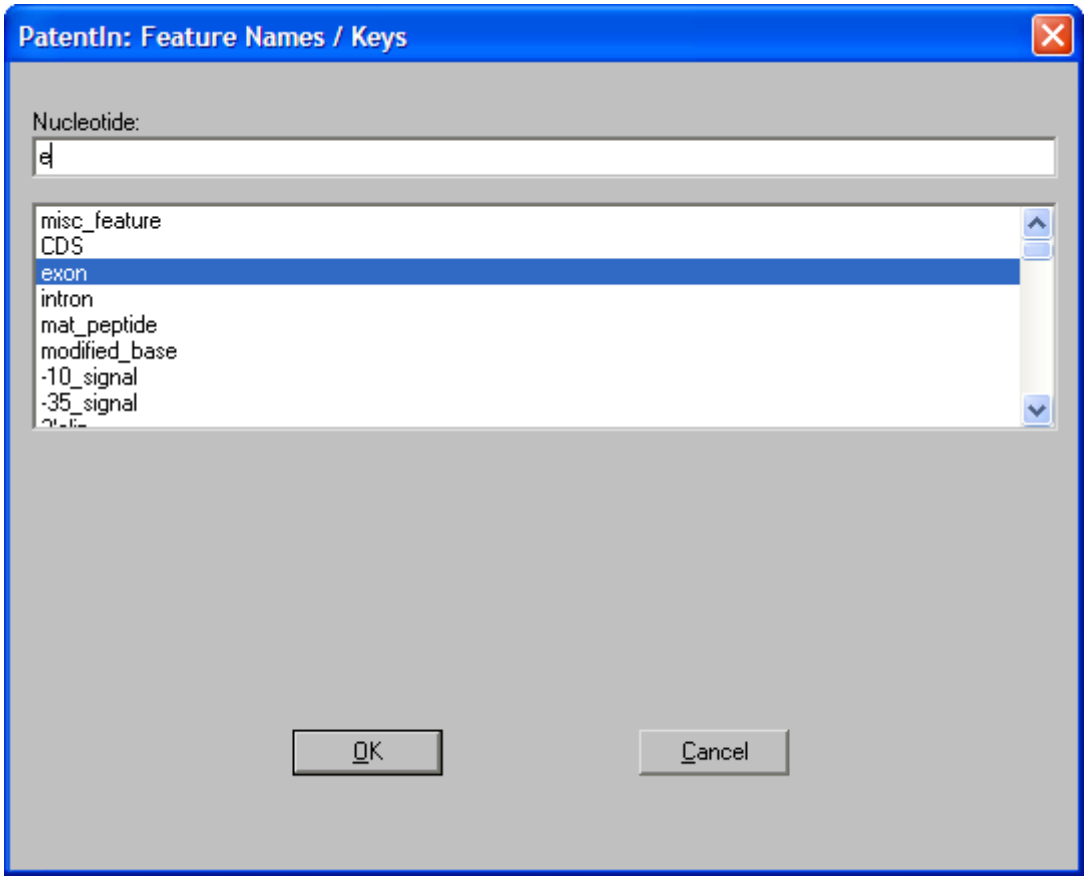

Figure 6-2: Feature Names/Key Selection Screen

To select a nucleotide name:

1. Select the **Names** button on the Feature Screen (Figure 6-1).
2. Begin typing on the **Nucleotide** field.
3. Or, 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 Feature 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.

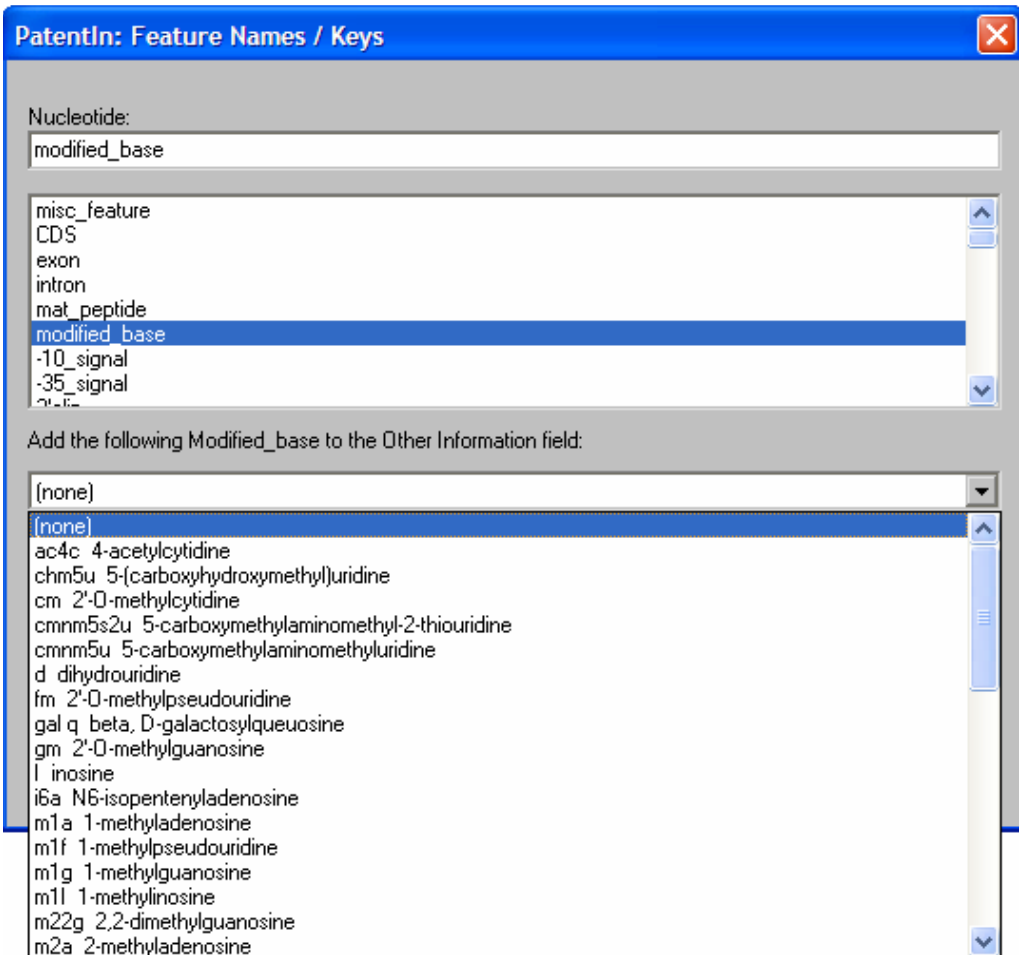

Figure 6-3: Feature Names/Key Selection Screen with Modified Base

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 6-3).

6.1.3 Additional Information on CDS (Coding Sequence)

When a CDS 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 3.5.1 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 Name/Key Selection Screen (Figure 6-4) automatically reveals an additional window when LIPID is selected.

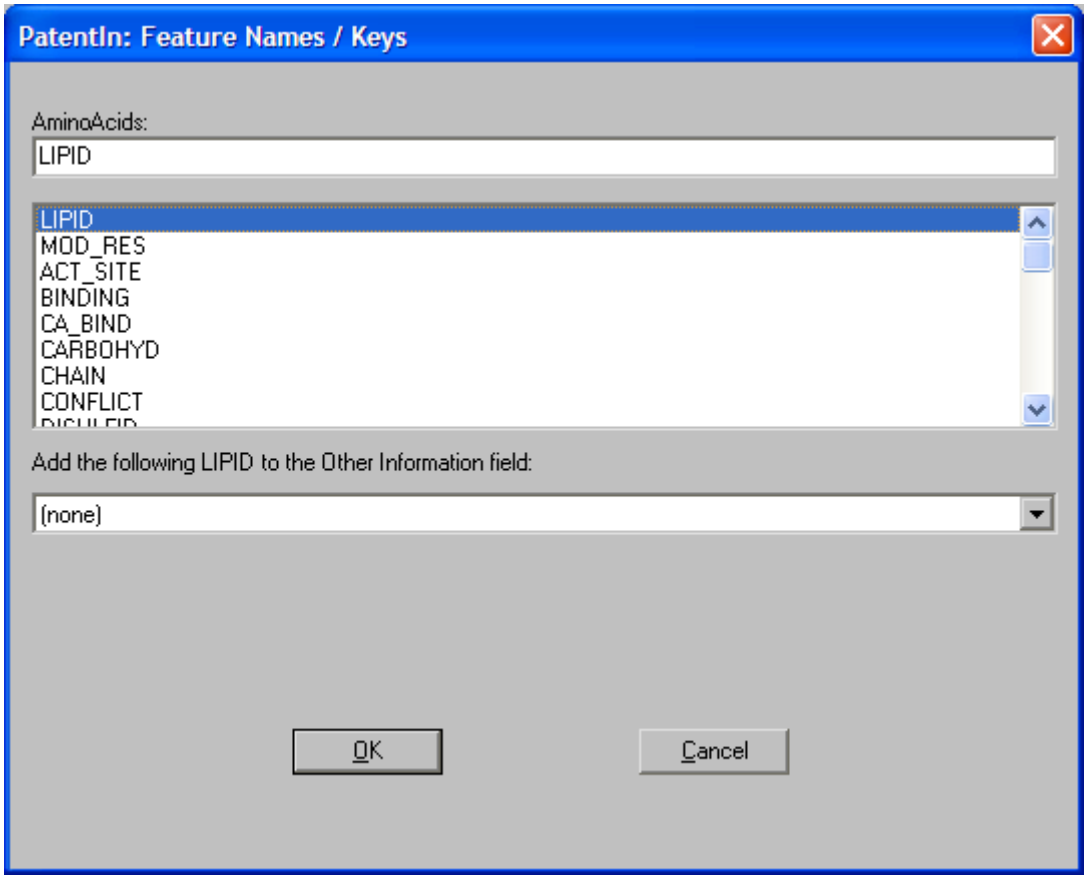

Figure 6-4: Feature Names/Key Selection Screen with LIPID Selected

To select an amino acid name:

1. Select the **Names** button on the Feature Screen (Figure 6-1).
2. Begin typing on the **Amino Acid** field.
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 Feature Screen (Figure 6-4).

To add information about a LIPID:

1. Select the pull down arrow on the box marked Add the following LIPID to the Other Information field.
2. Select the LIPID information from the list.
3. Click on the **OK** button (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.

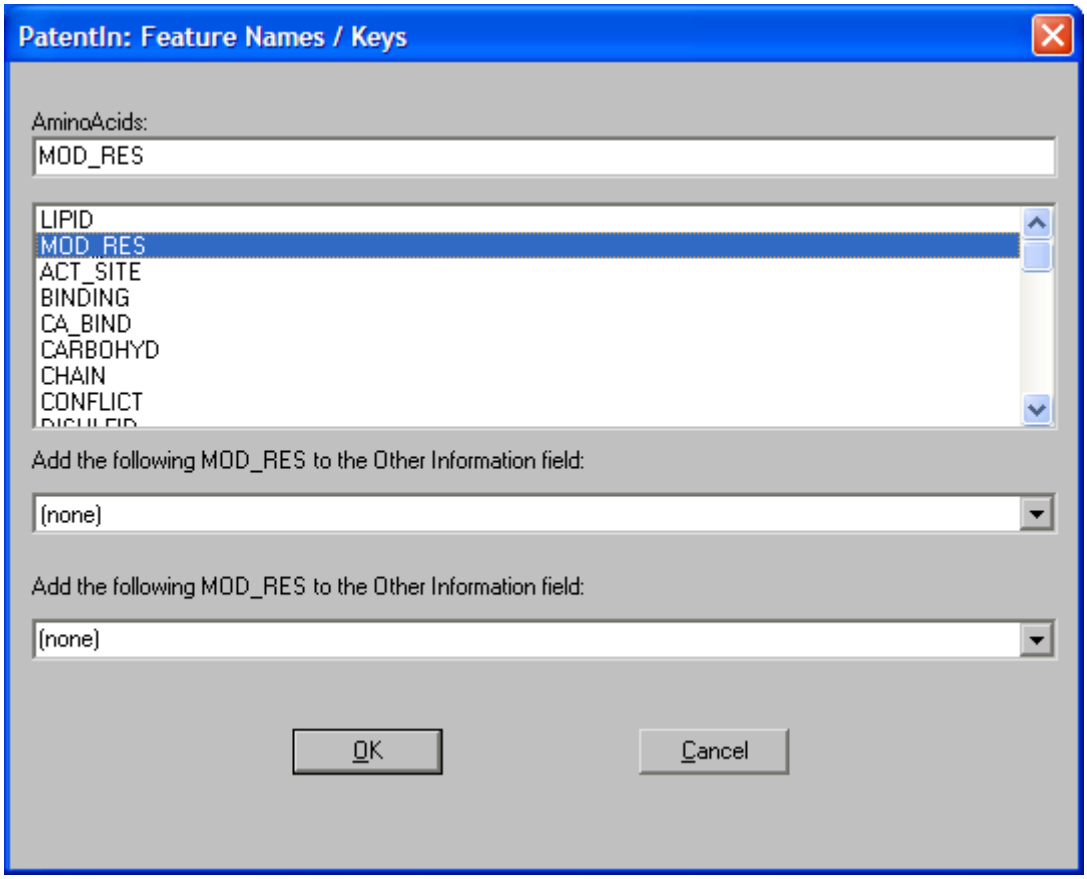

Figure 6-5: Feature Names/Key Selection Screen with MOD_RES Selected

To select an amino acid name:

1. Select the **Names** button on the Feature Screen (Figure 6-1).
2. Begin typing on the **Amino Acid** field.
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 the **OK** button (Figure 6-5).

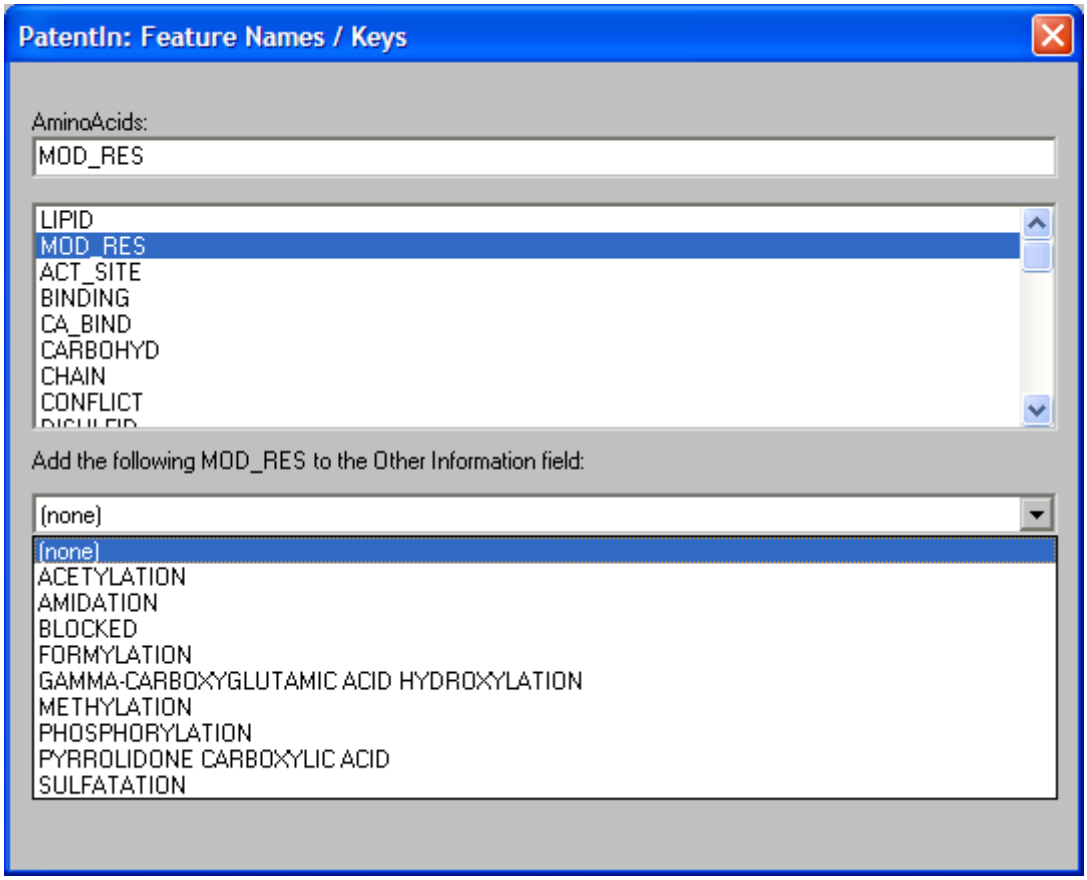

Figure 6-6: First MOD_RES Pull Down List

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 the **OK** button (Figure 6-5).

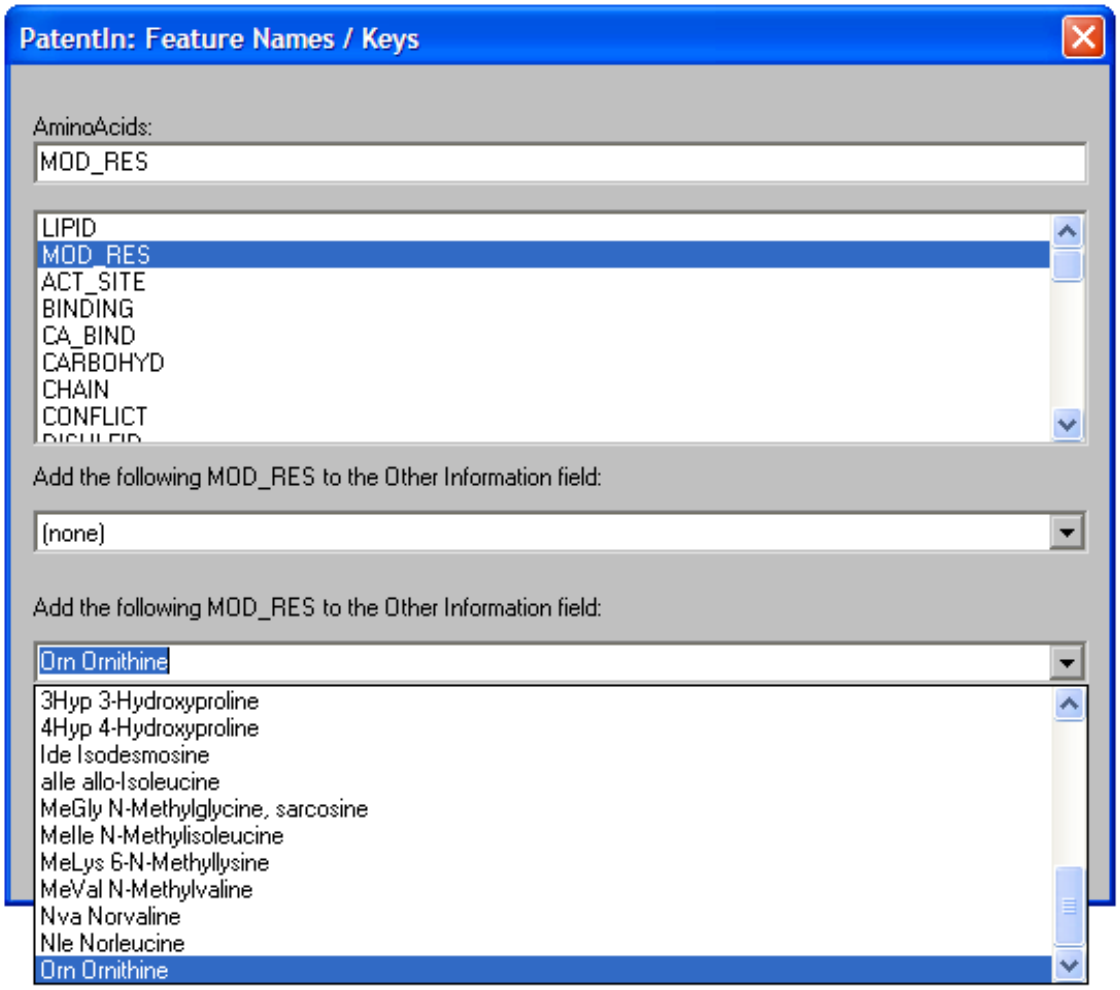

Figure 6-7: Second MOD_RES Pull Down List

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.

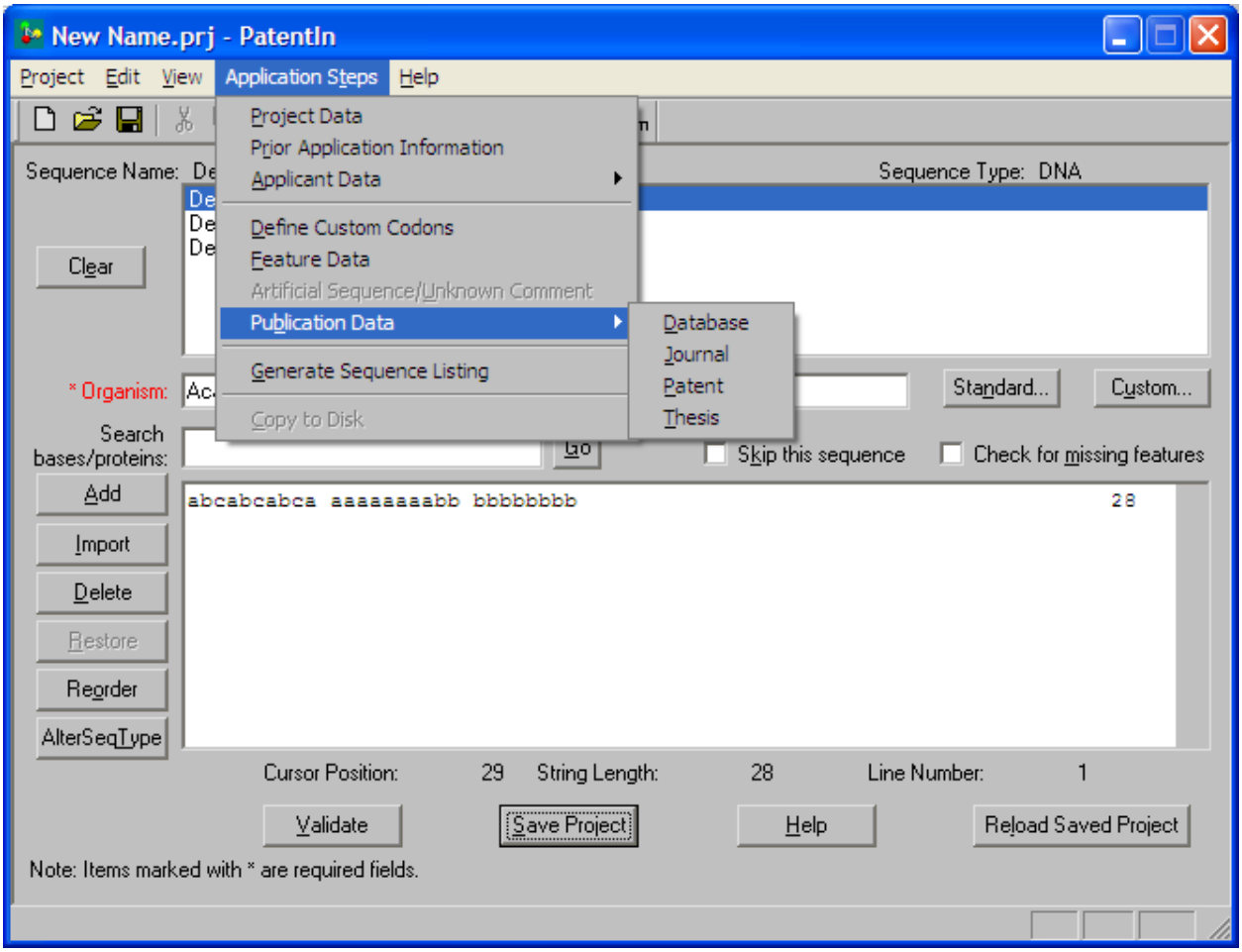

Figure 7-1: Publication Type Screen

To select a publications type:

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

7.2 Journals Publication Information

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

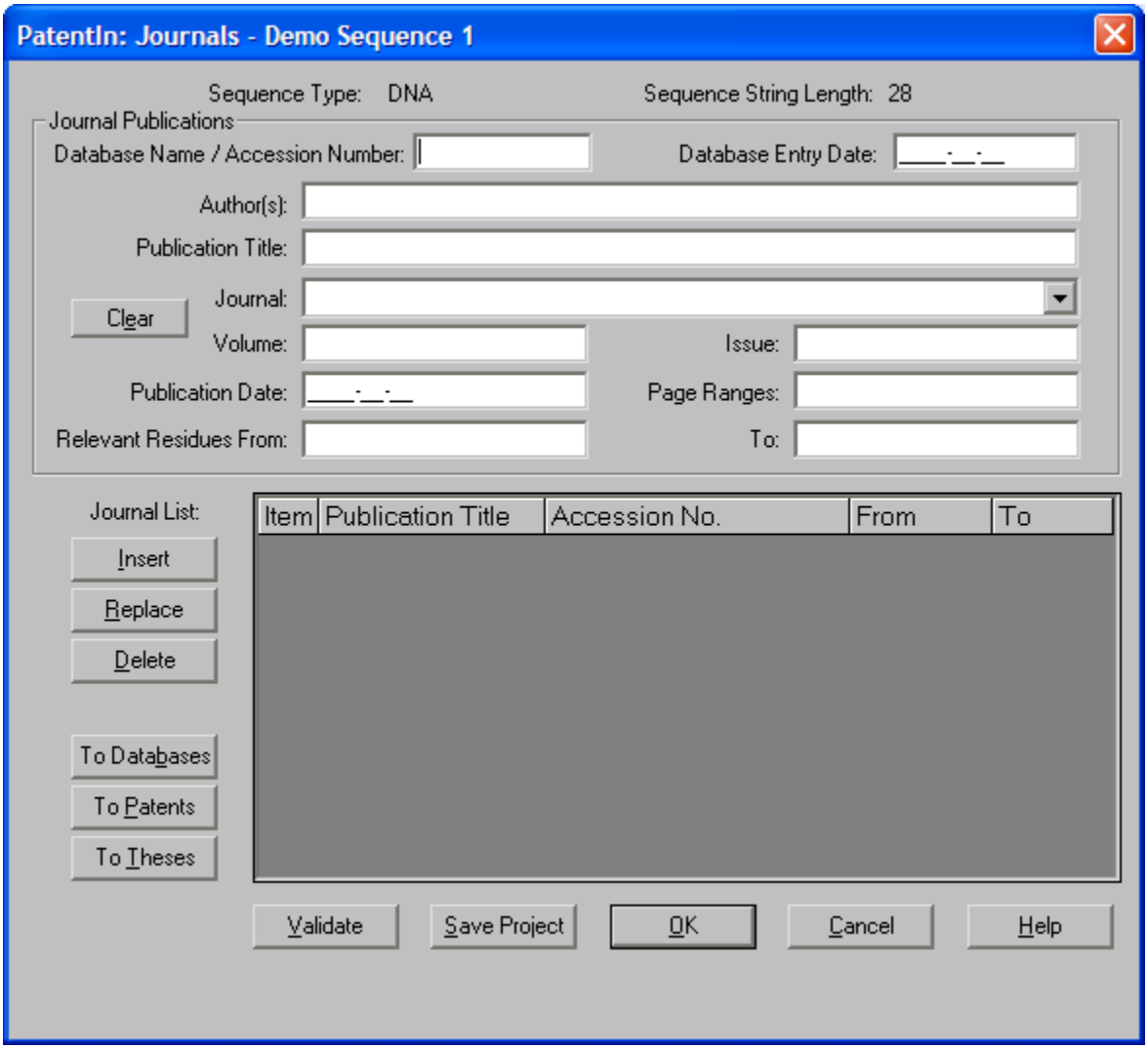

Figure 7-2: Journals Publication Information Screen

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. 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 other **Publication Data**, click on the **To Databases, To Patents or To Theses** 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.

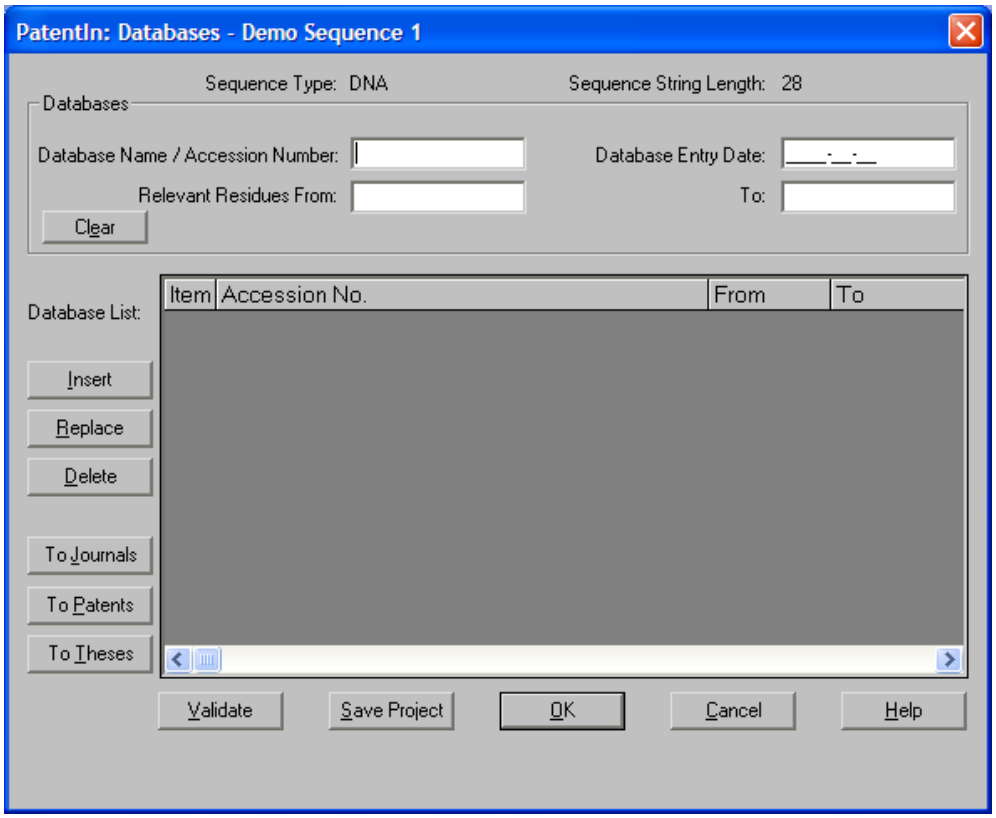

Figure 7-3: Databases Publication Information Screen

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**, 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 proceed to other Publication Data, click on the **To Journals**, **To Patents,** or **To Theses** 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.

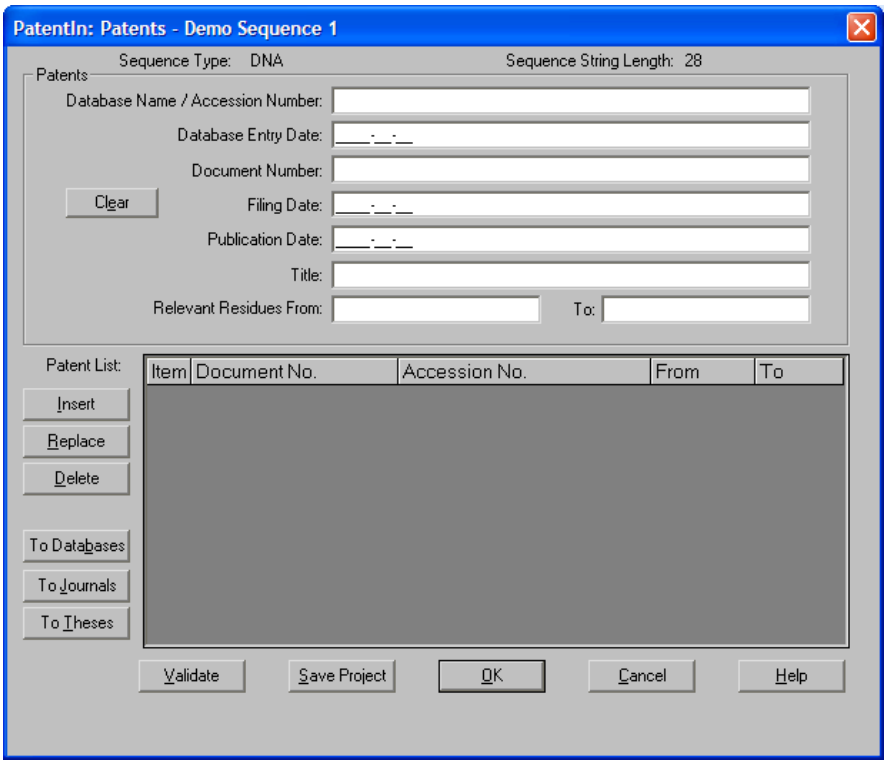

Figure 7-4: Patents Publication Information Screen

To enter information about a patent publication:

1. From the Application Steps select Publications Data, and 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 proceed to other **Publication Data**, click on the **To Databases, To Journals** or **To Theses** button.

7.5 Theses Publication Information

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

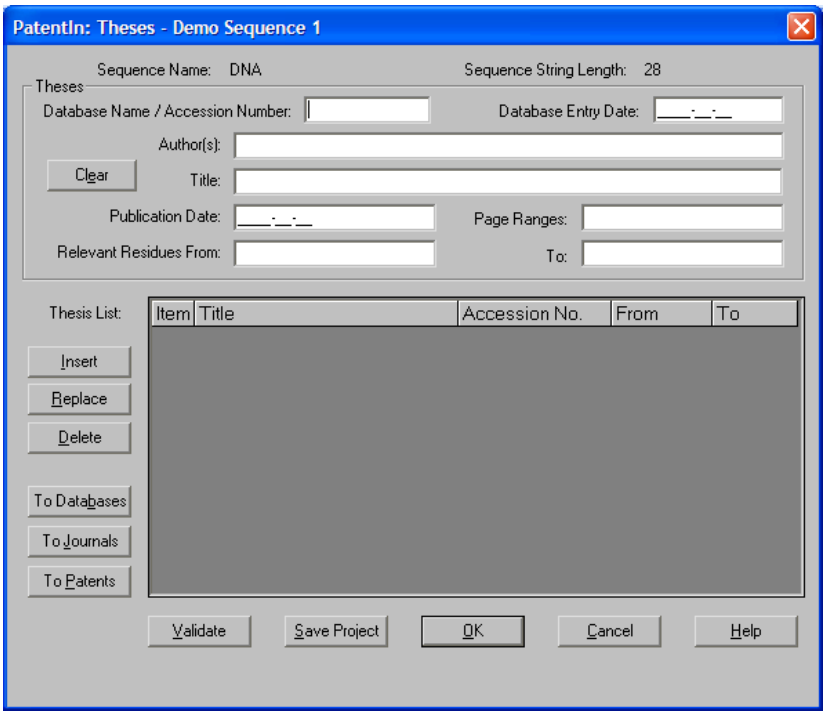

Figure 7-5: Theses Publications Information Screen

To enter information about a thesis publication:

1. From the Application Steps select Publications Data, and then select Thesis.
2. Enter the **Database Name/Accession Number**.
3. Enter the **Database Entry Date**.
4. Enter the name of the **Author(s)**.
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 proceed to other **Publication Data**, click on the **To Databases**, **To Journals**, or **To Patents** button.

Section 8 Creating a Sequence Listing Project File

8.1 Sequence Listing File

The sequence listing file includes all the information required by ST.25. PatentIn 3.5.1 will generate a sequence listing in text format. The name of the sequence listing is consisted of the current PatentIn project name appended with "_ST25.txt." 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.

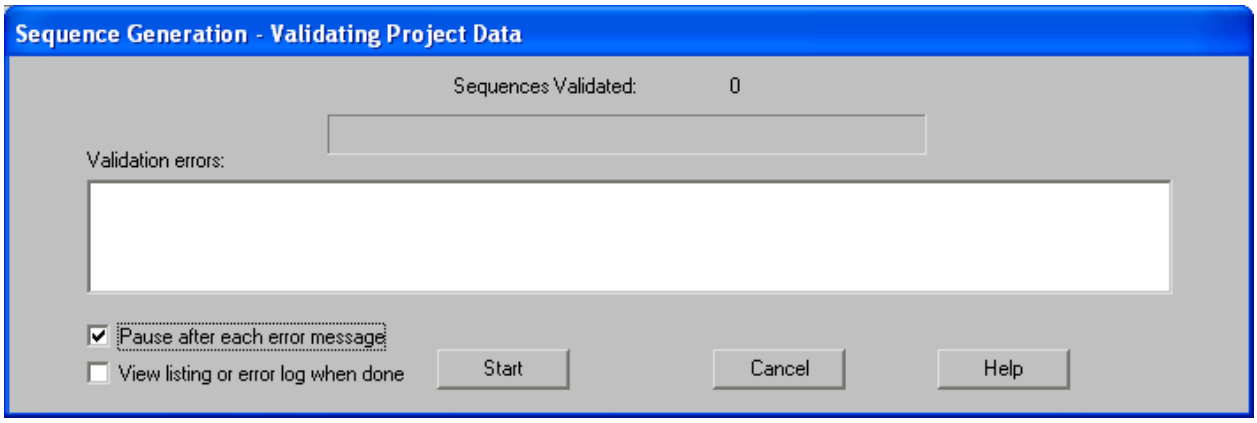

Figure 8-1: Sequence Generation Screen

To generate a sequence listing:

1. Select **Generate Sequence Listing** from the **Application Steps** menu in Figure 4-1 or by selecting the **Gen** button on the PatentIn toolbar.
2. Click in the box next to the "Pause after each error" 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.
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.

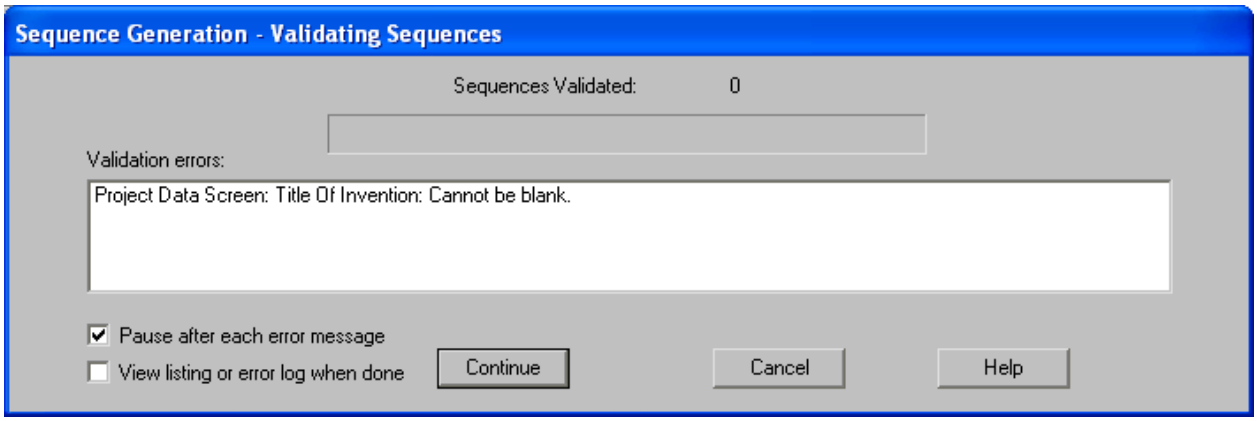

Figure 8-2: Second Sequence Generation Screen

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/Print Sequence Listing** or **View/Print Error Report** from the **Project** menu.

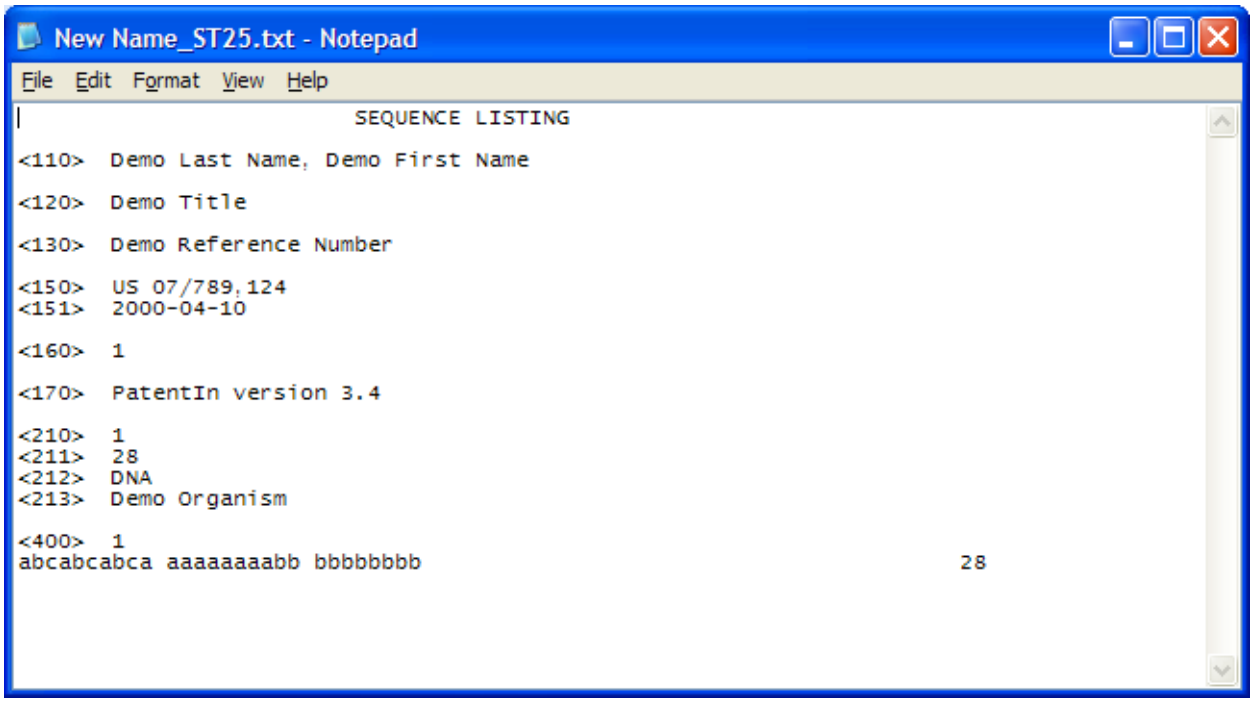

Figure 8-3: View Results Window

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 the version is 7.2 (SR-2). USPTO successfully viewed 60 MB and 120 MB files using an earlier version of the V software and that it was on a laptop with Windows 98 (LocalAdmin may be required for

installation). The USPTO 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.

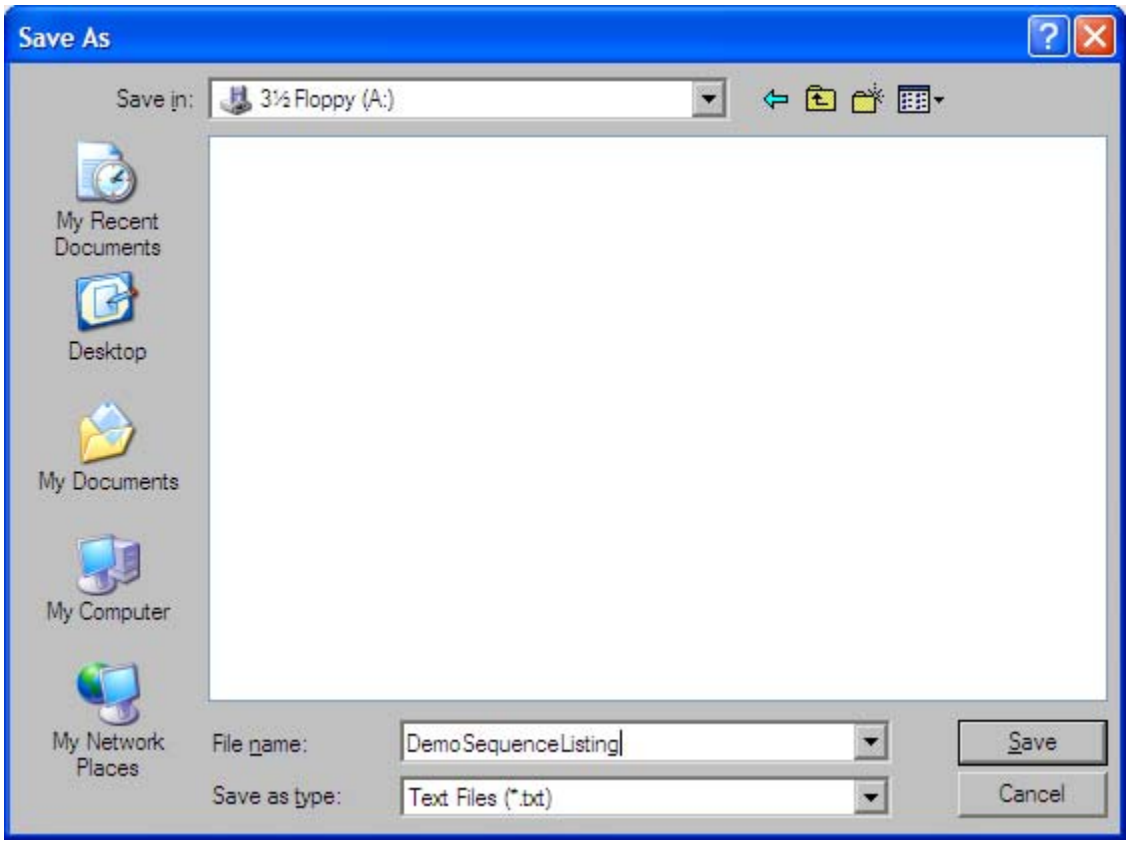

Figure 8-4: Copy to Disk Screen

To copy a sequence listing:

1. Select **Copy to Disk** from the **Application Steps** menu in 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: The .zip file works only with floppy disk(s), hard drive, and network drive.

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 Compact Disk (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.

Section 9 List of Acronyms

Table 9-1 lists the specific acronyms that are used within this document.

Table 9-1: Acronyms

Acronyms	Definitions
AN	Applicant Name
ASCII	American Standard Code for Information Interchange
CDS	Coding Sequence
DID	Data Item Description
DLL	Dynamic Link Library
DNA	Deoxyribonucleic Acid
DOS	Disc Operating System
EPO	European Patent Office
ET	Equatorial
FQT	Functional Qualification Test
LTR	Long Terminal Repeat
MB	Megabytes
MS	Microsoft
PC	Personal Computer
RBS	Ribosome Binding Site
STS	Sequence Tagged Site
UM	User's Manual
US	United States
USPTO	United States Patent and Trademark Office
WIPO	World Intellectual Property Organization
WPI	Web PatentIn
XP	Microsoft eXPerience

Appendix A Field Identifiers, Lengths and Types

Table A-1 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 A-1: Field Name, Identifier, Size, and Type

Field Identifier	Field Name	Field Length	Field Type 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
<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

Appendix B Country Codes

Table B-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 B-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
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

Code	Country
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
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
HK	Hong Kong
HU	Hungary

Code	Country
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
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

Code	Country
NO	Norway
OM	Oman
PK	Pakistan
PA	Panama
PG	Papua New Guinea
PY	Paraguay
PE	Peru
PH	Philippines
PL	Poland
PT	Portugal
QA	Qatar
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
TH	Thailand
TG	Togo
TO	Tonga
TT	Trinidad & Tobago

Code	Country
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

Appendix C Conversion Table Between Nucleotide Triplets (Codons) and One- and Three Letter Amino Acid Codes

Table C-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 C-1: Exchange Nucleotide Characters and Amino Characters

PRT/1	PRT/3	Nucleotide Equivalent
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
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”

Appendix D Nucleotide Sequence Features

Table D-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 D-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
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=gggcgg
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

Key	Description
	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)
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'

Key	Description
	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)
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
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

Key	Description
	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
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 [ ]

Appendix E Amino Acid Sequence Features

Table E-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 E-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
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

Key	Description
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)
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

Appendix F Data Tables for Mod_Res Sequence Features

Appendix F 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 F-1) and second **Add the following MOD-RES to the Other Information field** for MOD_RES (Table F-2) sequence features appears. You can select from either one or both of the **Add the following MOD-RES to the Other Information field**s, and the data will appear in the **Other Information** field (<223>).

Table F-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 F-2: Second Data Table for MOD_RES Sequences

Symbol	Meaning
(none)	Blank space (default option)
Aad	2-Aminoadipic acid
bAad	3-Aminoadipic acid
bAla	beta-Alanine, beta-Aminopropionic acid
Abu	2-Aminobutyric acid
4Abu	4-Aminobutyric acid, piperidinic acid
Acp	6-Aminocaproic acid
Ahe	2-Aminoheptanoic acid
Aib	2-Aminoisobutyric acid
bAib	3-Aminoisobutyric acid
Apm	2-Aminopimelic acid
Dbu	2,4 Diaminobutyric acid
Des	Desmosine
Dpm	2,2’-Diaminopimelic acid

Symbol	Meaning
Dpr	2,3-Diaminopropionic acid
EtGly	N-Ethylglycine
EtAsn	N-Ethylasparagine
Hyl	Hydroxylysine
aHyl	allo-Hydroxylysine
3Hyp	3-Hydroxyproline
4Hyp	4-Hydroxyproline
Ide	Isodesmosine
aIle	allo-Isoleucine
MeGly	N-Methylglycine, sarcosine
MeIle	N-Methylisoleucine
MeLys	6-N-Methyllysine
MeVal	N-Methylvaline
Nva	Norvaline
Nle	Norleucine
Orn	Ornithine

Appendix G Additional Lipid Sequence Features

Table G-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 G-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

Appendix H Acceptable Characters In The Sequence Description Field

Table H-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 H-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

Appendix I Additional Modified_Base Sequence Features

Table I-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 I-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
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

Symbol	Meaning
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

Appendix J Technical Notes

J.1 Microsoft® Access Notes

(Section J.1 is only applicable to PatentIn with versions prior to PatentIn 3.4.)

PatentIn 3.1, PatentIn 3.2 and PatentIn 3.3 were packaged with a Microsoft® Access 97 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.

J.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).

J.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.

J.2 General Hints and Tips

1. The installation requires some Dynamic Link Library (DLL) registration for each machine.
2. PatentIn 3.5.1 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 XP defaults to not showing the underlines on the toolbar without pressing the Alt-key.
8. To access PatentIn menu without a mouse, do the following:
    a. Press the Alt key to activate the menu.
    b. Release the Alt key.
    c. Press an access key from the main menu to select a pull-down menu.
    d. Press an access key from the pull-down menu to complete the action.
9. To activate a button on PatentIn screen without a mouse, do the following:
    a. Press and hold the Alt key.
    b. Press the access key of the button.

Although no future enhancement releases are envisioned, maintenance releases may be necessary; however, unreported problems cannot be fixed.

J.3 Installation and Testing Notes

J.3.1 PatentIn Installation and Repair

PatentIn can be downloaded onto your computer from the USPTO web page http://www.uspto.gov/web/offices/pac/patin/patentinrel.htm. Before the installation of PatentIn 3.5.1, be sure to close all running applications on your computer and then follow the instructions found on the Web page to download PatentIn 3.5.1 and install it on your PC.

A user will be asked to reinstall PatentIn 3.5.1 while trying to run PatentIn 3.5.1. This occurs when PatentIn 3.5.1 detects the registry keys required to run PatentIn 3.5.1 are missing. This may be due to uninstalling an earlier version of PatentIn. Reinstalling PatentIn 3.5.1 will rebuild the required registry keys to run PatentIn 3.5.1 and it will not cause you to loose any previously saved PatentIn data. If a user starts the PatentIn 3.5.1 reinstallation without uninstalling PatentIn 3.5.1 first, the following screen will be prompted to the user. Select the "Repair" radio button as shown in Figure J-1, press the "Next" button and follow the installation instructions to complete the PatentIn repair operation. A user may also choose to uninstall PatentIn 3.5.1 first and then reinstall the PatentIn 3.5.1 software (Figure J-1).

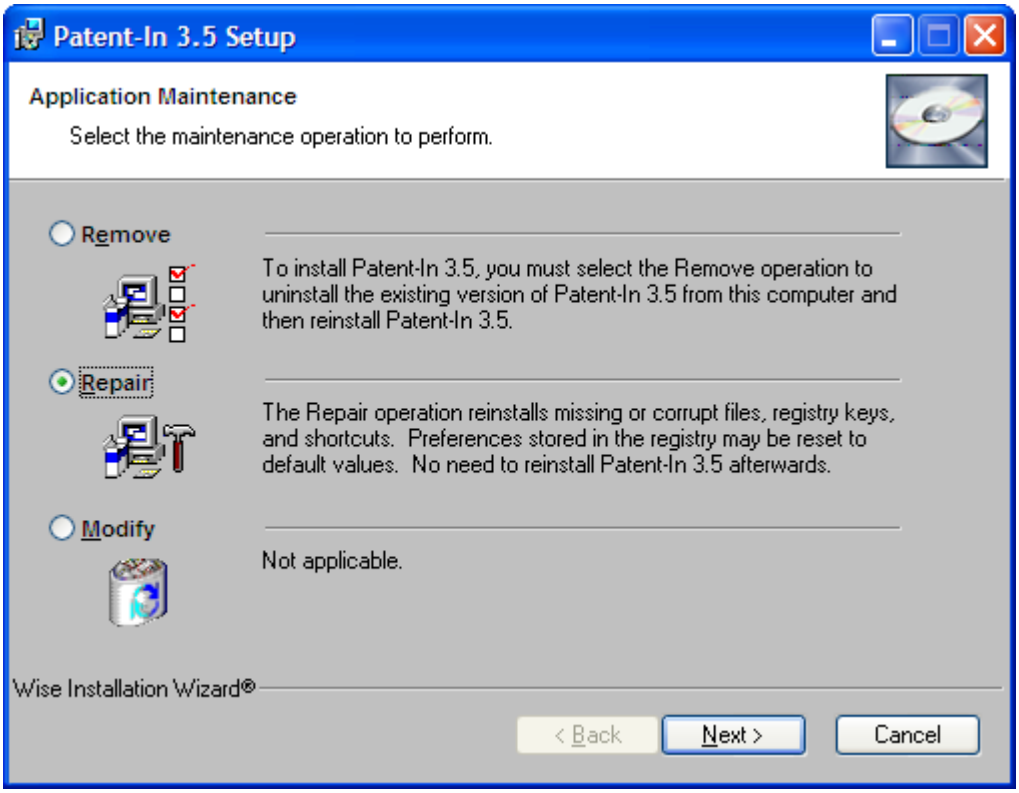

Figure J-1: PatentIn Repair and Uninstall

J.3.2 Testing Configurations

During testing the following configurations were used:

- Microsoft® Windows XP Professional

- Version 2002
- Internet Explorer 6.0.2800.1106.

- Microsoft® Windows Vista Ultimate, Microsoft® Windows Vista Home Basic and Microsoft® Windows Home Premium

- Processor : Intel Core 2 CPU 6400 @ 2.13GHz
- Memory : 1006MB
- System Type : 32-bit Operating System
- Internet Explorer 6.0.2800.1106.

- Microsoft® Windows 7 Enterprise

  - Version : 2009
  - Processor : Intel® Core™2 CPU 6400 @ 2.13GHz
  - Memory : 1006MB
  - System Type : 64-bit Operating System

- Internet Explorer : 8.0.7600.16385.

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

J.3.3 Internet Explorer Considerations

Internet Explorer ships with several of the DLLs that either Checker 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 MS Windows 98. The user quickly isolated the problem and informed USPTO that an upgrade to Internet Explorer 5.0 eliminated the problem.