BRUKER

topspin

Bruker BioSpin

Protein Dynamics Center

Software Manual
Version: 1.1.3

thinkforward

NMR Spectroscopy

Protein Dynamics Center

Version: 1.1.3

Software Manual

P/N: H121563

November 30, 2010

Bruker software support is available via phone, fax, e-mail, Internet, or ISDN. Please contact your local office, or directly:

Address:	Bruker BioSpin GmbH Service & Support Department Silberstreifen D-76287 Rheinstetten Germany
Phone:	+49 (721) 5161 6456
Fax:	+49 (721) 5161 91 6456
E-mail:	nmr-software-support@bruker.de
WWW:	*www.bruker-biospin.com*
FTP:	ftp.bruker.de / ftp.bruker.com

Computer typset by Bruker BioSpin GmbH, Rheinstetten 2007.

Contents

Contents

Contents

1 Introduction

The **Protein Dynamics Center** is designed for the analysis of series if 2D or pseudo 3D spectra which yield information about protein dynamics on various time scales. Fast motions (ps - ns) arising from bond motions or motions of individual atoms relate to the relaxation parameters Hetero-NOE, T_1, T_2 and T_{1rho}. Intermediate motions (ns - us) arising from motions of larger groups of atoms and slower motions (us - ms) arising from conformational changes relate to R_{ex}.

The analysis of these dynamic parameters is done in a 'method oriented' way, i.e. there is a NOE method, T_1 method etc. Each method consists of 6 steps: **Sample** set-up**, Data** selection, **Analysis**, **Viewing**, **Reporting** and **Exporting**. All methods are accessible in a **method tree** in which nodes represent the individual methods and leafs the individual steps,

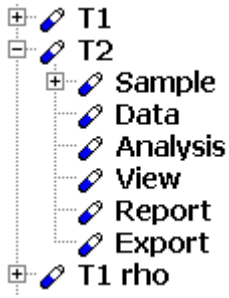

Figure 1.1 A view to the method tree, the node of the T_2 method opened

Some steps need to be executed (Sample, Data, Analysis), others are optional (View, Report, Export).

A method explicitly applied to a sample and data is called a **project**. Projects typically include details of calculations, selected parameters and results. Projects can be saved to project files and can be reloaded. Several projects can be executed at the same time.

The Protein Dynamics Center is a pure Java application. The GUI shows standard components like the method tree, file explorer tree and menu. Embedded is a very powerful multiple object viewer which is used for all visualization tasks. Furthermore contained (but not visible) is a calculation engine used for the numerical parts, IO functions and converters for import and export tasks.

Availability and Outlook

The development of the Protein Dynamics Center is ongoing. Two versions have been made available so far.

version 1.0, basic relaxation experiment analysis

version 1.1, advanced analysis, model free modelling

The major limitation so far is that the model free modelling only uses data at one field strength. Using multiple field strength as well as other improvements like the usage of user defined spectral density functions will be offered with the next version of the software.

2 How to install the Protein Dynamics Center

2.1 Availability

As a pure Java application the Protein Dynamics Center runs on all platforms able to execute a java run time environment (**jre**). Available versions exist for Windows XP, Vista, Win7 and Linux CentOS-5 systems. The software was compiled with java 1.6.xxx. As java runtime versions we recommend jre_1.6_16 or later.

2.2 Installation from a Bruker software DVD

The Protein Dynamics Center is a component of the Bruker Software Suite and will be available*) on the Bruker software DVD. It can be installed completely independently from any other components available on this DVD. This type of installation includes a suitable version of jre.

*) planned later in 2010

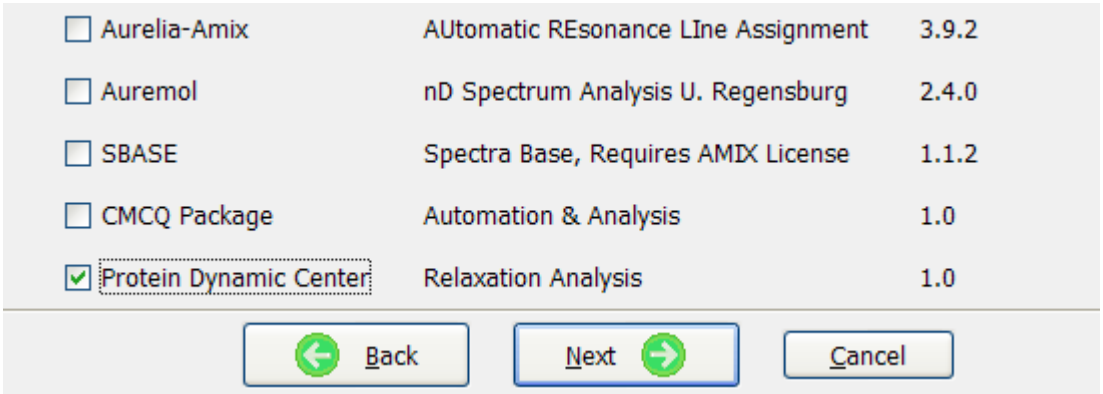

Figure 2.1 The Protein Dynamics Center is among the Bruker software products

2.3 Installation from the internet

The Protein Dynamics Center can be downloaded from the Bruker web site. **(http://www.bruker-biospin.com/software_nmr.html).** Check for PC Protein Dynamics Center and Linux Protein Dynamics Center.

Available download files are:

Windows	***ProteinDynamicCenter1.0_jre.exe***	full version including jre
	ProteinDynamicCenter1.0.exe	full version
Linux	***ProteinDynamicCenter1.0_jre.sh***	full version including jre
	ProteinDynamicCenter1.0.sh	full version

Jre is the java run time environment. It is usually existing on the computer (e.g. also part of the TopSpin installation). But to be sure that the correct version of jre is used it is mandatory to download the full version of the Protein Dynamics Center including jre once and install it. In the following downloads without jre are sufficient. All downloads include the documentation and a number of third party libraries.

To install the software just execute the downloaded file. You may need administrative rights to do so. An installer pops up which guides you through the installation. The Protein Dynamics Center will be installed in the folder you specify (e.g. C:\Bruker\ProteinDynamicCenter), the registration is performed and an entry in the **Start/Programs/Bruker TopSpin** (Windows) or **Applications/Bruker TopSpin** (Linux) folder is created.

2.4 Help

Manual, About and License information is available at run-time by clicking at the ***Help*** Button

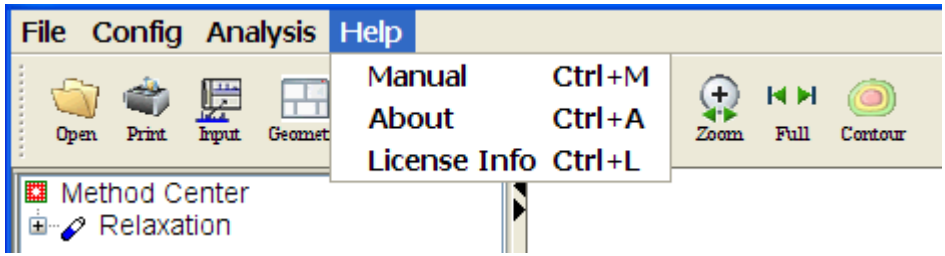

Figure 2.2 The Help pull-down menu

The **Adobe Reader** is used for the display of any pdf files. If not existing on the computer you may use ***Config/Preferences/Default PDF viewer*** and define an alternative program to view pdf files, e.g. ***xpdf*** on Linux systems. The manuals itself are stored in /xxx/docu/ where xxx is the installation directory of the Protein Dynamics Center.

Due to our company policy a ***License Agreement*** is provided. It contains license information of Bruker software products as well as third party products. The corresponding text file is topspin_license.txt and is stored in the license folder which is part of the Protein Dynamics Center Installation folder.

3 License Information

The Protein Dynamics Center is protected by a **FLEXlm** (www.macrovision.com) license. The **license key** is called **PROTEIN_DYNAMIC1**. It can be obtained through the Bruker web site **https://www.bruker-biospin.com/nmr_license_requests.html** or from the Bruker license department (**license@bruker.de**).

The license is currently a **free, node locked demo** license which expires April 30th, 2011. Skip reading sections below, especially 3.3, 3.6, 3.7, that refer to floating licenses and license managers.

3.1 General

There are two types of licenses. **Node locked licenses** are made for individual computers. The licensed software can then run only on these.The license checking is done by the Protein Dynamics Center itself. A node locked license is ready to be used if it is put into a particular license file (see below)

Floating licenses are network licenses. One (or three) computers in the network are defined as license servers. The floating licenses are bound to the license server(s) but any computers in the network can use the licenses as long as free licenses are available. The FLEXlm license manager that controls the licenses must run on the server(s). Usually, the installation of FLEXlm installs and starts a corresponding service. FLEXlm must also be installed on other computers in the network if license shall be used from these. A floating licenses is ready to be used if FLEXlm is installed on all involved computers and the license is put into a particular license file on all these computers (see below)

3.2 The license file (license.dat)

The license file (license.dat) is a text file with a particular format and stored at a defined location on disk. Usually this is ***/usr/local/flexlm/Bruker/licenses/license.dat*** on Linux systems and ***C:\Flexlm|Bruker\licenses\license.dat*** on Windows systems. A default license file is created when installing FLEXlm.It only needs to be updated if licenses are added to it. An example of a license file may look like

:

```
SERVER   nike   6906d704 1700

DAEMON   bruker_ls   /usr/local/flexlm/Bruker

FEATURE  PROTEIN_DYNAMIC1 0.0 26-apr-2010 1\
BBCE384950ACB123476CBE ""
```

The first line identifies the license server by name and hostid. Lines 2 to N contain individual licenses. Line 1 and 2 must be available but are not evaluated if only node locked licenses follow.

If you want to keep the license file not in the default location but somewhere else on disk you must define an environment variable and set it correspondingly.

Linux:

(Bourne Shell):

```
LM_LICENSE_FILE=/u/Bruker/licenses/license.dat
export LM_LICENSE_FILE
```

or (C Shell)

```
setenv LM_LICENSE_FILE  "/u/Bruker/licenses/license.dat"
```

Windows:

Execute **Start/Settingd/Control Panel/System** and click the Advanced tab
Define **LM_LICENSE_FILE** among the environment variable**s**

3.3 Starting the license manager)

The license manager only needs to run on license server(s) if floating licenses are to be used. It is usually started during the installation. If it needed it can be started by hand:

Linux:

```
cd /usr/local/flexlm/Bruker

lmgrd start -c /usr/x/flexlm/Bruker/licenses/license.dat
```

If the license manager already runs and you want to stop and restart it, then

```
lmgrd stop -c /usr/x/flexlm/Bruker/licenses/license.dat

lmgrd start -c /usr/x/flexlm/Bruker/licenses/license.dat
```

can be executed.

Windows:
Use ***Services*** in the ***Control Panel*** (possibly among the **Administrative Tools** (Windows 2000)) to stop and restart the FLEXlm License Server. Normally, the license manager would show up with the **Startup** attribute set to **Automatic**: A double click to a service allows starting and stopping the service or the definition of startup parameters.

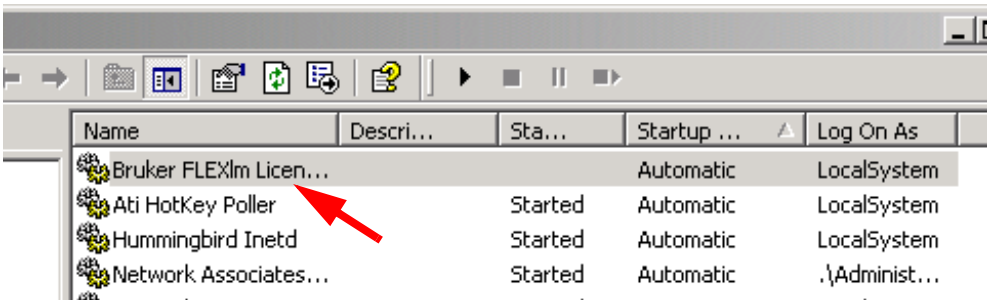

Figure 3.1 Listing of installed system services

There is also a tool called **lmtools.exe** coming with the FLEXlm installation. It resides in the FLEXlm installation directory, e.g. c:\flexlm\bruker and offers a graphical user interface. All actions like changing settings, re-reading license files, stopping/starting servers etc. can be done elegantly with this tool

3.4 Apply for licenses, first get the correct hostid

In case of node locked licenses the **hostid** of the computer on which the Protein Dynamics Center shall run must be supplied. In case of floating licenses the hostid of the license server(s) is needed. Most easily, first install the Protein Dynamics Center, start it and check Help/About. The correct hostid is shown. Alternatively, you may use system tools, e.g. ***ipconfig /all*** on Windows systems. The hostid shows up under **Physical Address**. On Linux systems the shell command **/etc/sysinfo** provides the hostid.

3.5 Obtain licenses, add them to the license file

Licenses are typically sent by e-mail together with instructions to add them into the license file. In principle you can use any text editor and add the licenses. When doing so make sure that you don't change the format of the file, especially don't add special (non-printing) characters. Such characters may accidentally come in via the e-mail programs. Also note that the licenses (so-called feature lines) are typically very long and line breaks (the \ character) are inserted for readability. These line breaks must not be taken out

In case of node locked licenses put the obtained license into the license file on the machine on which the Protein Dynamics Center shall run. It is ready to use immediately.

In case of floating licenses put the obtained license into the license file(s) on the license server(s). The obtained license must also be put into the license file(s) on all computers in the network on which the Protein Dynamics Center shall run. The running license manager on the license servers must then be notified that the license files changed. Newer versions of FLEXlm handle this automatically. If for some rea-

son this seems to not happen, execute the ***lmutil lmrerad*** tool, see below.

3.6 FLEXlm version numbers

With TOSPIN 2.0 and later FLEXlm version **9.2** is delivered. The Protein Dynamics Center is prepared correspondingly. In general, the version of the license server must be larger or equal than the version of the client libraries linked to the Protein Dynamics Center, i.e. must be **>= 9.2**.

3.7 FLEXlm Diagnostics

Unfortunately, problems with licenses are quite frequent. A typical example is that users first work with a node locked demo license then buy a floating license, install it and it doesn't work. Analysis of problems in the past years shows that the main reasons are

- The license file or license string got corrupt when adding a license
- Not all relevant license files have been updated
- The license manager is not running at all
- The license manager is running but didn't see the change of the license file

Less frequent reasons are

- The license string is corrupt as such
- The user supplied the wrong hostid when requesting the license
- There is a clash in version numbers

The FLEXlm installation folder (e.g. c:\flexlm\bruker) contains a utility tool called **lmutil**. It can be executed from a command shell and fed with several arguments, e.g.

lmutil lmreread can be used to inform the license manager if license.dat changed
lmutil lmstat -a display the current status, e.g. which licenses are checked out
lmutil lmdiag provides detailed diagnostic output

If the interactive **lmtools** program is running on your system you may use it instead.

We recommend to check **Help/About** in the Protein Dynamics Center. It shows the hostid of the computer, the location of the license file and available licenses. Comparing this with the output of lmutil lmdiag may already explain the problem.

If the problem persists you may send the output of lmutil lmdiag to Bruker via e-Mal (for e-mail addresses see Help/About) and briefly describe the problem. If the problem can still not be solved Bruker will send back a more detailed diagnostic tool which you again execute and send the output to Bruker.

Finally, there is documentation provided by FLEXlm itself, e.g. the document ***enduser.pdf*** which is the ***help*** folder in the FLEXlm installation directory. You may also check the Macrovision Web pages, www.macrovision.com.

4 How to start the Protein Dynamics Center

4.1 Starting the Protein Dynamics Center on Linux workstations

On Linux computers click at the program icon **Protein Dynamics Center** on the **desktop**, or execute ***Applications/Bruker TopSpin/Protein Dynamics Center***

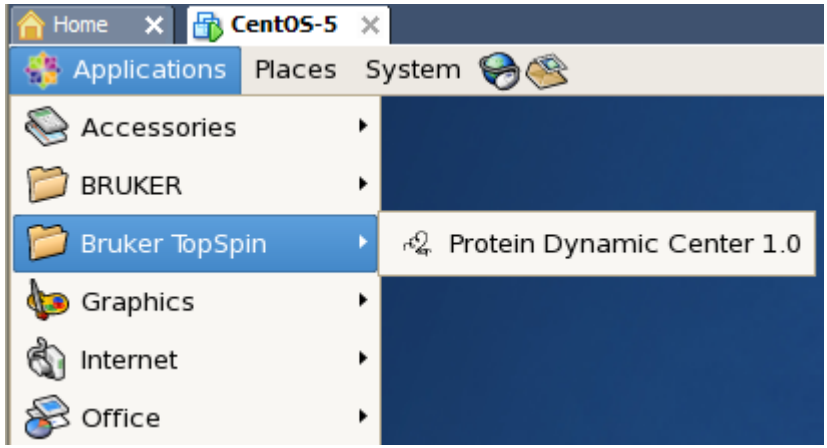

Figure 4.1 Starting the ***Protein Dynamics Center*** via Applications/Bruker TopSpin

Alternatively, one may use an **Explorer** (or shell window), navigate to the installation folder (e.g. /opt/Bruker/ProteinDynamicCenter) and execute the start-up script ***runDynamics***.

4.2 Starting the Protein Dynamics Center on the PC

On the PC click at the program icon **Protein Dynamics Center** on the **desktop**, or execute ***Start/Programs/Bruker TopSpin/Protein Dynamics Center***

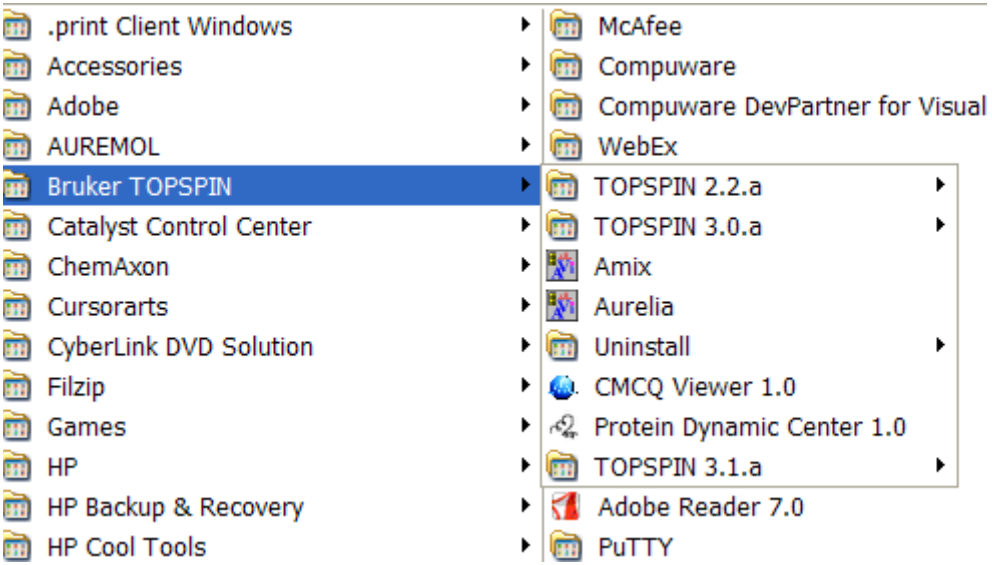

Figure 4.2 Starting the ***Protein Dynamics Center*** via Start Menu/Programs

Alternatively, one may use the **Explorer** (or **cmd** shell), navigate to the installation folder (e.g. c:\Bruker\ProteinDynamicCenter) and execute the start-up script ***runDynamics***.

5 Working with the Protein Dynamics Center

5.1 General aspects

The Protein Dynamics Center offers **t**he analysis of relaxation data in a method oriented way. Different methods (Hetero-NOE, T_1, T_2 and T_{1rho}, $R_{ex,}$ NOE/T_1/T_2) are offered in the **method tree** shown in the method center.

It is possible to switch between the **method tree** and a **file system tree** by clicking at the corresponding tabs in the **Tab Area**. The system tab offers a filtered file tree explorer. Only files understood by the Protein Dynamics Center appear. The file explorer may be used if additional spectra not related to a particular relaxation analysis method shall be loaded.

The **main menu bar** and **icon tool bar** are only used for general purposes such as setting preferences. Not many entries are offered there since the method tree in combination with spectral popup menus offer all tools needed.

> To analyze relaxation data just use the **Method Center** and execute the methods there. The main menu and file system tree are only needed for special actions, e.g. to load other data for comparison or for configuration purposes.

The biggest part of the Protein Dynamics Center window is used for **spectral displays.** Depending on the chosen view settings additional (external) display windows are dynamically created.

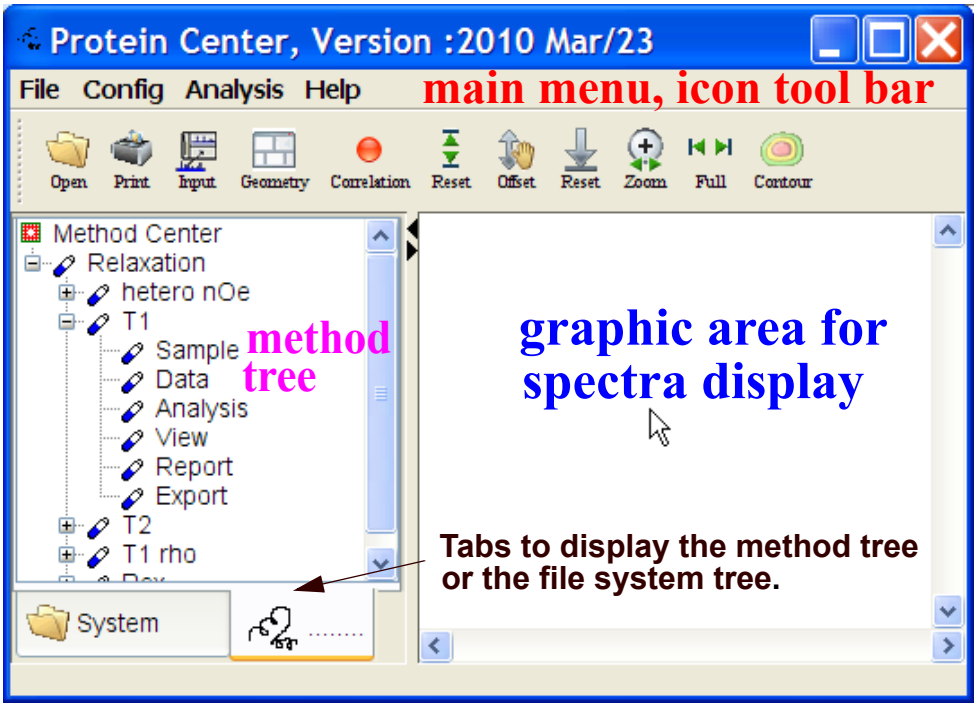

Figure 5.1 The Protein Dynamics Center

5.2 Working with the method tree, methods & projects

To open the method tree click at the ***Relaxation tab***. The method tree operates like a regular file tree explorer. To open/close a method node move the cursor to the +/- fields and click the **left** mouse button. After opening a node the components **Sample**, **Data**, **Analysis**, **View**, **Report** and **Export** become visible. If these are activated with left mouse button clicks dialog windows come up and displays/ calculations are done. All methods contain the same components but are customized to the respective needs.

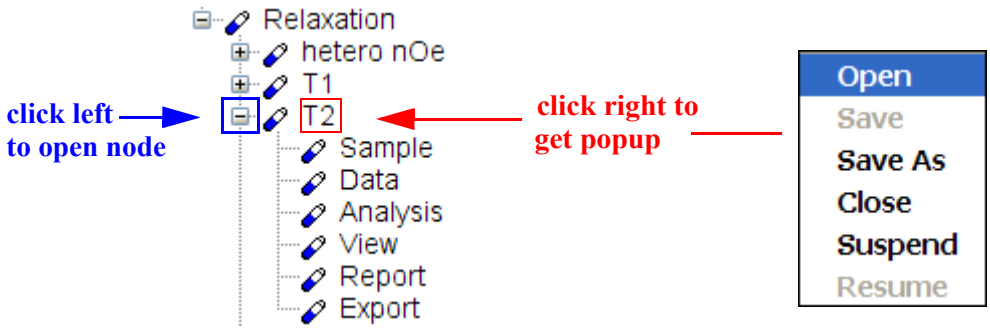

Figure 5.2 Left and right click to the method tree

A method that is applied to a sample and its data is called a **relaxation analysis project**, or for simplicity just **project**. Unfortunately, the word project is used by different software tools with different meanings. When saved to project files different formats and contents are created. In case of the Protein Dynamics Center the project files are **xml files** and contain a large amount of information. Typically, they are not intended to be read by the user but are used to save the status of an analysis and later load it again, e.g. to continue the analysis or just recall the results. Actions that can be performed with projects are available on popup menus. To get the popup menu of a method, move the cursor to the name of the method node and click the **right** mouse button. The following actions are currently known:

- **Open**
  The name of a project file to which a project has been saved with the ***Save*** command must be specified. The project then gets loaded. When activating the components **Sample**, **Data**, etc. on the method tree you get exactly the settings that have been saved.
  It is possible to work with several methods/projects at the same time, e.g. a T_1 and T_2 project but it is not possible to have two projects open for the same method, e.g. two T_1 projects.
  It is allowed to open a project file of a different method, e.g. execute ***Open*** for

T_2 but then specify ***T1.project*** as the project file. This makes sense if the sample is the same and one does not want input this information again. Details in the **Data**, **Analysis** etc. components are typically different and would have to be checked/corrected when traversing the method tree.

- **Save**
  A project will be saved to a specified file. File names and extension may be freely chosen, the default extension is.***project***. We recommend to gather different project files in a folder of your choice, for example c:\tmp\ on Windows systems. A typical example of a project file name would then be c:\tmp\T1.project.

- **Save As**
  This command can be used to save a project to a different file.

- **Close**
  The life time of a project is often identical to the life time of the Protein Dynamics Center, i.e. all data are kept in memory. However, it gets automatically replaced if another ***Open*** to the same method is performed. Thus, it is for example not possible to keep two T_1 projects at the same time. ***Close*** could explicitly be executed to save memory resources.

- **Suspend**
  Working with projects usually includes the display of data and results. Often three display windows are used per method. When working with several methods in parallel the number of display windows increases and the sizes of these windows get smaller and smaller. If, momentarily the windows of a particular method are not really needed on the display one can ***Suspend*** the project. All data and results are kept in memory but the display windows disappear until ***Resume*** is executed.

- **Resume**
  This is the opposite to the ***Suspend*** command, see previous section.

Some general remarks:
What exactly is contained in the project file and what is re-calculated at run-time is determined by the software. Part of the stored information is related to spectra names which are currently kept with absolute path names. If a project file and data are copied to another computer it is advised to maintain the spectra names.

The popup menu is context sensitive. If a project is for example loaded and nothing has been changed, ***Save*** is inactive, or if a project is suspended only ***Wake up*** is active.

5.3 Working with the system tree

To open the method tree click at the ***System tab***. The regular file system tree is initially shown but only shows folders which contain spectra, optionally molecules and peak lists. A popup menu is shown after a right mouse button click on the tree:

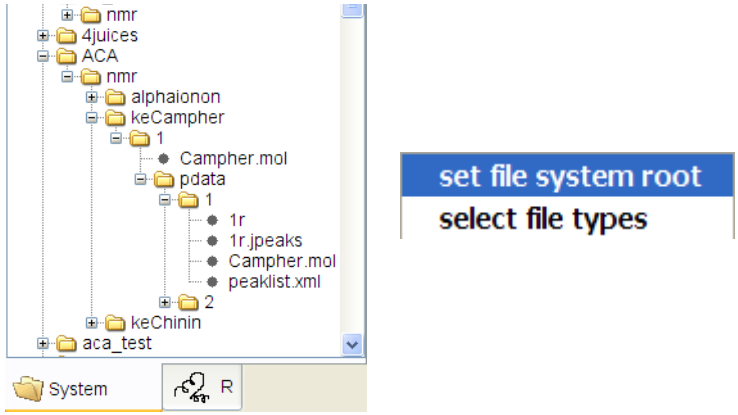

Figure 5.3 File system tree with popup menu

- **set file system root**
  Can be used to change the root of the tree. If for example all data are stored in d:\data and further sub-folders in there you could set the root to d:\data. The tree performs faster if contains less components.
- **select file type**
  Spectra (1r, 2rr, 3rrr) are automatically contained in the tree. Mol files (.mol, sdf) and/or peak lists (.xml. .str, .peaks) can be selected optionally.

5.4 General display features

The display is generated by a powerful multiple object viewer which can handle internal and external windows, different types of overlays and geometry layouts. Many features are not necessarily needed when doing a method oriented relaxation analysis but the user might want to use them in specific cases, e.g. if he wants to drag & drop another spectrum that is not part of the analysis into the display.

5.4.1 Cross-hair cursor

Whenever the mouse pointer is moved into the spectrum display it turns into a **cross-hair** cursor. When moving this cross-hair, the corresponding ppm positions are displayed in the upper left part of the display. If several internal spectral display windows are visible the cross-hair automatically comes as a linked cursor if possible.

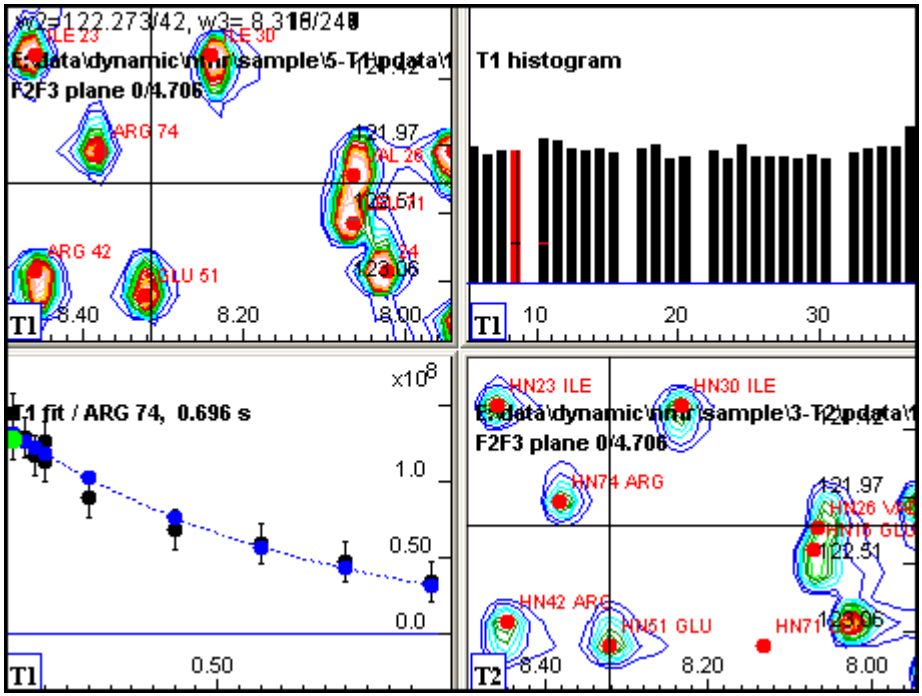

Figure 5.4 Linked cross-hair cursor

In the above example the linked cross-hair cursor is shown in the 2 spectra windows.

5.4.2 Drag & Drop

Spectra may be dragged and dropped from the internal file explorer or desktop explorer. Suitable objects are spectra, mol files and peak lists. The resulting display depends on the target position of the drop and on the type of target objects.

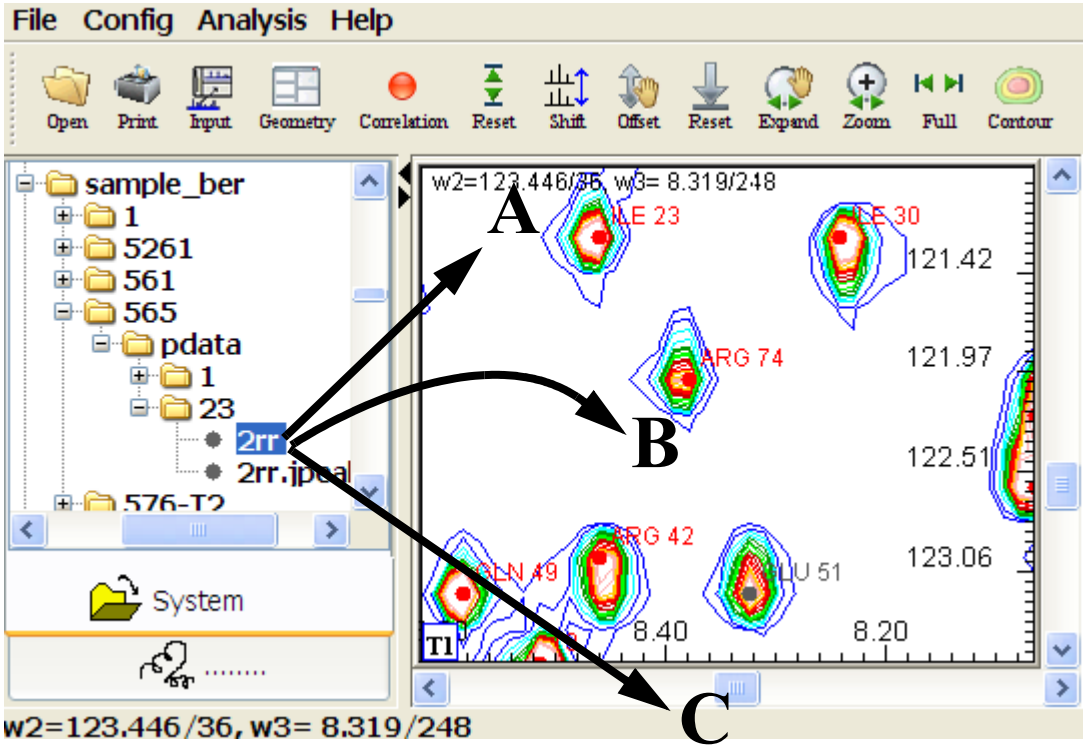

Figure 5.5 Drag & Drop target positions

If the drop target is in **area A** an upper-left overlay display is generated. This is often used to show molecules on a spectrum. If the drop target is **area B** a full overlay display is generated. Spectral regions are adjusted automatically. If the drop target is in **area C** a new display window is generated.

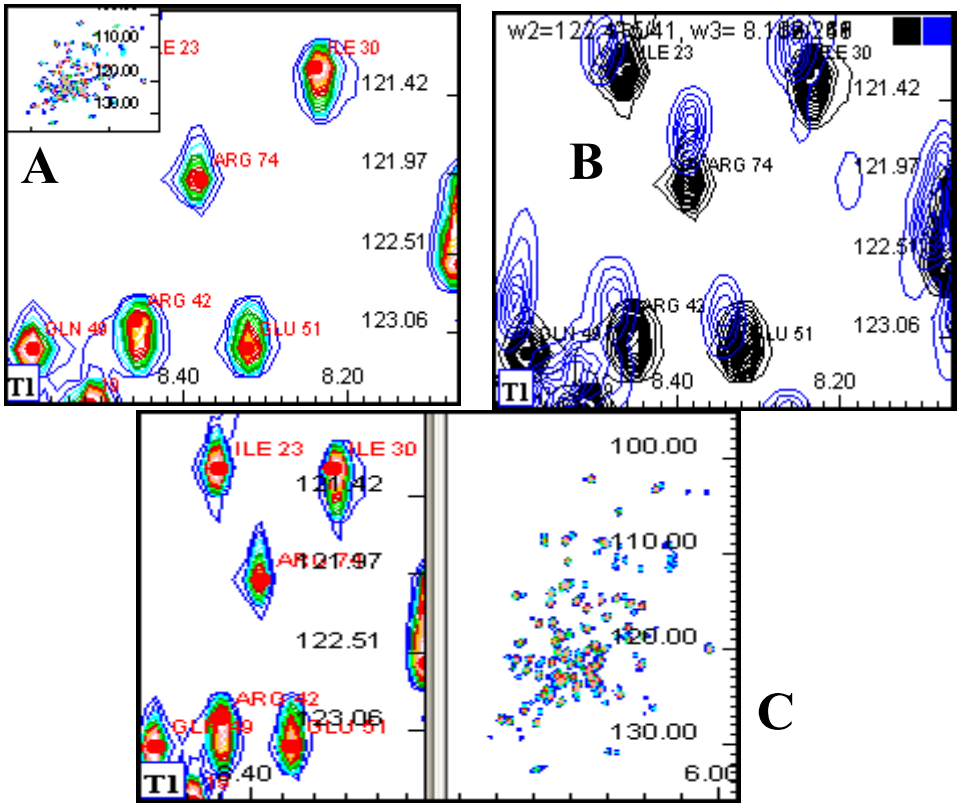

Figure 5.6 Possible results after drag & drop

There are some internal restrictions which can lead to a different behavior. A full overlay of different molecules or spectra of different dimension is for example not possible.

5.4.3 Direct display actions

Many user interactions are related to change the display, The most frequent activi-

ties can directly be done with the mouse:

- **zoom into a region of interest**

  click left button --- drag --- release

- **overview (full spectrum)**

  double click left button (anywhere on spectrum)

- **change contour levels**

  turn mouse wheel

If the display window contains several internal spectral windows one of them is always active.

- **define one internal display window as active**

  single left button click to internal window

5.4.4 Actions using the icon tool bar

The icon tool bar contains a set of icons mostly used to change the display. Several styles of tool bars are available, ***small***, ***large***, ***topspin***, ***topspin-new.*** The user can select among these styles by executing ***Config/Preferences/Change tool bar icon set***. With the availability of direct actions as described in the previous section, the context sensitive popup menus (see below) and the methods available in the method tree, tool bar icons are not really needed for practical work.

Figure 5.7 Example of the icon tool bar

The usage of the icons should be self explaining (also see the tool tip text when the cursor is moved into an icon), only short explanations are given here.

open

A standard file open dialog is launched. By navigating down to a spectrum, e.g. e:\data\guest\nmr\xxx\1\pdata\1\2rr the spectrum gets selected an loaded. More elegantly you can also just use **drag & drop** from the explorer. Open is not used to select spectra used during the relaxation methods since these are selected via the **Data** component on the method tree.

print

A PDF copy of the currently visible objects is generated and stored to the specified file. To get suitable output of relaxation results to use the **Report** component on the method tree.

input options

The Protein Dynamics Center uses a very powerful multiple object viewer to generate displays. It can handle internal and external windows and different types of overlays. ***Input options*** is used for selection. Data loaded via ***Open***, ***Open TopSpin*** follow exactly this selection. Drag & Drop has its own rules as described above. The relaxation methods also have some built-in rules to guarantee proper displays.

geometry

The Protein Dynamics Center uses a very powerful multiple object viewer to generate displays. It can handle different types of layout geometries. ***Geometry*** is used for selection.

correlation

This is understood as zoom correlation. If activated it means that whenever zoom is performed on an object it will automatically be performed on other objects as far as possible.

reset

If the vertical scale has been changed, e.g. with the mouse wheel or via the vertical scale icon (which is invisible by default but can be switched on, see ***Config/Preferences/Select individual icons***), the original vertical scale gets restored.

vertical offset

This changes the vertical offset. After having clicked at the icon, move the cursor to the desired object, click the left mouse button and move the mouse up/down while keeping the mouse button clicked. Release when finished. In case of 2D contour displays constant values are too or subtracted to all contour levels.

vertical reset

The original vertical offset gets restored.

zoom

To zoom into a region of interest, move the mouse to a corner of the region, **click the left mouse button - drag the mouse - release** when finished. Since this is such a frequent operation it can directly be done spectral objects without first clicking at the zoom icon.

full display

The full object display (e.g. full spectrum, complete histogram, etc.) is shown. Since this such a frequent operation you may also just left **double click left button** to an object.

contour levels

A dialog window to set contour levels explicitly is shown. Some special functions like automatic **ramp filling** or **Bezier-smoothing** contours to get nicer displays are also offered. For quick vertical scaling just turn the mouse wheel.

5.4.5 Actions using context sensitive popup menus

Context sensitive popup menus appear after a right mouse button click to an object **Context sensitive** not only means that there are different popup menus depending on the cursor position (e.g. near a peak or not near a peak), it also means that the behavior is context sensitive. If individual 2D spectra are for example loaded via ***drag & drop*** or ***Open TopSpin 2D***, actions refer to the spectra individually. If a peak is for example deleted in one spectrum it would not be deleted in another as well. However, if the spectra were loaded as part of a relaxation analysis method, the software knows that the spectra belong together. Peak manipulations are then updated in all participating spectra.

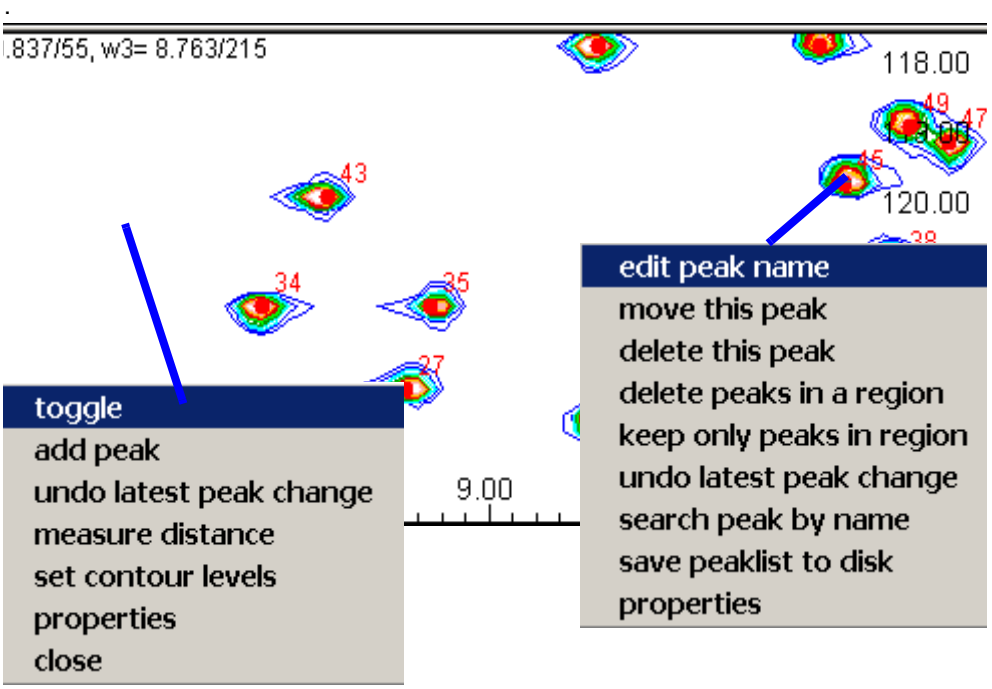

Figure 5.8 Click **right mouse button** at empty area or near peaks

Description of the popup menu items:

If the right click is done in an empty spectral region a popup with the following options appears:

- **toggle**
  The currently selected display object is momentarily toggled to full screen display. That is often needed if several display windows are shown in the viewer and individual object become quite small. All other objects remain in memory and re-appear if *toggle* is activated again.

- **undo last peak change**
  Manipulating peaks is offered with the peak popup, e.g. you may delete all peaks. To be able to undo such manipulations the ***undo last peak change*** command is offered in all popup menus.
- **measure distance**
  Distance measurements behave like zoom: **click left button-drag mouse-release** when finished. The distance between current cursor position and starting point is shown.
- **set contour levels**
  A dialog window to set contour levels explicitly is shown. Some special functions like automatic **ramp filling** or **Bezier-smoothing** contours to get nicer displays are also offered. For quick vertical scaling just turn the mouse wheel.
- **properties**
  Some properties of the current object are shown.

If the right mouse button is clicked near a peak (the peak is highlighted in this case) a popup with options related to peaks appears:

- **edit peak name**
  A name can be assigned to a peak.There is no limitation in length of the name. Naturally, peak names refer to the measured nuclei, e.g. HN in a hetero nuclear ^{1}H ^{15}N HSQC spectrum. If known, the residue number should be included, like HN10 or HN-10 or HN/10 or HN 10 etc. There is no particular naming convention. Only if residue numbers are available other useful displays, e.g. T_1 relaxation times vs. residue number can be generated. If the spectrum is a pseudo 3D or a series of 2D spectra the peak name is also transferred to the other planes/spectra.
- **move this peak**
  A peak can be moved with the mouse After the command is activated near a peak: **click left button - drag - release** when finished. The peak follows the cursor movement. If the spectrum is a pseudo 3D or a series of 2D spectra the peak movement is also done in the other planes/spectra.
- **delete this peak**
  The peak at which the right click was done gets removed from the peak list. If

the spectrum is a pseudo 3D or a series of 2D spectra the peak is also deleted in the other planes/spectra.

- **delete peaks in a region**
  A region can be defined with the mouse (as usual: **click left button-drag-release**). All peaks inside this region get removed from the peak list. If the spectrum is a pseudo 3D or a series of 2D spectra peaks are also deleted in the other planes/spectra.
- **keep only peaks in a region**
  This is the opposite to the ***delete peaks in region*** command. This time all others are deleted.
- **undo last peak change**
  The most recent peak manipulation can be un-done. If the spectrum is a pseudo 3D or a series of 2D spectra this also applies to the other planes/spectra.
- **search peak by name**
  The user specifies a search string which is searched for in all peak names.The first peak that matches causes a new region of interest to be displayed with the peak in the center of it.
- **save peak list to disk**
  All peak manipulations are done in memory. The peak list on disk is not changed until a ***save peak list to disk*** ia applied. Also, when a spectrum is closed, peak lists are saved to disk. Currently, you have to confirm if the peak list on disk shall be really overwritten.
- **properties**
  Some peak properties are shown.

5.5 Tools available from the main menu bar

When working with relaxation methods offered in the method tree, the tools provided in the main menu bar are not needed with the exception of the ***Config*** pull-down menu which contains several entries to change display settings, fonts etc. The main menu bar would however be used if besides the relaxation methods some other spectra are loaded, displayed and analyzed. A minimum set of tools is avail-

able.

5.5.1 File

Open

With Open you can navigate to the location of a spectrum (down to 1r, 2rr, 3rrr) and load it into the Protein Dynamics Center. A spectrum loaded this way can also be analyzed (see *Analysis*) but it does not participate in a relaxation analysis as offered in the method tree.

This is also indicated on the display where little markers show in which method a spectrum is involved. In the following example the spectrum in the upper left, the histogram in the upper right and the fit curve in the lower left all have a T_{1rho} marker and are currently used with the T_{1rho} method. The spectrum in the lower right has no marker and is used for other purposes.

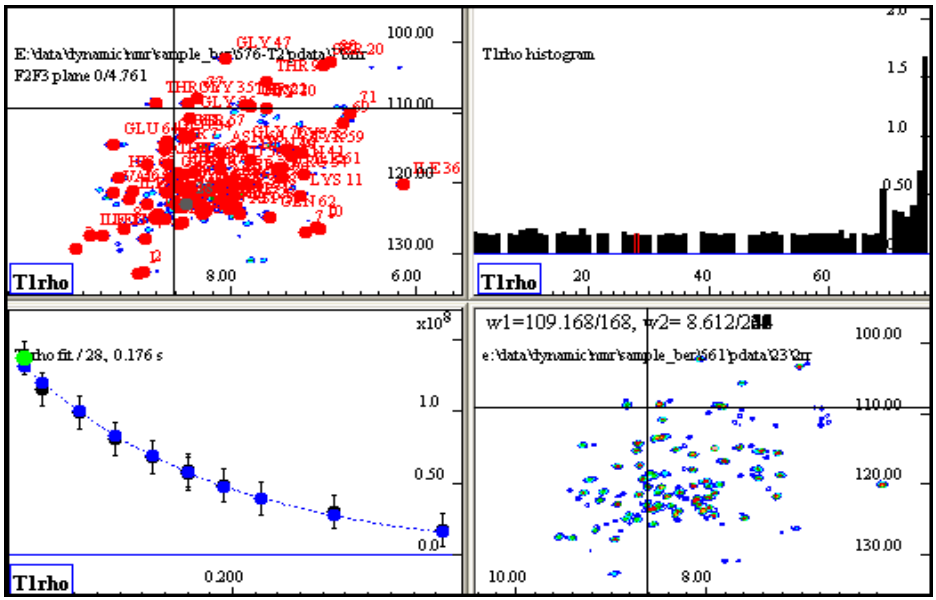

Figure 5.9 Display markers indicate where objects belong to.

Open TopSpin 1D, 2D, 3D
These commands allow the selection of spectra by specifying individual file name components of the path name. A complete path name such as *e:\data\dynamic\nmr\sample_ber\561\pdata\23\2rr* gets split up into its components as shown.

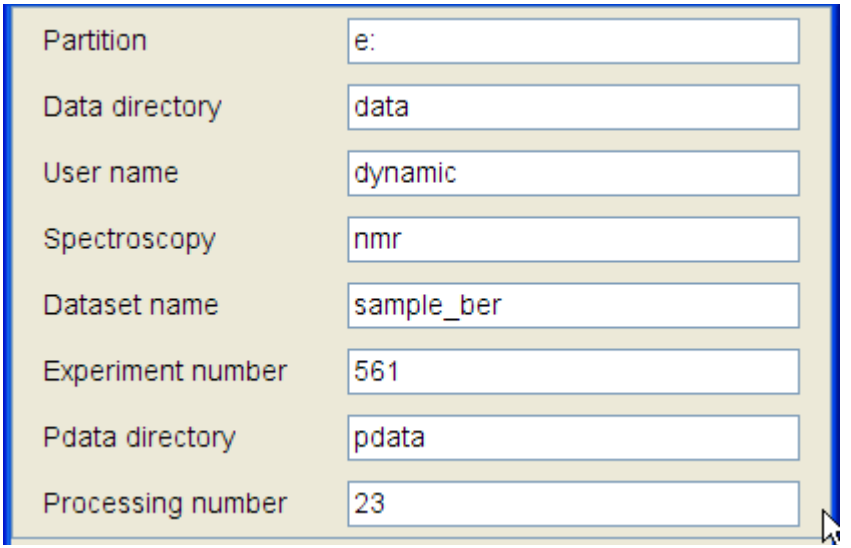

Figure 5.10 Open TopSpin dialog window

The advantage is that individual fields may contain * as **wild cards**. In the following example wild cards are entered such that all spectra with data set names starting with sample and any experiment or processing number are searched for. These are then shown in a list from which the selection can be done with left mouse button clicks (multiple selections with **ctrl** or **shift** keys are possible).

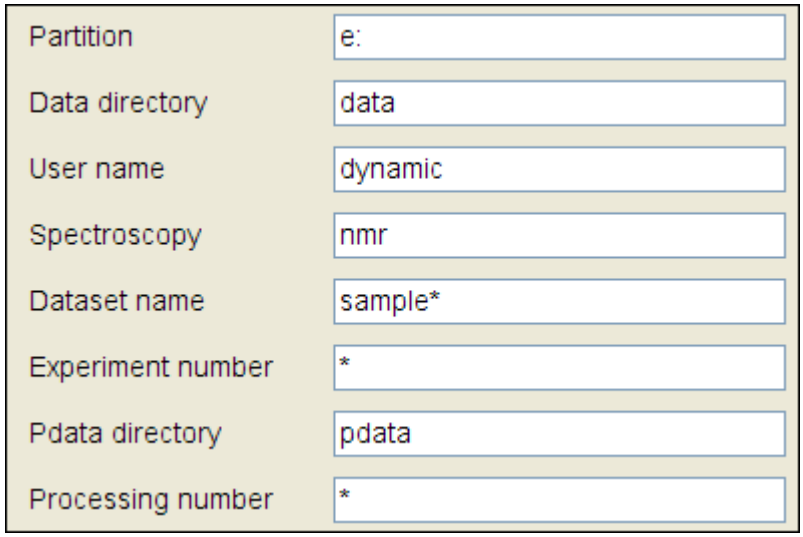

Figure 5.11 Usage of wild cards to search for multiple spectra

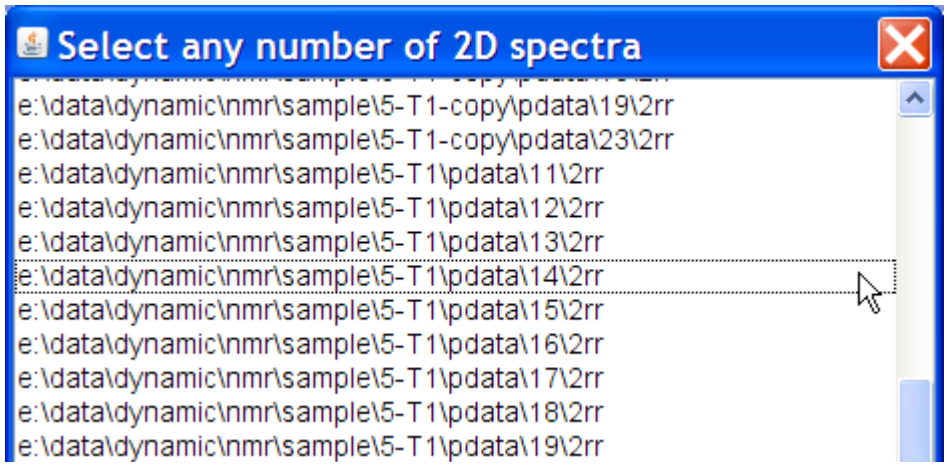

Figure 5.12 Selection of spectra from a file search list

Close all objects
Any object can be closed via the popup menu that appears after a right click to it (unless you are near a peak, then the peak popup appears). When working with the relaxation methods there is also a ***close*** available in the popup menu on the method tree. It closes all objects belonging to a project and should normally be used. ***Close all objects*** is a fast way to close all objects no matter whether they participate in a project or not.

Properties
Properties of the currently loaded objects are shown.

Print
A PDF copy of the currently visible objects is generated and stored to the specified file.

Copy to clipboard
Puts a copy of the currently visible objects into the system clipboard.

Exit
Closes the Protein Dynamics Center. If needed, save projects before using ***Save*** on the method tree.

5.5.2 Config

Input Options
The Protein Dynamics Center uses a very powerful multiple object viewer to generate displays. It can handle internal and external windows and different types of overlays. *Input options* is used for selection. Data loaded via ***Open***, ***Open TopSpin 1D, 2D, 3D*** follow exactly this selection. Drag & Drop has its own rules as described above. The relaxation methods also have some built-in rules to guarantee proper displays.

Geometry
The Protein Dynamics Center uses a very powerful multiple object viewer to generate displays. It can handle different types of layout geometries. ***Geometry*** is used for selection. Data loaded via ***Open***, ***Open TopSpin 1D, 2D, 3D*** follow exactly this selection. Drag & Drop has its own rules as described above. The relaxation meth-

ods also have some built-in rules to guarantee proper displays.

Scaling
If spectra are shown in overlay vertical scaling may get important. Several options are offered. Best results are obtained if different spectra contain a reference peak each to which the data can be scaled.

Display Options
A few simple display options, e.g. if spectra names shall be displayed or not, can be set.

Arrange
If objects are already loaded, e.g. side-by-side, Arrange allows you to change the display without closing and re-loading objects, e.g. pick objects with the cursor and get them overlaid. There are however limitations in so far as only objects of the same type can participate. You can for example not overlay a histogram to a 2D spectrum.

Correlation
This is understood as zoom correlation. If activated it means that whenever zoom is performed on an object it will automatically be performed on other objects as far as possible.

Preferences
A number of options are offered to customize the display. Different sets of icons can be selected out of which individual icons can be put into the tool bar. Many other options refer to font settings. Finally, you can select an alternative PDF viewer if for example no **Adobe Reader** is available on your system.

User Interface
Two possible user interface styles are offered.

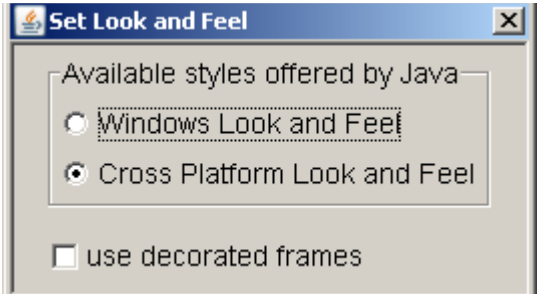

Figure 5.13 Available user interface types

The first type should show the user interface compatible to the current platform, e.g. on Windows XP systems it would look different compared to Linux. The second option creates a user interface that is platform independent and looks the same on all systems. Changing the user interface within a running program can cause update problems therefore you have to confirm that all display objects are first closed. You have to reload them afterwards. Independent on the type of interface you can decide if the internal windows come with decorations or not. Using decorations sometimes looks quite ugly and wastes a lot of space. The only advantage of decorations are that they contain buttons to iconify internal windows. However for real work this seems of minor importance. Furthermore, the ***Toggle*** option in the popup menus of displayed objects allow full screen displays which is often used. Finally remember that the method tree offers ***Suspend*** and ***Resume*** per project to change the number of visible objects in screen.

5.5.3 Analysis

Peak Display Options
The display of peaks, peak names and peak integral values can be switched on/off. In case of 1D spectra integral curves can also be displayed.

Manual Peak Analysis
The tools provided here would be used for a simple analysis of spectra which do not

participate in any of the relaxation methods. Since a general spectrum analysis is not the goal of the Protein Dynamics Center only a few very basic options are available.

- **Pick all peaks above threshold**
  In case of 2D or (pseudo) 3D spectra enter a peak picking threshold. All peaks with absolute intensities above this threshold will be picked. If 0 is entered an automated peak picking threshold calculation is done for each column and each row. Peaks with absolute intensities larger than the intersecting row/column threshold are picked.

- **Pick multiple peaks manually**
  The peak picking cursor gets active and a help window is displayed. Move the cursor to any locations in the spectrum and click the left mouse button to get peaks picked. While doing so the right mouse button is also active and provides a small specialized popup menu. If only a single peak needs to be picked the regular ***Add single peak*** provided on the spectrum popup menu is sufficient. picking peaks with the cursor may not necessarily locate true peaks. Automated peak integration then typically fails. It is therefore better to perform he ***Pick all peaks above threshold*** command first and then clean up the peak list if needed. Corresponding tools are offered in the peak popup menu.

- **Integrate all peaks**
  A line shape analysis from peak height to half height can be performed for each peak individually. Reasonable peak integrals are obtained if peak overlap does not occur within this range. The shape analysis fails if peak maxima are not properly picked. In this case a simple area integration should be performed.

Peak report to text file

The peak list of a spectrum can be written to a text file. A simple format which includes peak positions, intensities, integrals and annotations is chosen.

5.5.4 Help

Manual

The Adobe Reader is launched to display this manual which is stored in pdf format in the ***docu*** folder inside the Protein Dynamics Center installation folder. If no Adobe Reader is found an error message appears. To define an alternative pdf viewer use ***Config/Preferences/Default PDF Viewer*** and specify an executable there.

About

Shows a short summary of names, version numbers, license status and contact address.

License info

Several license information files are shown in a list. The contents gets displayed if selected with a left mouse button click.

6 Basic Relaxation analysis with the Protein Dynamics Center

6.1 Introduction

Each relaxation method shown on the method tree contains the same components **Sample**, **Data**, **Analysis**, **View**, **Report**, **Export**. These components have to be executed in this order. After a successful execution the corresponding nodes get a green color. A red color indicates that a component is currently executing, e.g. doing a longer calculation. Any following components can only be executed if the previous component is shown in green.

In the example below state **A** shows the T_2 method just opened. State **B** indicates that **Sample** and **Data** have been successfully executed while **Analysis** is currently in progress. The next following component (**View**) can only be executed after Analysis has turned into green. It is however possible to execute components of other methods in parallel.

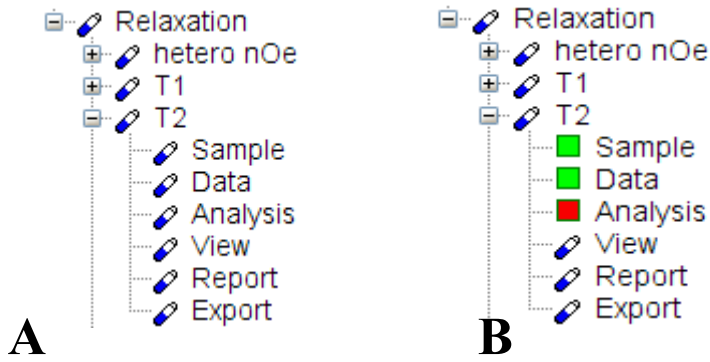

Figure 6.1 Color codes indicate the state of a method

6.2 Sample

The ***Sample*** component is used to provide information about the used sample.

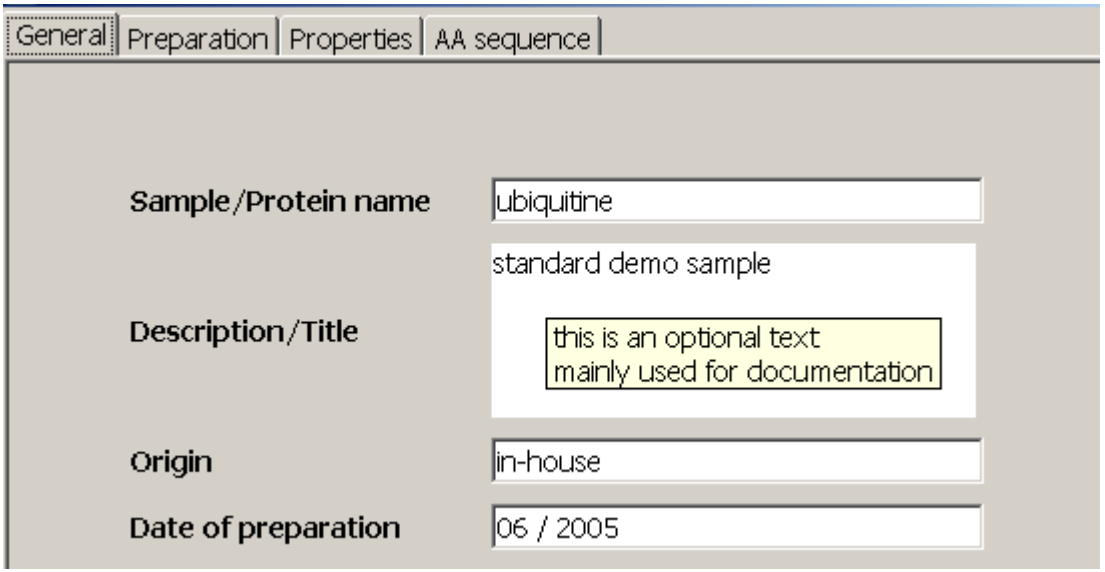

Figure 6.2 Tab oriented dialog window to describe the sample

The information to be provided is mostly used for report purposes. When moving the cursor into a field a tool-tip text provides some additional information. Some of the given sample information can be used during the analysis or for display purposes. In particular this holds for the **AA sequence tab** under which the format (**FASTA**, **SEQ**) and name of an amino acid sequence file can be specified. All relaxation methods offer a histogram display of results (T_1, T_2, etc.) versus sequence. Since the results relate to individual peaks the peak annotations must contain residue information otherwise proper histogram displays are not possible (also see section View below). Proper peak annotation would for example be ALA 10 or ALA/10 or ALA [10].

6.3 Data

If the **Data** component is executed a dialog window is shown to define details related to the spectra.

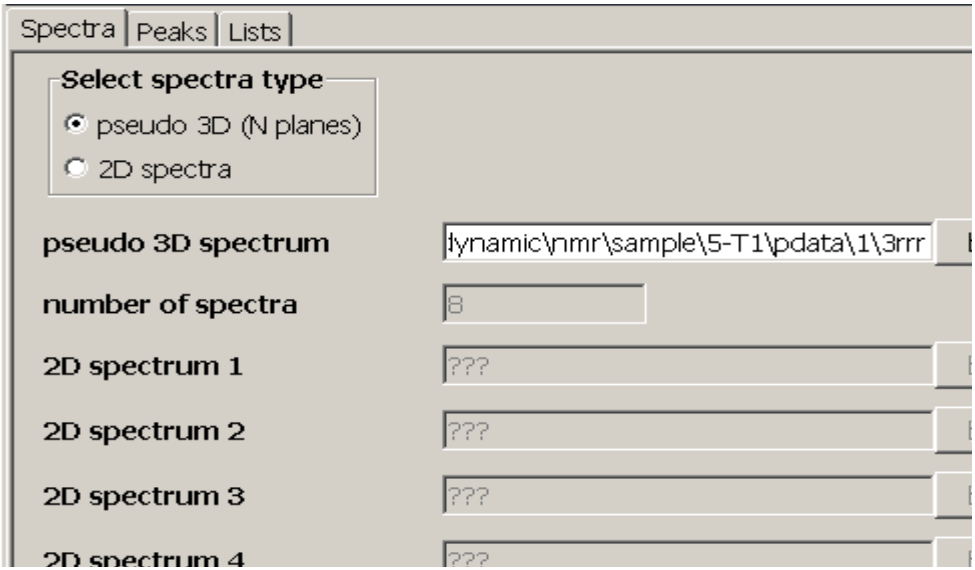

Figure 6.3 Data Dialog for the selection of spectra

Bruker has released pulse programs for all relaxation methods, they all generate pseudo 3D spectra. Mixing times (or field strengths) **are then automatically retrieved** from spectra parameters (further details below). However, if you provide a series of 2D spectra names (currently limited to 16) you must also provide the mixing times (or field strength) in the dialog window. The order of the planes is as stored in the pseudo 3D, the order of the 2D spectra as specified in the **Data** dialog window. In both cases planes or spectra do not need to be sorted, e.g. according to mixing times.

All relaxation analysis methods are based on **peak integrals**. Peak picking and peak integral calculation therefore play a central role. The **Peaks** tab in the **Data** dialog window offers several choices. Peaks can be automatically picked or existing xml peak lists can be used. The automated picking has a built-in **adaptive peak picking threshold** estimation. Noise levels are determined for each data point. Simple descriptive statistics is furthermore used to check if columns or rows in the spectrum contain an unusual number of peaks. If a valid sequence file was specified the overall number of peaks can be well estimated. The most important source of information is however that the spectra are available N times (N planes of a pseudo 3D or N 2D spectra). The signal-to-noise in these spectra differs according to experimental parameters (e.g. mixing time) but a peak should occur in the majority of spectra.

Attention:
To optimize a peak list by hand and use it for relaxation analysis, use the options in the peak popup menu to optimize and save to disk. Then chose the option **use peaklist at spectrum** here.

If peak lists are imported that have not been generated by the Protein Dynamics Center it might be advised to allow so-called **peak snapping**. This means that imported peaks are moved to near experimental local maxima if possible. The snapping algorithm combines a global shift and local nearest neighbor search. If snapping is not activated or imported peaks do not fit to the data a subsequent peak integral calculation gives false results. T

For relaxation parameter determination peak intensities or peak integrals calculated by **region integration** (3 x 3 points), **shape integration** (shapes down to half height) or **peak deconvolution** can be used. In most cases, especially if peaks are isolated, there is no big influence to the fitted parameters. The peak deconvolution method uses already picked peaks. A careful (manually optimized) peak picking is therefore advised. Different shapes (**Gaussian**, **Lorentzian**, mixed) can be applied. Suitable start parameters for the peak line widths should be given. The spectra popup menu contains ***measure distance*** as a graphical tool to estimate it. The used deconvolution algorithm assumes no correlation of parameters of different peaks which speeds up the calculation. Nevertheless, 20-25 minutes are typically

needed for a full pseudo 3D spectrum with 12 planes and about 100 peaks each with **Gaussian** shapes which is recommended as default. The variable Gaussian/ Lorentzian shape needs 2-3 times longer. While not finished you cannot continue with **Analysis** of the current method. You may however work with other methods in parallel. Illegal operations, e.g. closing the spectrum while the deconvolution is calculating are blocked. It is possible to cancel the peak deconvolution but the current plane of a pseudo 3D spectrum or current 2D spectrum of a series of spectra gets finished before the deconvolution is really stopped.

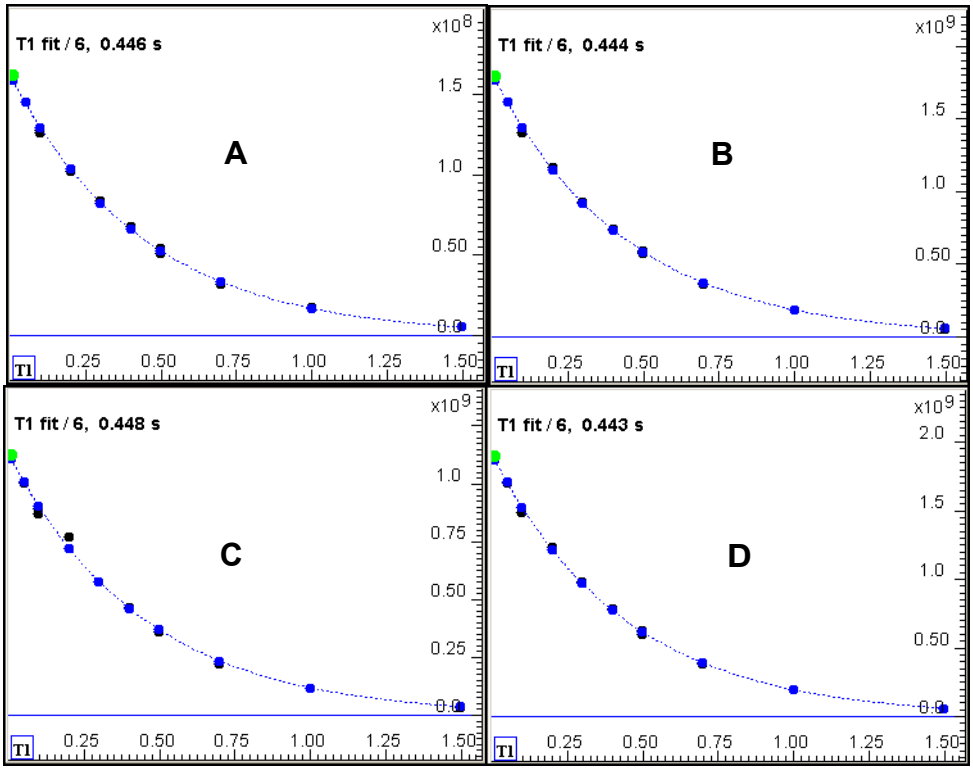

Figure 6.4 Comparison of a T_1 fit based on intensities (A), area integrals (B), shape integrals (C) and deconvolution (D), differences of T1 are less than 1%

Attention
Currently, TopSpin does not allow that peak integrals are stored in peaklist.xml. Peak integrals therefore have to be re-calculated each time a project is opened. This will be changed in near future.

If assignments are available from elsewhere they can be imported. **BMRB** (.str) or **XEASY** (.peaks) formats are supported. BMRB files may have 8 or 9 columns in the assignment section (lines following the **_Atom_shift_assign_ID** to **_Chem_shift_ambiguity_code** lines). XEASY files may contain peak annotations in each following line of a peak, for a description see for example http://nmr.uth-scsa.edu/~ahinck/html/soft_packs/xeasy/xeasy_m3.html#PeakList Only a simplified automated peak snapping is applied to find the closest peaks if assignments are imported.

The **Lists tab** finally relates to handling the mixing times (or field strength) and is strongly context sensitive. In case of T_1, T_2,T_{1rho} and R_{ex} data the mixing times (or field strengths) are either taken from **vdlists** (if Bruker pulse programs have been used) or the user has entered the values by hand under the **Spectra** tab.

In case of T_2 data the mixing times need to be converted from **vclists** into **vdlists** if Bruker pulse programs have been used. This conversion uses the number of mixing times, loop duration and constant duration which are read from the acquisition parameters and depend on the pulse program. When using **hsqct2etf3gpsi3d** for example the loop duration is equal to **D31** but when using **trt2etf3gpsi3d** it is calculated from **P2** and **D20**. If the user defines a series of 2D spectra and enters mixing times in the Spectra tab no conversions need to be done.

In case of R_{ex} spectra peak intensities/integrals need to be converted into **decay rates** first, see e.g. Mulder et al., J. Am.Chem. Soc., Vol 123, 967-975 (2001). This conversion needs the total mixing time (length of **CPMG pulse train**) and is read from the acquisition parameters if Bruker pulse programs have been used, otherwise it must be entered here.

The question is how to treat repetition experiments. It is recommended to measure the one or other mixing time more than once to check for reproducibility of the data. Repeated experiments can be kept or can be collapsed into mean data. Repetition experiments are also evaluated when calculating uncertainties of peak intensities or peak integrals (see next section). Currently, 2 different options are offered to esti-

mate systematic errors. The estimate can be based on variance averaging using peak intensities/integrals. All repetition experiments and all peaks are taken into account and the square root of the average variance is then assigned to all peaks at all mixing times. The other alternative is to use all repetition experiments to calculate the largest difference of peak intensities/integrals per peak and assign this as a systematic error per peak at all mixing times. Variance averaging balances over- and underestimates of systematic errors whereas the difference method slightly overestimates systematic errors. This may have an impact on the model free modeling.

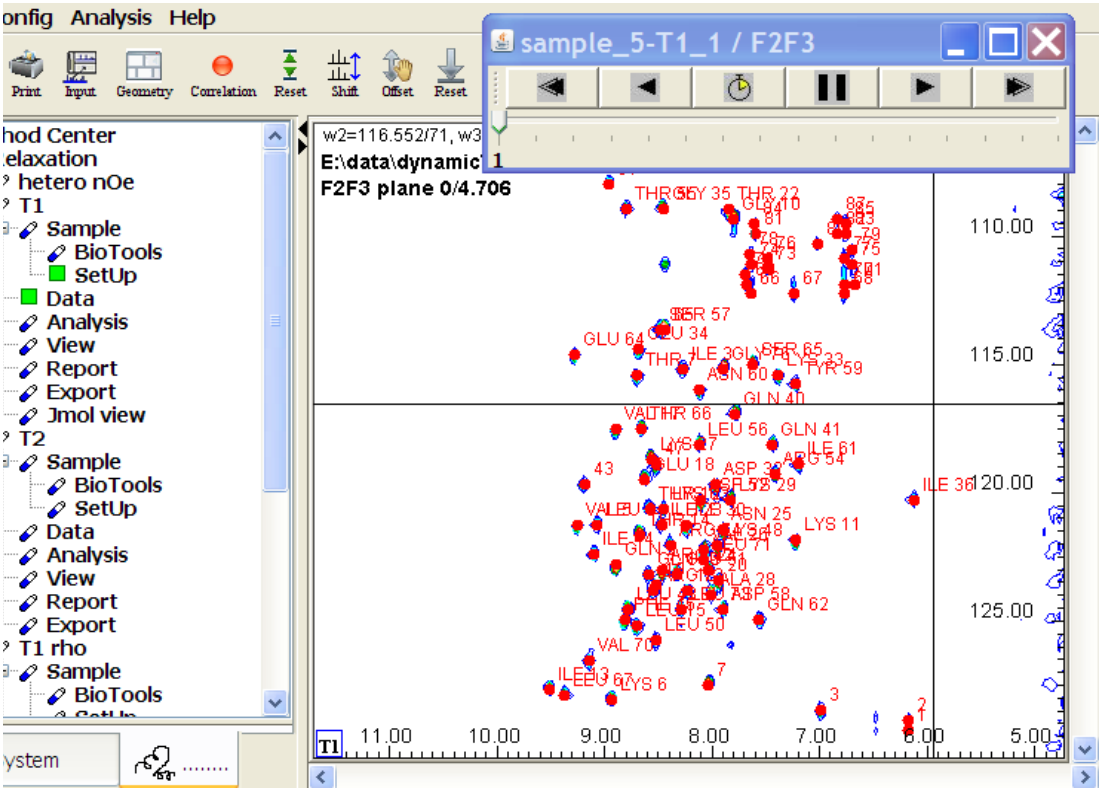

Figure 6.5 Data display with slider

Finally, the specified data are loaded and partially displayed. In case of pseudo 3D spectra the first plane and in case of a series of 2D spectra the first spectrum are shown. A **data-slider** which allows you to select other planes/spectra appears.

The **slider** can be regarded as a player which contains buttons **forward**/**backward play**, **fast forward**, **fast backward** and **stop**. One further button allows a few **set tings**. One can also use the slider arrow to navigate to another plane or spectrum. If several methods are active with several data displays open then several sliders will be shown, one slider for each method. These sliders normally work independently of each other but can also be correlated, see options offered via the settings button.

6.4 Analysis

The ***Analysis*** component of a method can be executed if the ***Data*** component was already successfully executed before. The main task is to fit the peak integrals to particular functions and extract the relevant relaxation parameters.

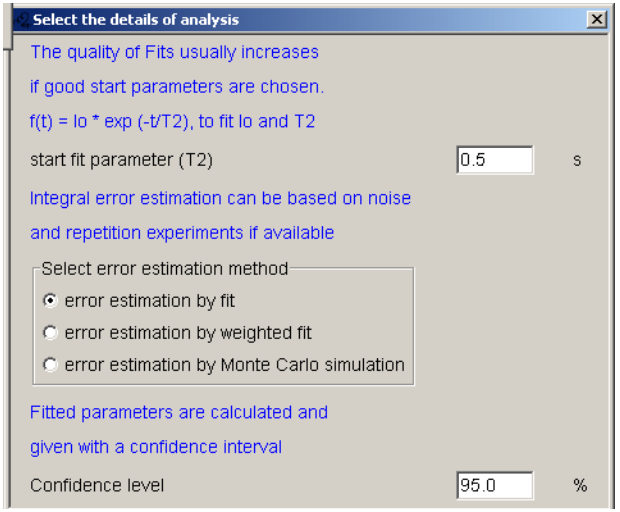

Figure 6.6 Example of the Analysis dialog window

The fit function is indicated in the dialog window. In case of the T_1 method there are three alternative functions in all other cases only a single function is offered at the moment. The parameters to fit are also shown. If possible, specify a reasonable start parameter e.g. for T_1. The non-linear **Marquardt** fit may otherwise get stuck in local minima of the fit parameter space.

Three options related to **error estimation** are offered. The first one is to use the Y data (peak intensities, integrals) under the assumption that their **uncertainties** are unknown but equal for all Y values. The non-linear fit determines errors of the fitted parameters from the inverse of the unweighted curvature matrix (second derivatives of chi-squared). The final **chi-squared** itself is arbitrary.

The second option applies if individual uncertainties of the Y values are known and are passed to the non-linear fit. The errors of the fitted parameters are then calculated from the inverse of the weighted curvature matrix. The uncertainties itself are derived from the standard deviation of the noise in each plane/spectrum and, if available, differences of Y values in repetition experiments. Since the number of repetition experiments is usually small and only available for a few mixing times the maximum difference is determined and assumed to be valid for all mixing times. This way the uncertainties appear to be an upper estimate rather than statistically validated. With the upcoming availability of non-uniform sampling techniques and shorter experiment times more repetition experiments and more elaborate uncertainty calculations will become possible.

The third option is to use a **Monte Carlo** simulation. The non-linear fit is performed 1000 times with the input Y data varied according to a normal distribution with a standard deviation equal to the uncertainties of the Y values.

In order to provide fitted parameters within a **confidence interval** a **confidence level** needs to be given. The confidence interval of a fitted parameter is then calculated by multiplying the error of the fitted parameter with a factor taken from the inverse of **Student-s-T** cumulative distribution at given confidence level and number of degrees of freedom. The confidence interval calculates as fitted parameter +/- fitted parameter error.

6.5 View

Viewing the obtained results is a central component of any method. Important is to correlate spectra/peaks with parameter fits shown as individual fit curves or together with the amino acid sequence. The user may select which types of displays he wants to see, the correlation between all display windows is provided automatically.

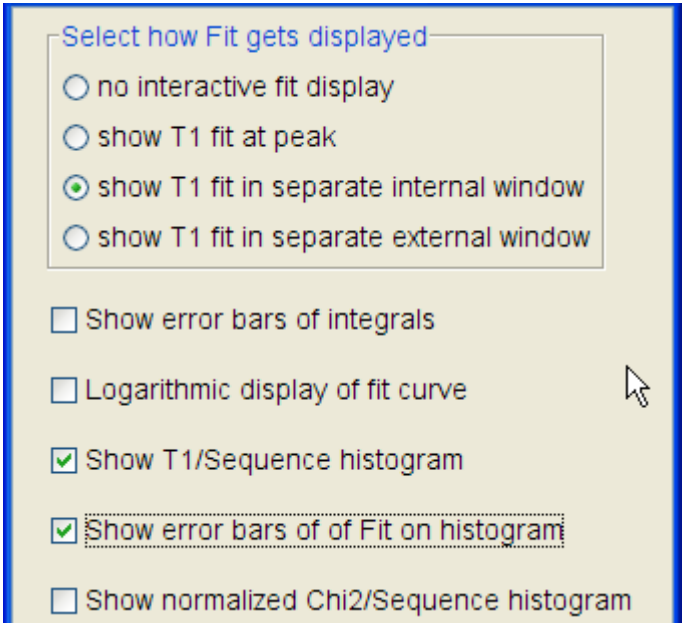

Figure 6.7 View dialog window to customize the result display

Show T_1 fit at peak means that a local display window pops up whenever the cursor is moved close to a peak. If this is too hectic in usage you may get the T_1 fit in a separate internal window or in an additional external window. **Error bars** (uncertainties, see previous section) can be drawn on the fit curves and the exponential decay curves can be shown in logarithmic form. The fit curve display is a regular display object and can be scaled, zoomed etc. as usual.

Fitted relaxation parameters versus amino acid sequence can be displayed in **histogram** style and may include error bars. The histogram is a regular display object and can be scaled, zoomed etc. as usual. Finally, chi-squared of each fit divided by the sum of all Y values taking part in that fit can be displayed as a histogram. It serves as a visual diagnostic tool to estimate the quality of input data and fit.
A typical display with spectra/peaks, individual fit curves and histogram looks as follows:

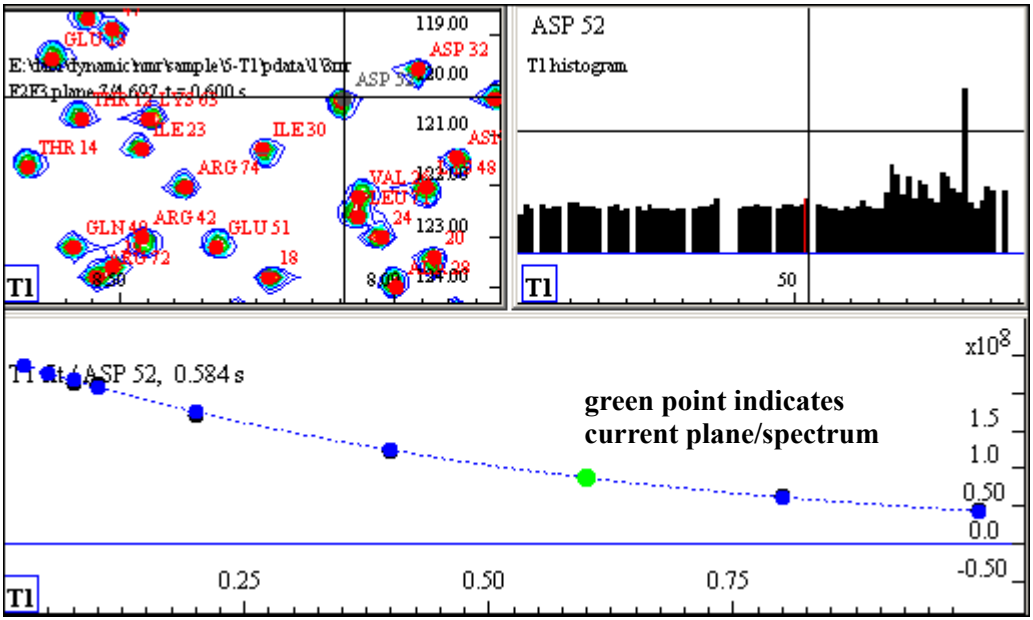

Figure 6.8 Typical result display with spectrum, histogram and fitted curve.

The cursor can be moved in the spectra display or on the histogram. Whenever close to a peak or histogram item the corresponding fit curve is loaded. The cursors itself are correlated. If multiple projects are open and multiple spectra/histogram displays are on screen they are all correlated.

The X variables (e.g. mixing times in the vdlist) may be in arbitrary order and corre-

spond to the order of planes/spectra. The fitting procedures need peak integrals in sorted order which is handled automatically. To see which plane/spectrum relates to which point in the fit curves, a **green marker** is shown. This marker jumps from point to point if you switch to other planes/spectra using the spectra slider.

As you move the cursor the fit displays are updated in separate internal or external windows depending which type you selected. However, you might want to move the cursor into the fit display without getting it changed if the cursor approaches another peak.

While a **key is hit at the keyboard**, the cursor can be moved out without updating other display windows.

Fit display objects

As mentioned the fit display objects are regular display objects and can be scaled, zoomed etc. The fit displays also contain context sensitive popup menus.

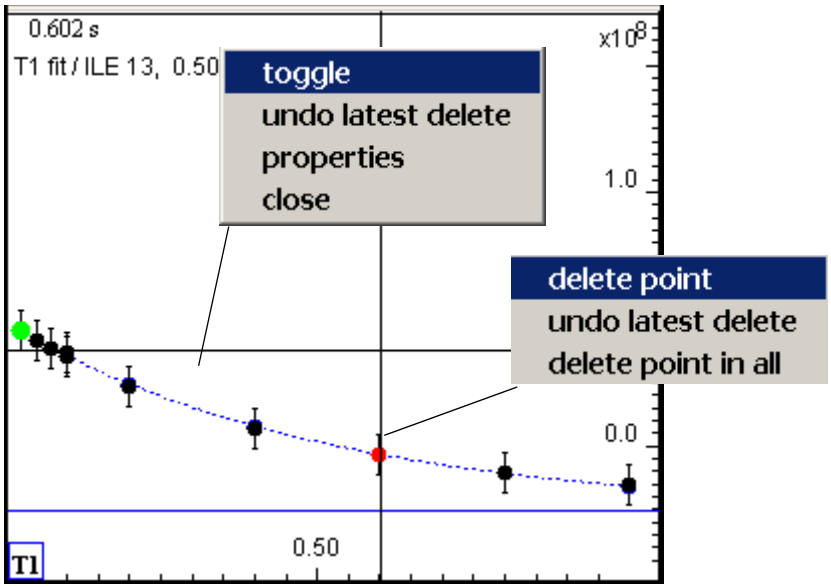

Figure 6.9 Fit display objects also have context sensitive popup menus.

Context sensitive popup menus are available after a right mouse button click. If the cursor was near a point in the fit curve (it gets highlighted with a red color) the popup menu offers:

- **delete point**
  The current point gets deleted and the fit curve is recalculated.
- **undo latest delete**
  The point deleted latest gets restored and the fit curve is recalculated. This can be repeated until all previously deleted points are restored.
- **delete point in all**
  A point in a fit curve would often be deleted if the data of that particular mixing time are bad. But then it makes sense to delete this point in all fit curves. Note, that there is no global undo that would restore a point in all curves.

With a right mouse button click not near any fit point (no red highlight visible) the popup menu offers:

- **toggle**
  This is the standard display toggle to show the fit curve in full screen.
- **undo latest delete**
  The point deleted latest gets restored and the fit curve is recalculated. This can be repeated until all previously deleted points are restored.
- **properties**
  Some properties of the current fit curve are shown. These include the used X and Y values, fit results, g**oodness of fit** and some peak information.
- **close**
  Closes the fit display. It will however come up again as soon as you move the cursor to another peak in the spectrum or another histogram item if visible on screen. To permanently switch off the fit curve display of a method execute the ***View*** component again and chose proper setting there.

Histogram display objects

The histogram displays show for example relaxation parameters versus amino acid sequence. This is possible if the peak to which the parameters belong contain residue numbers as part of their annotation, e.g. ALA 10 or ALA [10] or even just 10. An amino acid sequence file must have been specified under ***Sample*** of a method. If no sequence has been specified or if it cannot be loaded properly a pseudo sequence 1...n is assumed and displayed. The correlation between histogram, spectrum and fit curve display still works. If however peak annotations do not contain any number then they cannot contribute to the histogram. They can also not contribute to the histogram if a sequence is properly loaded (and thus defines how many items the histogram has) but the peak annotation contains a higher number.

6.6 Report

A **pdf report** can that contains loaded information as well as results can be generated in pdf format. Graphical components include the current spectrum, histogram and fit curves display. Numerical components include the sample, peak and fit information The user may select wanted components in a dialog window. The name of the pdf file must also be specified. At the end of the report generation **AcroRead** is automatically launched to display the report. Some versions of AcroRead do not display the pdf file if it was not specified with its absolute path name, e.g. c:\tmp\test.pdf. If you definitively don't have AcroRead on your computer but alternative pdf display programs you may define them as default pdf viewers under ***Config/Preferences/Default PDF Viewer***.

6.7 Export

Information, especially peak information and fit results can be exported to a file on disk. The supported formats are **text** and **xls**. This allows users to present data with standard tools, e.g. chart diagrams in **EXCEL** or to load peak integrals into other software packages to repeat the fit calculations or fit other functions not offered in the Protein Dynamics Center.

6.8 Recommended working procedures

The flexibility of the Protein Dynamics Center might lead to unexpected behaviors. You can for example mix individual spectra displays and analyses with the tools from the main menu bar and at the same time run several relaxation analysis methods and even use the same spectra. Since the software treats the data differently in both cases even results might be different. If spectra are loaded via a method the peak picking for example takes place in all spectra (or all planes) simultaneously and makes the assumption that real peaks should occur in all spectra. This yields better results as compared to an individual analysis.

> We therefore recommend to do the relaxation analysis completely inside a relaxation method and use the global main menu bar tools for other spectra if needed.

Again, remember that the context sensitive popup menus behave differently depending on whether spectra are loaded via a method or via drag & drop or the main menu.

Regarding peak manipulations, one would first check the peak quality before running ***Analysis*** and for example correct peak positions or delete unwanted peaks (see popup menu after right click near a peak). It is however possible to manipulate peaks at any stage, e.g. after inspecting results. In general, recalculations need to be done to update the internal status correspondingly. The re-calculations are done automatically if possible otherwise the status on the method tree is color coded such that you know from where to continue by hand.

> Since the re-calculations may take a longer time (e.g. if peak deconvolution was activated!) we recommend to do peak manipulations as early as possible.

The software has to keep track of an ongoing analysis. A lot of information is stored in memory as long as spectra are loaded. When closing spectra this information is lost even if the spectrum gets reloaded.

We recommend not to confuse a relaxation analysis by closing or loading the participating spectra with tools other than provided by the method.

6.9 Further information

6.9.1 Recommended pulse programs

Bruker has released a number of pulse programs to acquire relaxation data. Examples are (all resulting in pseudo 3D spectra)

15N / Channel f3	HSQC version	TROSY version
NOE	hsqcnoef3gpsi	trnoef3gpsi
T_1	hsqct1etf3gpsi3d	trt1etf3gpsi3d
T_2	hsqct2etf3gpsi3d	trt2etf3gpsi3d
T_{1rho}	hsqctretf3gpsi3d	trtretf3gpsi3d
R_{ex}	hsqcrexetf3gpsi3d	trrexetf3gpsi3d

6.9.2 Summary of fit functions

The following shows the used fit functions, parameters that are fitted and typical start values.

T_2

$$I(t) = I_0 \cdot e^{-t/T_2}$$

I(t) is a decay curve of y-values, i.e. peak intensities/integrals
t is the x-variable, i.e. mixing time
I_0 (amplitude at t=0) and T_2 (decay constant) are fitted
start parameter for I_0 is the y-value at lowest mixing time (automatically chosen)
start parameter for T_2 is given by the user (e.g. 0.2s)

T_1

exponential decay

$$I(t) = I_0 \cdot e^{-t/T_1}$$

saturation recovery

$$I(t) = I_0 \cdot (1 - e^{-t/T_1})$$

inversion recovery

$$I(t) = I_0 \cdot (1 - 2e^{-t/T_1})$$

I(t) is a decay curve of y-values, i.e. peak intensities/integrals
t is the x-variable, i.e. mixing time
I_0 (amplitude at t=0) and T_1 (decay constant) are fitted
start parameter for I_0 is the y-value at lowest mixing time (automatically chosen)
start parameter for T_1 is given by the user (e.g. 0.5s)
Three available fit functions are: exponential decay, saturation recovery and inversion recovery.

T_{1rho}

exponential decay

$$I(t) = I_0 \cdot e^{-t/T_1}$$

I(t) is a decay curve of y-values, i.e. peak intensities/integrals
t is the x-variable, i.e. mixing time
I_0 (amplitude at t=0) and **T_1** (decay constant) are fitted
start parameter for I_0 is the y-value at lowest mixing time (automatically chosen)
start parameter for T_1 is given by the user (e.g. 0.15s)

As part of the T_{1rho} analysis T_2 values are also approximately calculated.

$$T_2 = \sin(\theta)^2 \cdot T_{1rho}$$

with

$$\theta = \operatorname{atan}(B_1/\omega)$$

see e.g. Palmer at al., Chem. Rev. 106, 1700-1719 (2006)

B_1 is the rf field strength, ω the resonance offset.

R_{ex}

$$R(\tau_{cp}) = R_0 + \tau_{ex} \cdot (1 - \operatorname{sinc}(\phi_{ex} \cdot \tau_{cp})/(\phi_{ex} \cdot \tau_{cp}))$$

$R(\tau_{cp})$ is a curve of y-values, i.e. decay rates obtained from peak intensities/integrals
τ_{cp} is the x-variable, i.e. derived from rf field strength
R_0 (amplitude at τ_{cp}=0), ϕ_{ex} (frequency difference of exchanging sites) and τ_{ex}

(exchange rate) are fitted
start parameter for R_0 is the y-value at lowest field strength (automatically chosen)
start parameter for ϕ_{ex} is 2000 Hz (automatically chosen)
start parameter for τ_{ex} is given by the user (e.g. 0.5 1/s)

The decay rates are obtained from peak intensities/integrals, see e.g. Mulder et al., J. Am.Chem. Soc., Vol 123, 967-975 (2001).

$$R(\tau_{cp}) = -(1/T) \cdot \ln(I(\tau_{cp})/I_0)$$

T/2 is the total mixing time.

7 Advanced analysis with the Protein Dynamics Center

7.1 Introduction

The basic relaxation methods include NOE, T_1, T_2, T_{1rho} and R_{ex} experiments. Following the strategy most often described in the literature (see e.g. Fushmann D., BioNMR in Drug research, 283 ff, 2002) a general analysis of the protein dynamics requires the combination of at least the NOE, T_1 and T_2 methods. This combined method is named **NOE/T1/T2** on the method tree.

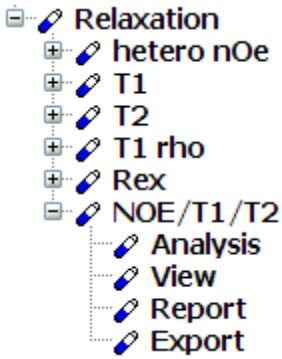

Figure 7.1 The combined NOE/T1/T2 method is part of the method tree

The difference compared to the basic methods is that this combined method uses the results already obtained by the NOE, T_1 and T_2 methods. Therefore, the sample and data information is already known and NOE/T1/T2 does not contain **Sample** and **Data** on the method tree.

7.2 How to execute the NOE/T1/T2 method

The NOE/T1/T2 method uses data and results of the **NOE**, **T_1** and **T_2** methods directly from memory. These three methods must therefore have been set-up or loaded and executed at least up to **Analysis.**

You may then continue with NOE/T1/T2/**Analysis** and select from the upcoming dialog window which calculations shall be done.

At any time you may apply changes to the **NOE**, **T_1** and **T_2** methods, e.g. change peaks. NOE/T1/T2/**Analysis** should then be executed again to synchronize results. This is currently not done automatically but corresponding warning messages are shown to remind you to do so.

The **View**, **Report** and **Export** components work as usual, some more details are given below.

As all other methods the NOE/T1/T2 method shows a popup menu after right clicking at it. It contains the regular entries Open, Save, Save As, Close, Suspend and Resume. If **Save** is executed, it is checked whether the individual, NOE, T1 and T2 methods have originally been loaded from corresponding project files. These are then also saved into the NOE/T1/T2 project file. If later **Open** of NOE/T1/T2 is applied, the NOE, T1 and T2 project files can be loaded automatically.

7.3 Extracting dynamic information from T_1, T_2 and NOE

R_1 (=1/T_1), R_2 (=1/T_2) and NOE are related to combinations of a spectral density functions J at 5 different frequencies (0), J(w_N), J(w_H). J(w_H+w_N), J(w_H-w_N) (see e.g. D. Fushman. BioNMR in Drug Research, p. 288 ff) of X-H bond motions. Since 5 unknown spectral densities cannot be determined from 3 experimental data points, so-called three reduced spectral densities J(0), J(w_N), J(0.87w_H) were introduced (see e.g. Farrow et al., J. Biomol NMR 6, 153-162, (1995)) which can directly be calculated from R_1, R_2 and NOE. With certain assumptions the difference between measured R_2 values and R_2 values predicted via the reduced spectral density functions can be explained as chemical exchange contributions to R_2 (see e.g. Henkels et al., Biochemstry, 46, 15062-15075, (2007)). Unfortunately, the inter-

pretation of the reduced spectral densities in terms of global and local motions is not straightforward even so a semi-manual method called **Lipari-Szabo mapping** exists (see Andrec et al., J. Biomolecular NMR, 18, 83-100, (2000)).

Most widely used in turn is the "**model-free**" formalism introduced by **Lipari** and **Szabo** (Lipari & Szabo, J. Am. Chem. Soc., 104, 4546-4559, (1982) which separates the global tumbling motion (described by a **global correlation time**) from a local motion of a X-H bond, characterized by an order parameter S^2 (ranging from 0 to 1, 1 meaning no local motion) and a **local correlation time**. The model-free formalism was later extended to allow two different local motions on two different time scales (see Clore et al., J. Am. Chem. Soc., 112, 4989-4991, (1990)) per X-H bond. Furthermore the model-free models depend on whether the global tumbling of the protein can be described by one isotropic diffusion constant or rather by an axially symmetric or even anisotropic diffusion tensor. In the later cases the components of the diffusion tensor must first be determined. A possible approach is to evaluate the ratio between $J(w_N)/J(0)$ (see e.g. Walker et al., J. Mag. Res., 168, 336-345, (2004)) which does not depend on local motions, or R_2/R_1 (see e.g. Tjandra et al., J. Am. Chem. Soc., 117, 12562-12566 (1995)). The protein structure is needed in these cases. If the protein structure is not available it is still possible to estimate the ratio of parallel and perpendicular components of the diffusion tensor $D_{II}/D_{\perp}$ (see e.g. see e.g. D. Fushman. BioNMR in Drug Research, p. 296 ff). If this ratio is small, say < 1.2, then assuming global isotropic tumbling is justified.

A substantial number of computer programs doing model-free analysis is available. It seems to be a common practise to offer sets of model free models and fit each model to each residue. The models have different numbers of parameters, usually up to 3 are determined by the fit, others are held constant.The model that fits best to a certain residue is regarded to be relevant for that residue. As a criterion not only χ^2 but for example the **AIC** value, $AIC = \chi^2 + 2 \cdot k$ where k is the number of fitted parameters is evaluated (see, e.g. Akaike, H., Proceedings of the 2nd International Symposium on Information Theory, 267-281, (1973).
It should be mentioned at this point that different major modelling strategies exist. The more simple one assumes that it is possible to find residues that do not show local motions or relaxation exchange by inspecting T_1/T_2 and NOE values. From these the global correlation is estimated and kept fixed. Another approach is to keep the global correlation time not fixed but modify it after each modelling step. A large number of iterations between modelling and global correlation time calculation is done in this case.

7.4 Performing Analysis with the NOE/T1/T2 method

A typical and simple protocol for protein backbone relaxation analysis reads as:

1. Calculate the three reduced spectral densities J(0), J(w_N), J(0.87w_H) . Use them to estimate local correlation times, order parameters and relaxation exchange constants for each residue.

2. Estimate the global isotropic correlation time τ_c and $D_{II}/D_{\perp}$ of the diffusion tensor using a subset of residues that do not show relaxation exchange.

a) Check NOE values to exclude residues that undergo fast local motions, e.g. those with NOE > 0.7

b) Check low T_2 values to exclude residues that show conformational exchange. Since low T_2 values can also be caused by anisotropic tumbling one may rather check combined T_1 and T_2 values (e.g. from http://www.biozentrum.unibas.ch/embo07_nmr/skripts/Tjandra/tjandra_practical_relax.pdf)
$(T_2 - \overline{T_2})/T_2 > f \cdot SD$ with for example f=1.0, SD is the standard deviation of T_2 and simultaneously $(T_2 - \overline{T_2})/T_2 > n \cdot (T_1 - \overline{T_1})/T_1$ with for example n=3.0.

3. Check if the isotropic diffusion model can be assumed or not.

4. **case A:**
Isotropic diffusion can be assumed. Fit individual models of motions to individual residues using T_1, T_2 and NOE values per residue. The estimated overall isotropic correlation time is held fixed. The models are called M1 - M5. No protein structure is required.
case B:
Anisotropic diffusion must be assumed. In this case the protein structure is needed since relaxation depends on the orientation of the NH bond vectors relative to the axes of the diffusion tensor. The diffusion tensor of the protein is first determined by fitting J(w_N)/J(0) or R_2/R_1 of all residues. The single overall correlation time used in M1 - M5 now gets replaced by five individual correlation times which are derived from the components of the diffusion tensor. These correlation times are then held fixed and are incorporated into the model free models which are called T1 - T5. These models also contain five coefficients that depend on the orientation of the NH bond vector relative to the diffusion

tensor. These coefficients have to be determined for each residue individually.

Remarks on the diffusion tensor:

Various alternative procedures are described in the literature. Besides **isotropic** and **anisotropic** diffusion an intermediate **axially symmetric** diffusion is for example considered. There are also different ways to determine the components of the diffusion tensor, e.g. in a **combined modelling** of local motions per residue and global modelling of the diffusion tensor (e.g. see Cole et al., J. Biomolecular NMR, 26, 203-213, (2003)). Another approach is to fit an **individual isotropic diffusion** model to each individual R_2/R_1 pair (e.g. see Lee et. al., J. Biomolecular NMR, 9, 287-298, (1997)).

Remarks on the model fitting:

There are also various ways of interpreting the model fitting. Most common is to regard a model as valid if it best fulfills some statistical criteria. It has been observed, even with simulated data, that various fitting algorithms yield different results. One reason is that a fit of a model function with two or three parameters to only three data points works less well as if for example 12 data points were available. Another reason is that the parameters are constraint and there are different ways to incorporate constraints into the fit algorithm. Fits should always incorporate uncertainties of the data points. The determination of the uncertainties can be done in different ways which influences the fit results. Finally, some models just do not fit. If for example a residue shows conformational exchange, models that do not contain an exchange parameter do not fit. Running the fit algorithm without constraints would then usually yield results that are obliviously false, e.g. an order parameter > 1. But since constraints are typically applied some other combinations of fit parameters are obtained that look reasonable at first sight. The hope is that the statistical evaluation select them as inacceptable results. I

The Protein Dynamics Center uses a combination of **Simplex** and **Levenberg-Marquardt** algorithms. Random selection of a larger number of start parameters, e.g. 1000, is possible and suggested. Constraints are defined but not used during the minimization. However, results that violate constraints are rejected. Thus, it happens that a model does not yield any reasonable result at all. This is a clear indication that the model is not applicable.

After having selected **Analysis** a dialog window to setup the calculations is shown. Different categories are accessible under different tabs.

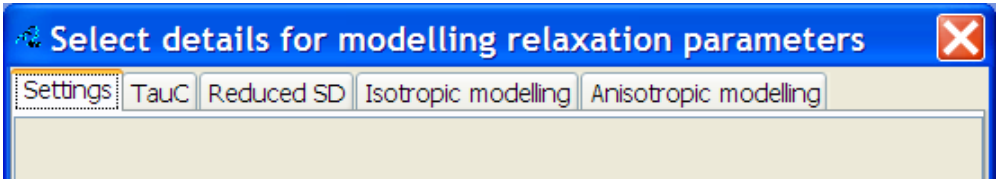

Figure 7.2 Tab oriented dialog window to select details for modelling

Settings

Some global settings can be set under this tab.

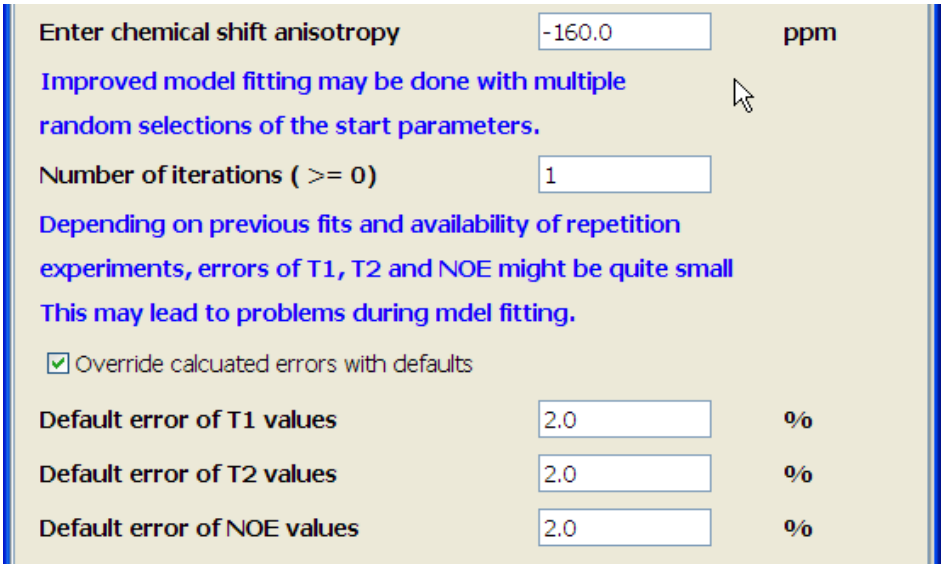

Figure 7.3 The **Settings** Tab is used to define some global parameters

The **XH bond length** (usually NH) is 1.02 Angstrom by default. The N chemical shift anisotropy, **CSA**, is often taken as an average of -160 ppm. In reality it varies from residue to residue and can cover values from -120 to 215 ppm.
The number of iterations refers to the random selection of start parameters when doing the actual modeling (see below). A value of 0 means that no extra selections are done, we recommend value of 500 or 1000.

Override calculated errors with defaults needs some explanations. The source of all calculations are the T_1, T_2 and NOE spectra. The errors of the extracted peak intensities or integrals depend on signal-to-noise and repetition experiments. Repetition experiments are usually available for some of the mixing times of the T_1 and T_2 experiments, but often not for the NOE experiment. The errors of T_1 and T_2 relaxation parameters are obtained via T_1 and T_2 fits which take into account the intensity/integral errors. The NOE value instead is only derived from a ratio of two peak intensities/integral and the error of the NOE value results from an error propagation calculation. A consequence of this procedure is that the NOE errors are often much smaller than the T_1 or T_2 errors. Also, the absolute errors may become quite small.

When running the modelling it is often seen that the NOE values are less well reproduced than T_1 or T_2 values. With small NOE errors they then dominate the final χ^2 values of the fits. In fact the final χ^2 values may go up dramatically like up to 10^6. Other models that slightly better reproduce the NOE values (but not necessarily the T_1 and T_2 values) come out with much smaller final χ^2 and would always dominate in a model comparing procedure.

An alternative provided here is to override the calculated errors with default errors which can be specified as percentage numbers separately for T_1, T_2 and NOE values. The errors are then just obtained by taking each T_1, T_2 and NOE value an by multiplying it with the percentage number, i.e. 0.01 if 2% was selected.

TauC, calculation of the global isotropic correlation time τ_c

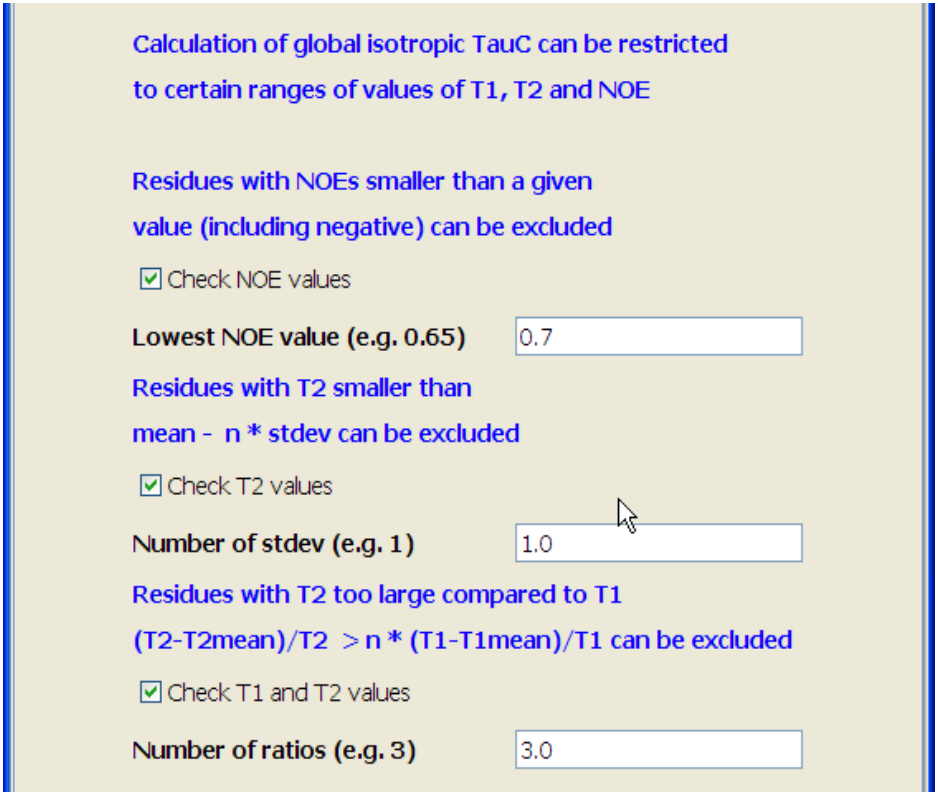

Figure 7.4 The **TauC** Tab is used to select proper residues for the global τ_c calculation.

Suitable residues can be selected according to 2b). τ_c is then obtained by fitting the selected T_1/T_2 ratios to

$$\frac{2}{5} \cdot \frac{\tau_c}{1 + (\omega \cdot \tau_c)^2} \qquad (7.4.1)$$

(see e.g. Kay et al., Biochemistry, Vol. 28, No.23, 1089, page 8974). The finally presented τ_c value is the average over all fitted values. Alternatively, τ_c is estimated as an average over local τ_c calculated for each residue as

$$\tau_c = \frac{1}{2 \cdot \omega_N} \cdot \sqrt{\frac{6 \cdot T_1}{T_2} - 7} \qquad (7.4.2)$$

(see e.g. Fushman et. al., J. BioMol NMR, 4, 61-78, (1994)). Again the selected residues are chosen. Other suggestions described in the literature, e.g. taking all residues but use R_1 and R_2 values corrected for high frequency components and/or exclude some local τ_c values from the final averaging are not used here.

Usually, both calculations yield very similar results.

Reduced SD, calculate reduced spectral densities, estimate $D_{II}/D_{\perp}$, R_{ex} and S^2.

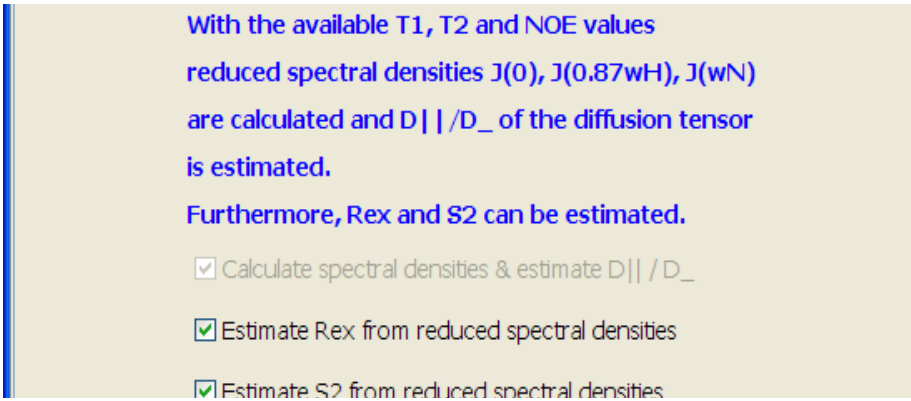

Figure 7.5 The **Reduced SD** Tab is related to spectral density calculations

The reduced spectral densities are always calculated since also used internally. Based on selected T_1/T_2 pairs they are also used to estimate the rotational anisotropy. A quantity ρ defined as

$$\rho = \left(\frac{2 \cdot R'_2}{R'_1} - 1\right)^{-1} \qquad (7.4.3)$$

in which R'_1 and R'_2 are taken from R_1 and R_2 corrected for high frequency motions

(see e.g. D. Fushman. BioNMR in Drug Research, p. 294) is calculated for all selected residues. The maximum and minimum found values for ρ correspond to NH vectors pointing parallel or perpendicular to the rotational diffusion axis. D_{II} and $D_{\perp}$ of the diffusion tensor can be calculated as

$$D_{\perp} = \frac{\omega_N}{3} \cdot \sqrt{\frac{\rho_{min}}{3 - (4 \cdot \rho_{min})}} \qquad (7.4.4)$$

and

$$D_{II}/D_{\perp} = 1 + \frac{9}{4} \cdot \frac{\rho_{max}/\rho_{min} - 1}{3 - 4 \cdot \rho_{min}} \qquad (7.4.5)$$

If $D_{II}/D_{\perp}$ is small, say < 1.2, then the overall motions can be considered as isotropic. Isotropic modelling does not need any information about the molecular structure.

The user can furthermore select if R_{ex} and S^2 shall be directly estimated from the reduced spectral densities for each residue.

Isotropic modelling, fitting NOE, R_1 and R_2 of each residue to one or more isotropic models

A number of model-free models have been proposed in the literature. Various software packages offer various numbers of these models for fitting. Following the majority we offer 5 most commonly used models for the spectral density functions.

M1 $$J(\omega) = \frac{2}{5} \cdot \tau_c \cdot \left[\frac{S^2}{(1 + (\tau_c \cdot \omega)^2)} \right]$$

.

S^2 fitted, global isotropic correlation time τ_c held fixed

M2 $$J(\omega) = \frac{2}{5} \cdot \tau_c \cdot \left[\frac{S^2}{(1 + (\tau_c \cdot \omega)^2)} + \frac{(1 - S^2) \cdot (\tau_c + \tau_e) \cdot \tau_e}{(\tau_c + \tau_e)^2 + (\omega \cdot \tau_c \cdot \tau_e)^2} \right]$$

S^2, τ_e fitted, global isotropic correlation time τ_c held fixed, formula given here is the numerically stabilized form (e.g. d'Auvergene, Protein Dynamics, p. 91)

M3, like M1 but an R_{ex} term is added to the R_2 calculation.

M4, like M2 but an R_{ex} term is added to the R_2 calculation.

M5

$$J(\omega) = \frac{2}{5} \cdot \tau_c \cdot \left[\frac{S^2}{(1 + (\tau_c \cdot \omega)^2)} + \frac{(1 - S_f^2) \cdot (\tau_c + \tau_f) \cdot \tau_f}{(\tau_c + \tau_f)^2 + (\omega \cdot \tau_c \cdot \tau_f)^2} \cdot \frac{(S^2_f - S^2) \cdot (\tau_c + \tau_s) \cdot \tau_s}{(\tau_c + \tau_s)^2 + (\omega \cdot \tau_c \cdot \tau_s)^2} \right]$$

As an approximation the fast local motion is assumed to be very fast, i.e., $\tau_f \to 0$, the global isotropic correlation time τ_c held fixed, fitted are S^2, S_f^2 and τ_s, numerically stabilized from.

From linear combinations of J(0), J(w_N), J(w_H), J(w_N+w_H) and J(w_H-w_N) the NOE, R_1 and R_2 can be calculated (see e.g. Abragam, 1961). The model fitting tries to adjust the fit parameters such that the calculated NOE, R_1 and R_2 values fit best to the corresponding experimental NOE, R_1 and R_2 values.

Anisotropic modelling, fitting NOE, R_1 and R_2 of each residue to one or more anisotropic models

If $D_{II}/D_{\perp}$ is not small, say >1.2, the assumption of isotropic diffusion may not be valid and anisotropic diffusion must be assumed. In this case the relaxation rates of individual NH bond vectors depend on their orientation relative to the axes of the diffusion tensor.

The anisotropic modelling requires a pdb file containing protons. The location of the pdb file must be defined via ***Sample*** in the **NOE**, **T1** or **T2** (or in all of them) methods on the method tree. Check the ***Structure*** tab in the dialog window that comes up. If you only have a pdb file without protons available use one of the existing software packages like **Reduce** developed by J. Michael Word in the lab of David and Jane Richardson at Duke university. For a description see Word, et. al. J. Mol. Biol. 285, 1733-1747, 1999.

The determination of the components of the diffusion tensor and the orientation of the tensor relative to a molecular frame can be done in different ways. One, also used here, is to take the spectral density function as (e.g. Woessner, D. E., J. Chem. Phys., 37, 647-654, (1962))

$$J(\omega) = \frac{2}{5} \cdot \sum_{j=1}^{5} A_j \cdot \left[\frac{\tau_j}{(1 + (\tau_j \cdot \omega)^2)} \right] \qquad (7.4.6)$$

The newly introduced five correlation times τ_j are linear combinations of components (D_{xx}, D_{yy}, D_{zz}) of the diffusion tensor. The coefficients Aj also depend on D_{xx}, D_{yy} and D_{zz} and additionally on the coordinates of the NH unit vectors in the axis frame of the diffusion tensor. Since the coordinates of the NH vectors (taken from the pdb file) are typically given in any other frame, they need to be rotated into the axis frame of the diffusion tensor. This rotation involves three Euler angles α, β, γ. Thus, we have a total of six unknowns (D_{xx}, D_{yy}, D_{zz}, α, β, γ).

From linear combinations of this spectral density function one can calculate R_1 and R_2 or R_2/R_1 for each residue and compare it to experimental R_2/R_1 values. A minimization of

$$\chi^2 = \sum_{\text{residues}} \left[\frac{R_2/R_1 - R'_2/R'_1}{\sigma} \right]^2$$

where the summation is over selected residues, R_2/R_1 are experimental values, R'_2/R'_1 are calculated and σ is the error determined for each R_2/R_1.Selected residues means that residues probably showing conformational exchange are excluded.

If the diffusion tensor is known, the above isotropic models can be rewritten by replacing the single global isotropic correlation time τ_c by a summation of individual anisotropic correlation times.

For the general anisotropic case we then get

TM1

$$J(\omega) = \frac{2}{5} \cdot \sum_{j=1}^{5} A_j \cdot \left[\frac{S^2 \cdot \tau_j}{(1 + (\tau_j \cdot \omega)^2)} \right]$$

.

S^2 is fitted, τ_j are calculated from the components of the diffusion tensor and are the same for each residue, Aj are calculated from components and orientation of the diffusion tensor and have to be calculated individually for each residue.

Instead of M2 we get

TM2

$$J(\omega) = \frac{2}{5} \cdot \sum_{i=1}^{5} A_j \cdot \left[\frac{\tau_j \cdot S^2}{(1 + (\tau_j \cdot \omega)^2)} + \frac{(1 - S^2) \cdot (\tau_j + \tau_e) \cdot \tau_e}{(\tau_j + \tau_e)^2 + (\omega \cdot \tau_j \cdot \tau_e)^2} \right]$$

S^2, τ_e are fitted.

TM3, like TM1 but an Rex term is added to the R_2 calculation.

TM4, like TM2 but an Rex term is added to the R_2 calculation.

TM5

$$J(\omega) = \frac{2}{5} \cdot \sum_{i=1}^{5} A_j \cdot \left[\frac{\tau_j \cdot S^2}{(1 + (\tau_j \cdot \omega)^2)} + \frac{(1 - S_f^2) \cdot (\tau_j + \tau_f) \cdot \tau_f}{(\tau_j + \tau_f)^2 + (\omega \cdot \tau_j \cdot \tau_f)^2} \cdot \frac{(S^2_f - S^2) \cdot (\tau_j + \tau_s) \cdot \tau_s}{(\tau_j + \tau_s)^2 + (\omega \cdot \tau_j \cdot \tau_s)^2} \right]$$

As an approximation the fast local motion is assumed to be very fast, i.e., $\tau_f \to 0$, fitted are S^2, S_f^2 and τ_s.

Progress bars indicate the ongoing fit calculations. The GUI remains active during this time and you change the display or work on another method. Changing any data (e.g. delete a peak) that are currently used in the calculation must not be done.

7.5 View results obtained with the NOE/T1/T2 method

Once the ***Analysis*** method has been successfully executed ***View*** can be used to define details of the result display and get corresponding objects on screen.

A tab oriented dialog window contains tabs related to general (T_1/T_2, spectral density), R_{ex}, S^2 and molecular displays. Individual check boxes are available depending on which calculations have been performed.

In the example below the R_{ex} tab is active and shows that R_{ex} histograms can be shown as estimated from spectral densities or via M3 and M5 modelling. These two model contain an R_{ex} term. Since anistropic modelling was not performed R_{ex} as obtained via modelling of TM3 and TM4 cannot be selected.

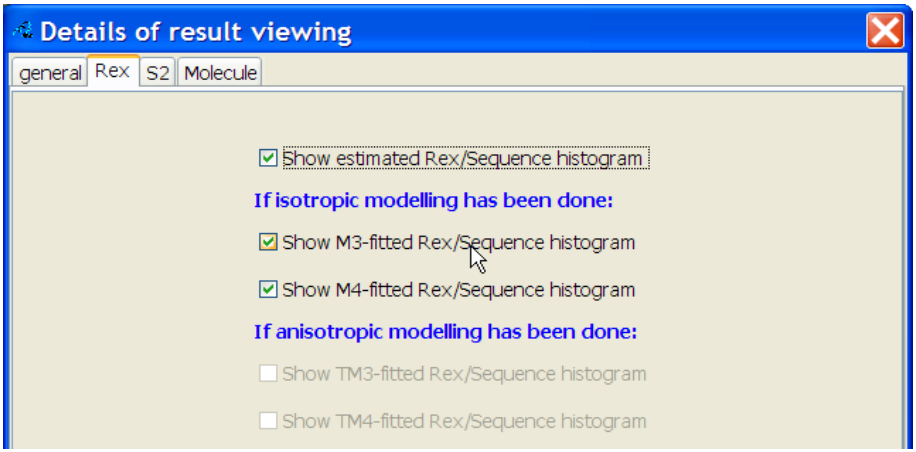

Figure 7.6 View dialog (Rex tab opened) the define details of the result display

The resulting display may contain numerous display objects as shown in the following example. As usual all objects are correlated as far as possible. If for example the cursor is moved to a certain residue in above histogram it is also moved to this residue in all other histograms.

Note, that some of the items in the histograms are shown in green instead of black color. This is due to a corresponding selection of a check box in the ***View*** dialog. All residues that were selected for the global isotropic τ_c calculation are shown in green.

The second example below shows the usefulness of the combined displays. A residue shows a low T_2 value (**a**). The reason could be a R_{ex} contribution. In the NOE histogram its value is relatively high (**b**) which means there is no fast local motion at this residue. T_1 is not low, thus T_1/T_2 high (**c**). Thus, the low T_2 value is probably not due to anisotropy. From a modelling of M1 a somewhat elevated R_{ex} value can be seen (**d**). All 4 result displays seem to be consistent with each other.

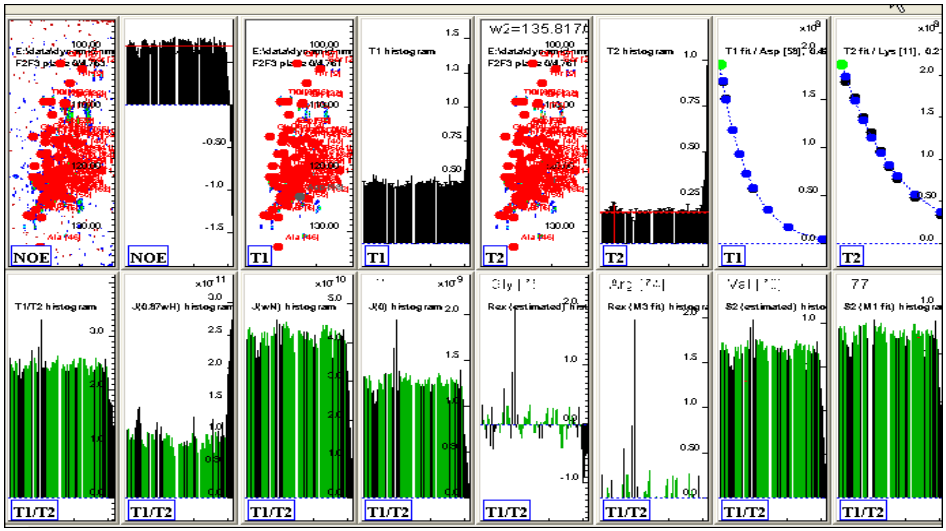

Figure 7.7 Example of a result display showing various types of objects

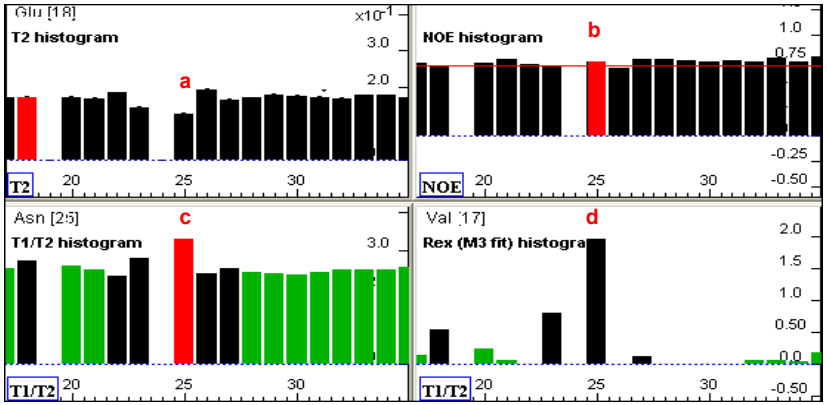

Figure 7.8 Example of a result display illustrating the combined interpretation

Apart from viewing the dynamic parameters via amino acid sequence histograms it is possible to display them directly on the 3-dimensional moldecular structure. Again this requires a pdb file of the protein which included protons. The location of the pdb file must be defined via ***Sample*** in the **NOE**, **T1** or **T2** (or in all of them) methods on the method tree. The 3-dimensional structure display is performed via **Jmol** that is part of the Protein Dynamics Center installation. A reference for Jmol is Jmol: an open-source Java viewer for chemical structures in 3D. http://www.jmol.org/. A very simple viewer that internally calls the Jmol viewer is provided.

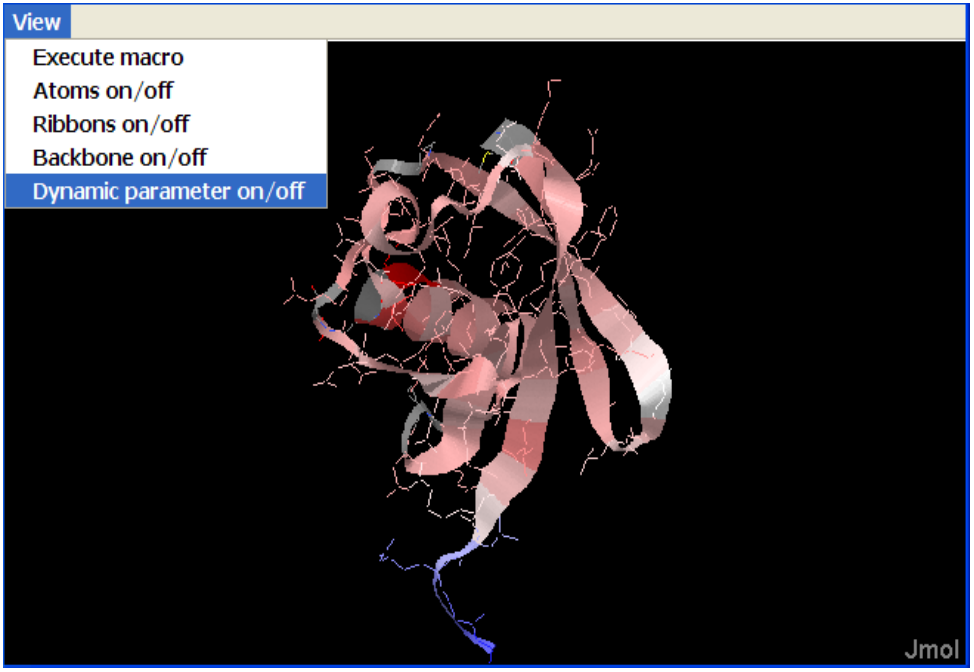

Figure 7.9 A simple strcuture viewer based on Jmol.

The viewer is pre-loaded with the given pdb file. A simple pull-down menu allows you to configure the display. One entry is called **Execute macro.** Any **RasMol**/ **Chime** script can be entered here, see for example http://jmol.sourceforge.net/

docs/JmolUserGuide/ch04.html. A second entry is **Dynamic parameters on/off**. This command presents a list of available dynamic parameters (T1/T2, spectral densities, order parameters etc.) that can be displayed as ***Positional Variability*** like for example B-factors. For a reference see for example http://jmol.sourceforge.net/jscolors/#Position%20along%20chain. The color translation used here would correspond to the so-called ***relativeTemperature*** scheme, i.e. the range between minimum and maximum value of a dynamic parameter, say J(0) is mapped into 30 colors ranging from blue (low values) to red (high values). The dynamic parameters are transferred in memory from the Protein Dynamics Center to the structure viewer and not placed into the pdb file first.

7.6 Report results obtained with the NOE/T1/T2 method

Since so many results are possibly generated with the calculations and modelling the reports contain numerous pages. Since one of the critical issues is which of the models shall be finally accepted, back calculated R_1, R_2 and NOE values are compared to their experimental counterparts.

T1/T2 Analysis

BRUKER

name	S2(M2)	error	TauE(M2)	error	dR1[%]	dR2[%]	dNOE[%]
Tyr [59]	0.834	0.000462	7.24e-11	3.55e-13	1.56	0.0848	16.2
Asn [60]	0.862	0.000281	7.01e-11	2.90e-13	1.10	0.0354	12.3
Ile [61]	0.845	0.000552	8.47e-11	5.61e-13	5.07	0.270	16.9
Gln [62]	0.713	0.000986	8.12e-11	5.23e-13	3.12	0.633	42.5
Lys [63]	0.869	0.000387	4.91e-11	2.84e-13	4.84	0.179	8.31
Glu [64]	-	-	-	-	-	-	-
Ser [65]	0.839	0.000626	9.41e-11	6.69e-13	1.21	0.169	19.5
Thr [66]	-	-	-	-	-	-	-
Leu [67]	0.865	0.000736	8.19e-11	7.43e-13	0.0497	0.00135	13.8
His [68]	0.836	0.000344	8.48e-11	4.20e-13	1.97	0.164	18.3

Figure 7.10 Partial view of a pdf report page

The above example shows a part of a page containing the M2 modelling results. Columns for $S^{2,}$ error of S^2, TauE and error of TauE. In addition there are 3 columns for **dR_1**, **dR_2** and **dNOE** all given in %. The percentage is calculated from $|R_1 - R_1^c|/R_1$, $|R_2 - R_2^c|/R_2$ and $|NOE - NOE^c|/NOE$,

There are further pages that compare the AIC values of all modelling results. Usually the model with the smallest AIC is regarded as the best one. But this is a dangerous assumption since the AIC values also strongly depend on the errors of the experimental R_1, R_2 and NOE values. Especially, if the modelling does not reproduce the NOE value quite well the overall AIC value of that model gets large, even though the more relevant R_1, R_2 and values might have been well modelled.

An often found strategy is to accept models on the basis of how well they reproduce R_1 and R_2 values neglecting the NOE values.

Therefore the user may specify an expected percentage number such that if dR_1, dR_2 (as mentioned above) are below this number the corresponding model gets marked with an asterisk. Independent of the AIC value this marker tells that the particular modelling was able to reproduce R_1, R_2 well.

AIC/M1	AIC/M2	AIC/M3	AIC/M4	AIC/M5
4.67e+04	202	4.66e+04 *	6.00 *	205
8.67e+04	-	8.65e+04 *	6.00 *	8.67e+04
8.17e+04*	-	-	6.00	8.17e+04 *
7.27e+04	150	-	-	7.27e+04
2.13e+05	-	-	-	2.13e+05
1.89e+05	4.04 *	-	6.00 *	1.89e+05

The above example in general shows very high AIC values for M1, M3 and M5 modelling. The internal analysis showed that the main reason were not well reproduced NOE values and small experimental errors in the determination of the NOE values.

The results therefore need some interpretation. For example in line 1 the AIC of model M2 has a low value (202), but it is not marked with the asterisk, whereas M3 with an AIC of $4.66e^4$ is marked. In terms of reproducibility of R_1 and R_2 M3 would be the better model. Model M4 has an AIC of 6.0 which is the lowest AIC for this residue. Since the number or parameters of M4 is 3 this AIC value ($AIC = \chi^2 + 2 \cdot k$)

shows that χ^2 of the fit is almost zero meaning that the fit exactly reproduced R_1, R_2 and NOE. As described in text books on statistics such perfect fits are rather unexpected and the fit is regarded as over fitting the model. But when checking the individually fitted parameters of Model M4 all of them appear to be reasonable.

7.7 Export results obtained with the NOE/T1/T2 method

Export provides the same information as ***Report*** in textual or xls format.

7.8 Validity of the NOE/T1/T2 modelling

The modelling of only 3 data points to models that contain up to 3 parameters causes principal and technical problems. In some cases the errors of the fitted parameters indicate that the fit is questionable, i.e. has not found the true global minimum. Resulting high AIC values may or may not indicate problems as described above.

Additionally, as can be seen in the report and export result tables there are cases where the modelling failed definitively. Instead of fitted model parameters or obtained AIC values only - is shown in the table cells.

There are 3 major reasons for a failure.

1. Minimization does just not find a minimum
This is rather a rare case and can be overcome if one allows a larger number of extra minimizations with randomly selected start parameters. 500 - 1000 extra minimizations are recommended.

2. Minimization fails due to numerical problems
During minimization matrix inversions are involved which may fail. Furthermore, internal quantities like the minimum increments of parameters might be exceeded.

3. Minimization works but results don't make sense

The minimization algorithms run in an unconstraint way but the obtained results are crosschecked with pre-defined constraints. If violations occur the minimization is regarded as having failed. Using 500 -1000 extra minimizations often lead to more results that are accepted. The current pre-definitions are as follows

parameter	start value	min value	max value
S^2	0.5	0.0	1.0
Sf^2	0.5	0.0	1.0
τ_e	$1x10^{-11}$	$1x10^{-15}$	$1x10^{-10}$
τ_s	$1x10^{-11}$	$1x10^{-15}$	$1x10^{-10}$
Rex	0	0	$1x10^{6}$

In addition, models M5 and TM5 are only accepted if $Sf^2 >= S^2$. In general models are only accepted if $\tau_e < \tau_c$ if any internal correlation time is involved.

If no extra iterations of start parameters are chosen repeated modelling of models M1 to M5 should always yield the same results, even on different computers. Extra iterations however are based on random start parameter selections and the random numbers are obtained from the system's **random generator**. As a consequence repeated modelling may yield different results. If the number of extra iterations is high, say 1000, the differences of the modelling results should be small. Vice versa, if the results are different, the modelling has obviously not yet found the global minimum and is not reliable.

The situation is somewhat different regarding models TM1 to TM5. These models

need the fitting of the diffusion tensor which also uses the specified number of random start parameter selections if this number is larger than 500. Otherwise a minimum of 500 iterations is used. The diffusion tensor fitting may thus yield different results and the modelling of TM1 to TM5 may then also be different.

8 Interfacing to relax

As mentioned in the previous chapter the strategy of modelling dynamic parameters from relaxation parameter is done such that residues with no fast internal mobility and no relaxation exchange are selected via their NOE and T_1/T_2 values. From these residues a global isotropic correlation time or the diffusion tensor are derived. Both are kept fixed when modelling M1--M2 and TM1--TM5. This strategy is simple and still used my many software packages in this field.

The software **relax** (see e.g. d'Auvergne, E. J. and Gooley, P. R., J. Biomol. NMR, 40(2), 107-119, 2008) uses a more advanced modelling scheme. As agreed with the authors, relax will be able to read the Protein Dynamics Center output of the NOE, T_1, T_2 and R_{ex} methods. Just execute these methods as described in chapter 6 and use ***Export*** to generate individual output files in **textual** form.

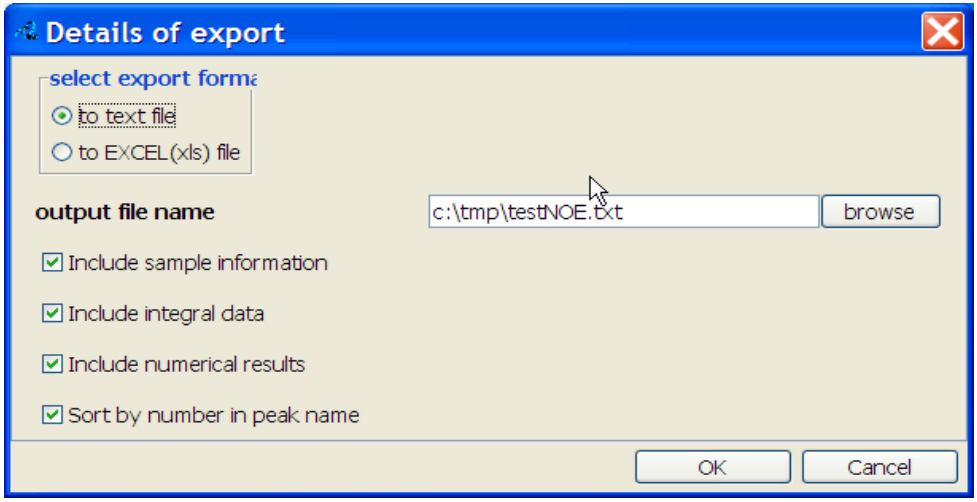

Figure 8.1 Standard ***Export*** dialog window of any relaxation method.

The output file names are arbitrary, for example NOE.txt, T1.txt, T2.txt, REX.txt. It is

recommended to activate all tic boxes in the export dialog to get complete outputs. If data at multiple fields are available analysis and export have to be repeated at each field strength using different output file names.

Relax which uses the T_1, T_2 and NOE data is used in combination with **NESSY** an open source (GPL) software. NESSY additionally analyses relaxation dispersion experiments at one or multiple magnetic fields. Data is individually fitted to different 2 states, 3 states and n-states models and model selection is performed according to Aikaike information criteria (AIC), AIC using second order correction for small sample size (AICc) or F-test.

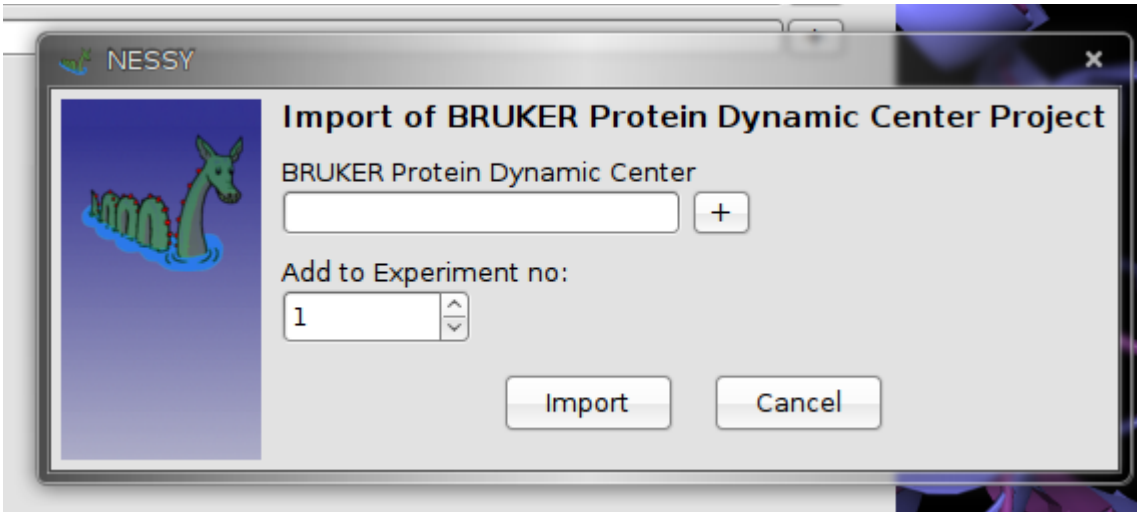

Figure 8.2 NESSY dialog window to import output from the Protein Dynamics Center.

Importing the Protein Dynamics Center output into relax is currently implemented and will be available with version >= 1.3.7 of relax.

9 Contact

Manufacturer:

Address:	Bruker BioSpin GmbH Service & Support Department Silberstreifen D-76287 Rheinstetten Germany
Phone:	+49 (721) 5161 6456
Fax:	+49 (721) 5161 91 6456
E-mail:	nmr-software-support@bruker.de
WWW:	http://www.bruker-biospin.com
author:	peter.neidig@bruker-biospin.de

Appendix A

A.1 Figures

A.3 Index