Preparing BioSAXS experiments with ISPyB

Contents

[1.1 Introduction 2](#_Toc324150763)

[1.1.1 Outline of users needs 2](#_Toc324150764)

[1.1.2 Aim of this document 2](#_Toc324150765)

[1.1.3 Example of a user project 2](#_Toc324150766)

[1.2 List of terms 3](#_Toc324150767)

[1.2.1 Samples 3](#_Toc324150768)

[1.2.2 Data classification hierarchy 3](#_Toc324150769)

[1.2.3 Data acquisition 3](#_Toc324150770)

[1.2.4 Data processing 4](#_Toc324150771)

[1.2.5 Shipping 4](#_Toc324150772)

[1.3 Definitions 4](#_Toc324150773)

[1.3.1 Assembly 4](#_Toc324150774)

[1.3.2 Macromolecule 5](#_Toc324150775)

[1.3.3 Sample 5](#_Toc324150776)

[1.3.4 Buffer 6](#_Toc324150777)

[1.3.5 Additive 6](#_Toc324150778)

[1.3.6 structure 6](#_Toc324150779)

[1.3.7 Session 6](#_Toc324150780)

[1.3.8 Stock solutions 7](#_Toc324150781)

[1.3.9 Sample Plate (MX Container) 7](#_Toc324150782)

[1.3.10 Sample Case (MX Dewar) 7](#_Toc324150783)

[1.3.11 Shipment 7](#_Toc324150784)

[1.4 Overview of heirarchy in data model 7](#_Toc324150785)

[1.4.1 Biological description of hierarchy from assembly to sample and buffer 7](#_Toc324150786)

[1.4.2 Data collection 8](#_Toc324150787)

[1.5 Screenshots 8](#_Toc324150788)

[1.5.1 Session List 8](#_Toc324150789)

[1.5.2 Experiment List 9](#_Toc324150790)

[1.5.3 Create / Edit Buffer 9](#_Toc324150791)

[1.5.4 Create / Edit Macromolecule 10](#_Toc324150792)

[1.5.5 Create / Edit Samples 10](#_Toc324150793)

[1.5.6 Create / Edit Assembly 11](#_Toc324150794)

[1.6 Wizards for streamlining user interaction with ISPyB 11](#_Toc324150795)

[1.6.1 Static 12](#_Toc324150796)

[1.6.2 Kinetic 14](#_Toc324150797)

[1.6.3 HPLC 15](#_Toc324150798)

[1.6.4 Edit Robot XML / Data Collection Array Sequence 16](#_Toc324150799)

## Introduction

### Outline of users needs

The aim of a BioSAXS experiment is to determine the low resolution shape of a macromolecule in solution under physiological conditions. Even in the simplest case, many similar samples (at different concentrations) and identical buffers need to be measured to complete a thorough dataset. Although the users know the macromolecules they wish to measure they may not always be aware of the number samples needed to undertake a thorough experiment. Not only in the design stage but often during data collection feedback from data collected is required to ensure maximum data quality is obtained. Thus not only does the database need to contain all information needed to collect data, but it is also needed to store results (raw data and processed results) to allow improved feedback. Also through its design the database user interface must help the users to provide all necessary information, ideally in a simple and intuitive way without the need for inputting the same information over and over again.

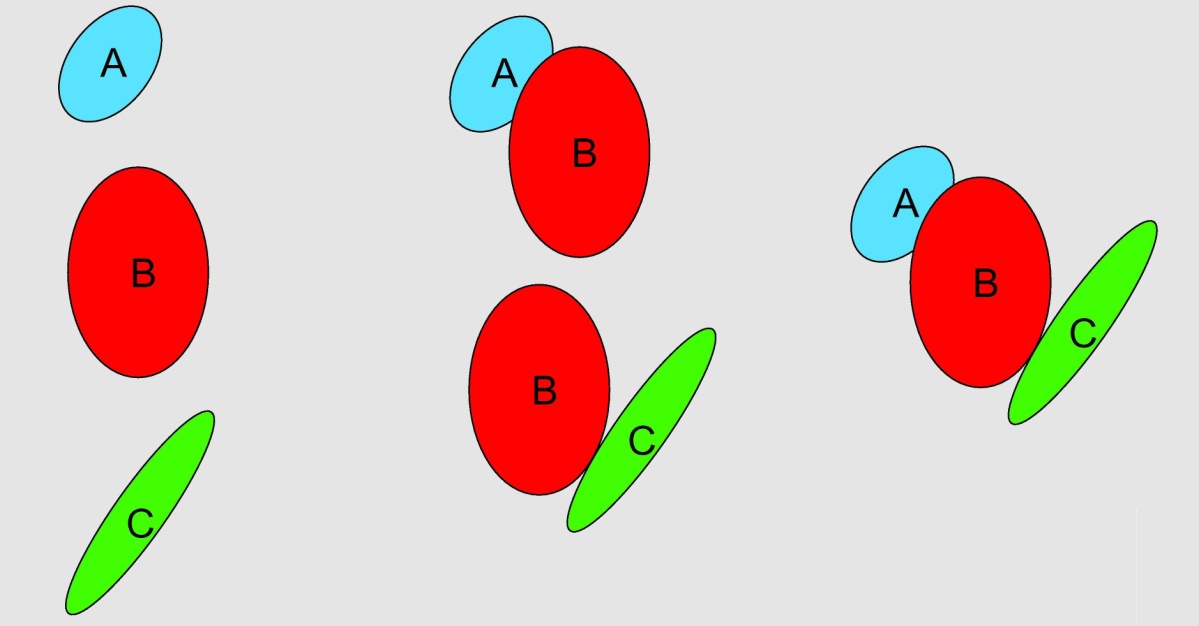
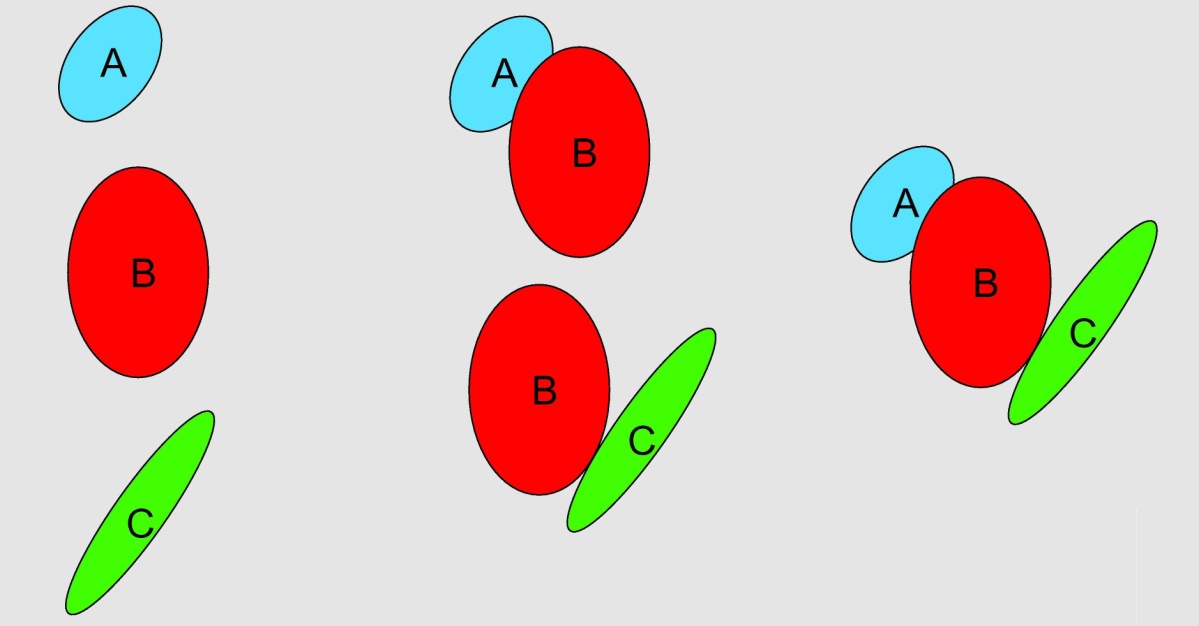
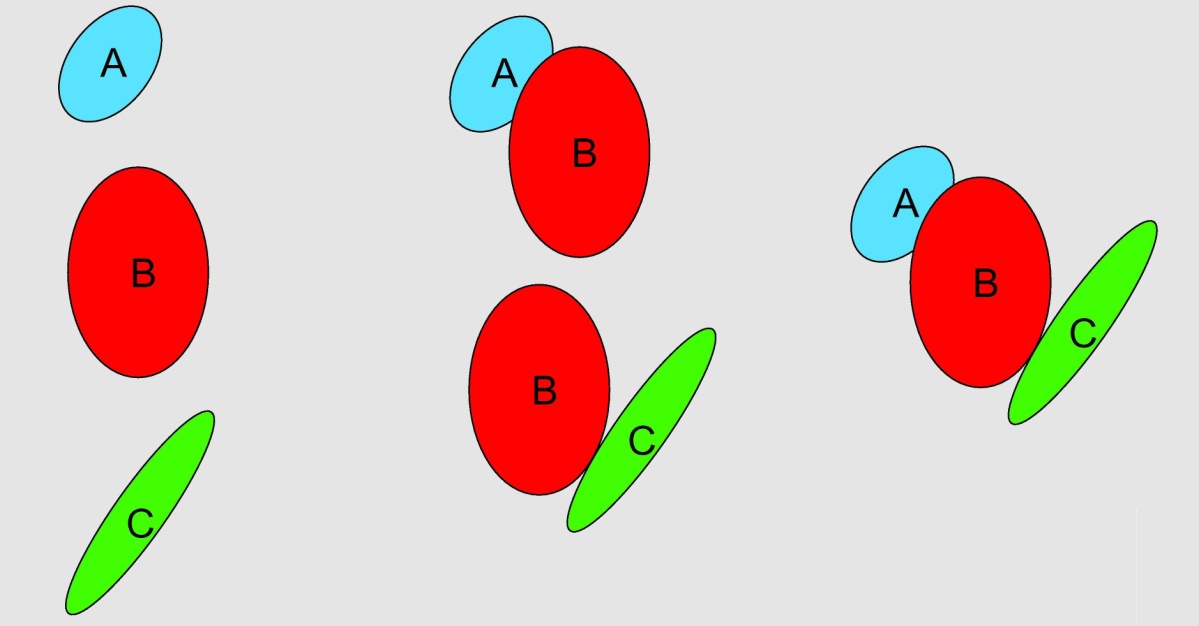
### Aim of this document

This document defines the groups which have been created to minimize the duplication of variables, their hierarchy and relationships. It is the basis for the data model proposed and as such explains is greater detail the tables and is a reference for the terms and names used.

### Example of a user project

An enzyme which in its functional form is comprised of 3 individual subunits (A, B and C). The first part of the Experiment ( P1) is to determine how the subunits fit together.

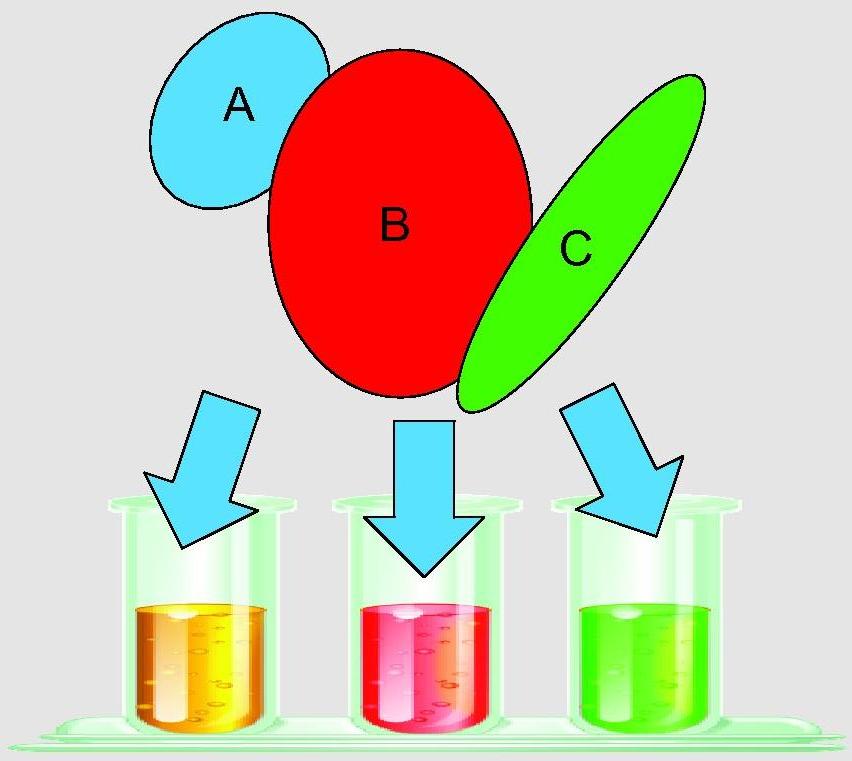
P1 will comprises data collections for:



A, B and C individually The dimeric complexes The trimeric complex

P2 is to understand how the enzyme functions and will comprises data collections for:

The trimeric complex ABC (same conditions as for P1) Plus a data colelction for each buffer condition



Ligands/additives required for activity and/or non hydrolysable homologues to isolate the various stages of the reaction.

## List of terms

### Samples

**Macromolecule** – Biological construct in solution for investigation

**Sample** –specific measurment details for the macromolecule (concentration, HPLC or position in SC) for all alliquots needed for the data collection group. *Thus a Macromolecule should have at least 3 samples.*

**Buffer** – The matched solution in which a sample is suspended

**Additive** – Any component of the buffer which will be varied in an experiment (salts, Ligands, detergents lipids, deuteration)

**Assembly** – Description (size, composition) of a complex containing multiple (list of contained) macromolecules (different combinations of individual macromolecules are considered as a separate macromlecules themselves (see example))

**Experiment** – List of related Assemblies, macromolecules and their results which are used in different sessions.

### Data classification hierarchy

**Frame** – One individual exposure of the detector (refers to both both the 2D detector image .edf and the reduced 1D curve .dat) *(should be linked with the logged Beamline parameters)*

**Measurment / Run** – all frames for an individual acquisition (buffer or individual macromolecule at one concentration) *(should be linked with the average of the logged Beamline parameters)*

**Data collection** – combination of runs (buffer-before, macromolecule at one concentration and buffer-after).

**Data collection group** – combined data collections for one macromolecule (minimum 3 concentrations). *In a sequential data collection group, by deffinition all samples are in the same buffer there is only one buffer measurment between samples which is classed as both the buffer-after in the first data collection and the buffer-before in the subsequent data collection .*

**Data collection array**– The data collection groups that are required to answer a biological question (describe an assembly).

**Session** – a slot of beamtime allocated to a particular proposal (may contain multiple data collection arrays)

Depending on the complexity of the system for some cases (structure validation for example) a data collection array will only contain one data collection group. However, for others there could be many groups in the array (see example)

### Data acquisition

**Number of Frames** – The desired number of frames to be acquired for each measurment / run

**Time per frame** – The duration of each individual **frame** in a measurment / run

**Transmission** – The % X-ray transmission to be used for eachmeasurment / run

**Energy** – the X-ray energy used for each measurment / run

Acquisition parameters should normally be kept the same for all measurements in a data collection group and all data collection groups in an data collection array. However, as some samples may have different radiation sensitivity it is possible that some macromolecules have to be treated differently. Furthermore as the deffinition of radiation sensitivity is impirical starting from the first sample measurment in the data collection group, based on analysis it could be seen that the acquisition parameters (including sample volume) need to be modified to improve data quality / confidence. Thus it is possible (though not ideal) that the data acquisition parameters will be modified for the subsequent measurment / run. However, it should be strongly encouraged that the acquisition parameters should be the same for a data collection group! As exposure time/intensity are normalized for and variations in energy should be corrected for by a change in the scattering vector (s) scaling, data acquisition variations should not block downstream processing.

### Data processing

**Subtraction** – Background corrected (averaged buffer subtracted) macromolecule measurement.

**Merged curve** – combination of all subtractions from a data collection group (corrected for interparticle scattering effects) which corresponds to the scattering from the macromolecule free from concentration effects.

**Structure** – complimentary structural information from other techniques (MX, NMR, EM, etc.) in PDB format for comparison to experimental data

### Shipping

(Note: in brackets are the terms used for MX experiments)

**Stock solutions** – set of samples and buffers (perhaps not diluted) for final preparation onsite in case of additional procedures necessary

**Sample Case (MX Dewar)** – Box with Sample Plates or stock solutions that will be physically sent to the ESRF or brought by the User.

**Sample plate (MX Container)** – Prepared (buffers and or samples in a) 96 wellplate (or other deffined option)

**Shipment** –Set of sample cases sent to the ESRF (or brought at the ESRF by the User) at the same time with the same courier company.

## Definitions

### Assembly

* A name
  + An acronym
* Biological details
  + Molecular mass (kD)
  + Number of individual:
    - Protein chains and residues in each
    - DNA strands and bases in each
    - RNA strands and bases in each
    - Each chain / strand should be given a unique name to allow it to be used for selection for macromolecules part of the assembly
  + Sequence
  + List of associated structures
    - Definition of which part of the assembly the structure is *(to minimize repetition should be linked with macromolecule definitions)* (additionally there could be a number of homologues for the same structure)
      * Could be from sequence information if available or
      * Residues (from – to)
      * Bases (from – to)
* A list of macromolecules which are part of the assembly
* As required by the safety, each assembly has to be described by a sample sheet in the proposal submission form (User Portal / SMIS system).

### Macromolecule

* A name
  + An acronym
* Biological details
  + Molecular mass (kD)
  + Extinction coefficient (M-1 cm-1) (for concentration scaling for UV Spectrophotometer)
  + Number of individual:
    - Protein chains and residues in each
    - DNA strands and bases in each
    - RNA strands and bases in each
  + Sequence
  + List of possible structures (as there could be a number of homologues)
    - Definition of which part of the macromolecule the structure is *(to minimize repetition should be linked with assembly definitions)*
      * Could be from sequence information if available or
      * Residues (from – to)
      * Bases (from – to)
* Assembly formation details
  + Which Assembly(s) this macromolecule belongs to:
  + Definition of which part of the assembly the macromolecule is (sequence, residues from - to or Bases if DNA/RNA) *(multiple regions must be allowed as it could be the combination of two chains in an assembly)*
  + Protonation/deuteration (labeling)

### Sample

* Its macromolecule acronym
* An associated buffer (by definition it is the conditions in which the sample is suspended and is thus the descriptor of its specific conditions and the buffer which should be measured to enable background subtractions) *i.e. if the buffer conditions change each new condition is defined as a new sample (copy paste)associated with the new buffer (with the correct additives).*
* Concentration (Mass or Protein/DNA/RNA) ( **Non 0** mgml-1)
* Measurement temperature (°C),
  + Set point
  + Measured value for each frame
* Its default name will then be:

**MACROMOLECULE acronym\_BUFFER acronym\_Concentration\_TempC \***

(if there are conflicts users should be prompted to distinguish the individual samples)

* Type of measurement,
  + HPLC
  + sample changer
    - viscosity (high/medium/low)
    - Locations (in large experiments multiple wells might be needed to provide the volume necessary for all measurements
      * plate name
      * row and column number
      * volume in well (ml)
* Volume to load (ml)
  + Flow (Yes / No)
* Comments: (text describing the sample)
* As required by the safety , each sample has to be described in a sample sheet in the proposal submission form (User Portal / SMIS system)

\* As a sample could be measured multiple times (for instance to find the best acquisition parameters for radiation damage) the output data (frames) will take the sample name and be appended by run number and frame number (or \_ave).

**MACROMOLECULE acronym\_BUFFER acronym\_Concentration\_TempC\_Run\_(Frame or ave)**

Thus following the example, Aase (Macromolecule A) in the base buffer (B1) at 10mg/ml and 4 °C if data was acquired at run 015 with 10 frames would give, 11 frames named sequentially: **A\_B1\_10\_4C\_015\_*x*** (where ***x*** is the incrementing frame number (1 to 10) and the average file (ave)

*Parameters from all Beamline monitoring devices (machine current, beamstop diode, SC temp, etc.) should be logged for each frame individually and the average.*

### Buffer

* A name
  + An acronym
* Concentration (Mass or Protein/DNA/RNA is **BY DEFFINITION = 0** mgml-1) **Non modifiable value**

*Value of 0 mgml-1 needs to be saved with the measurement details for downstream processing and verification.*

* Measurement Temperature
  + Set point **(Non modifiable by USER).**  *For each buffer measurement it is set by the sample thus you only need to define one buffer for a temperature series.*
  + Measured value for each frame
* Type of measurement,
  + HPLC
  + sample changer
    - viscosity (high/medium/low)
    - Locations (in large experiments multiple wells might be needed to provide the volume necessary for all measurements
      * plate name
      * row and column number
      * volume in well (ml)
* Volume to load (ml) (by default should match that of the sample)
  + Flow (Yes / No) (by default should match that of the sample)
* Composition
  + Type
    - PBS
    - TRIS
    - HEPES
    - Etc.
  + pH
* List of additives

Note: A buffer has no measurement temperature defined but the temperature it was measured at (defined by the sample it matches) must be recorded for verification. This temperature should be used in the naming for the frames collected for the buffer in the same way for samples. For a buffer the naming convention should be:

**BUFFER acronym\_TempC\_Run\_(Frame or ave)**

Thus all like buffers can be easily grouped and checked for consistency and will always be associated with the correct samples as the run numbers will increment.

*Parameters from all Beamline monitoring devices (machine current, beamstop diode, SC temp, etc.) should be logged for each frame individually and the average.*

### Additive

* A name
* A Quantity (mM or if the additive is deuterium %)

### structure

* A PDB file or accession code
* List of what data this structure is associated with
  + Assembly
    - Macromolecule

### Session

* Dates
  + start
  + end
* a number of shifts
* a beamline
* a local contact

Note: 1 Beamtime proposal may have multiple sessions, each session could have multiple data collection arrays

### Stock solutions

* Will need instructions for storage once they have arrived
  + Room temp (20°C)
  + Fridge (4°C)
  + Freezer (-20°C)
  + Cryo-cooling (-80°C)
  + Comments for special instructions
* Extra instructions for preparation of dilutions ( to make individual samples / Sample plate)

### Sample Plate (MX Container)

* Storage Temperature (°C) should be defined for each plate to set SC thermostat for correct conditions (as 3 plates can be stored at once if there is a difference in the plates in the SC a warning should be given for the users to decide to either override (set new temp for all plates in SC) or to remove plates from SC)

### Sample Case (MX Dewar)

* List of Stock solutions and Sample plates contained
* Will need instructions for storage once they have arrived
  + Room temp (aprox 20°C)
  + Fridge (4°C)
  + Freezer (-20°C)
  + Cryo-cooling (-80°C)
  + Comments for special instructions

### Shipment

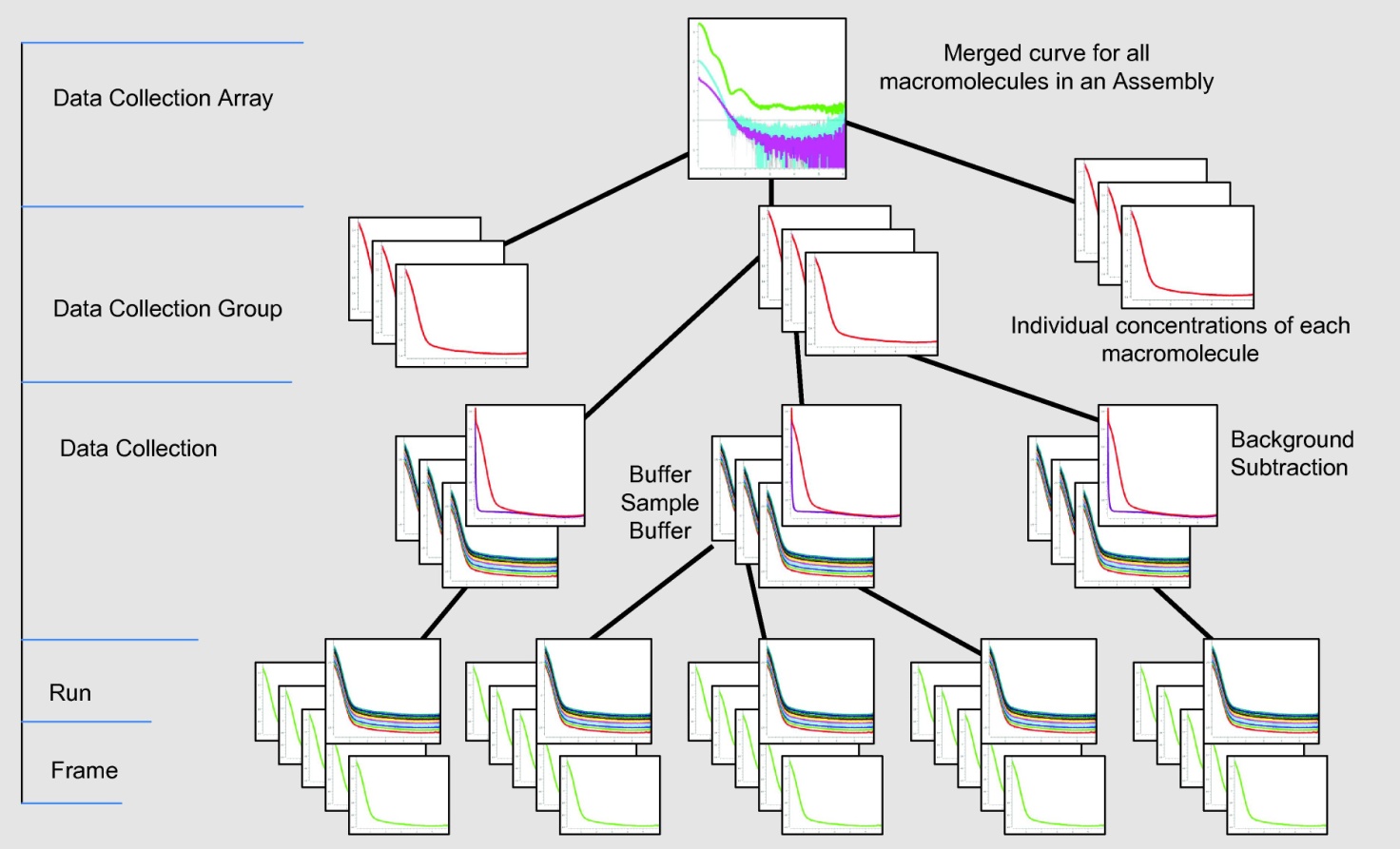
* User (return/Contact) details
* List of Sample cases included in the shipment

## Overview of heirarchy in data model

### Biological description of hierarchy from assembly to sample and buffer

### heirarchy2_crp.JPG

### Data collection

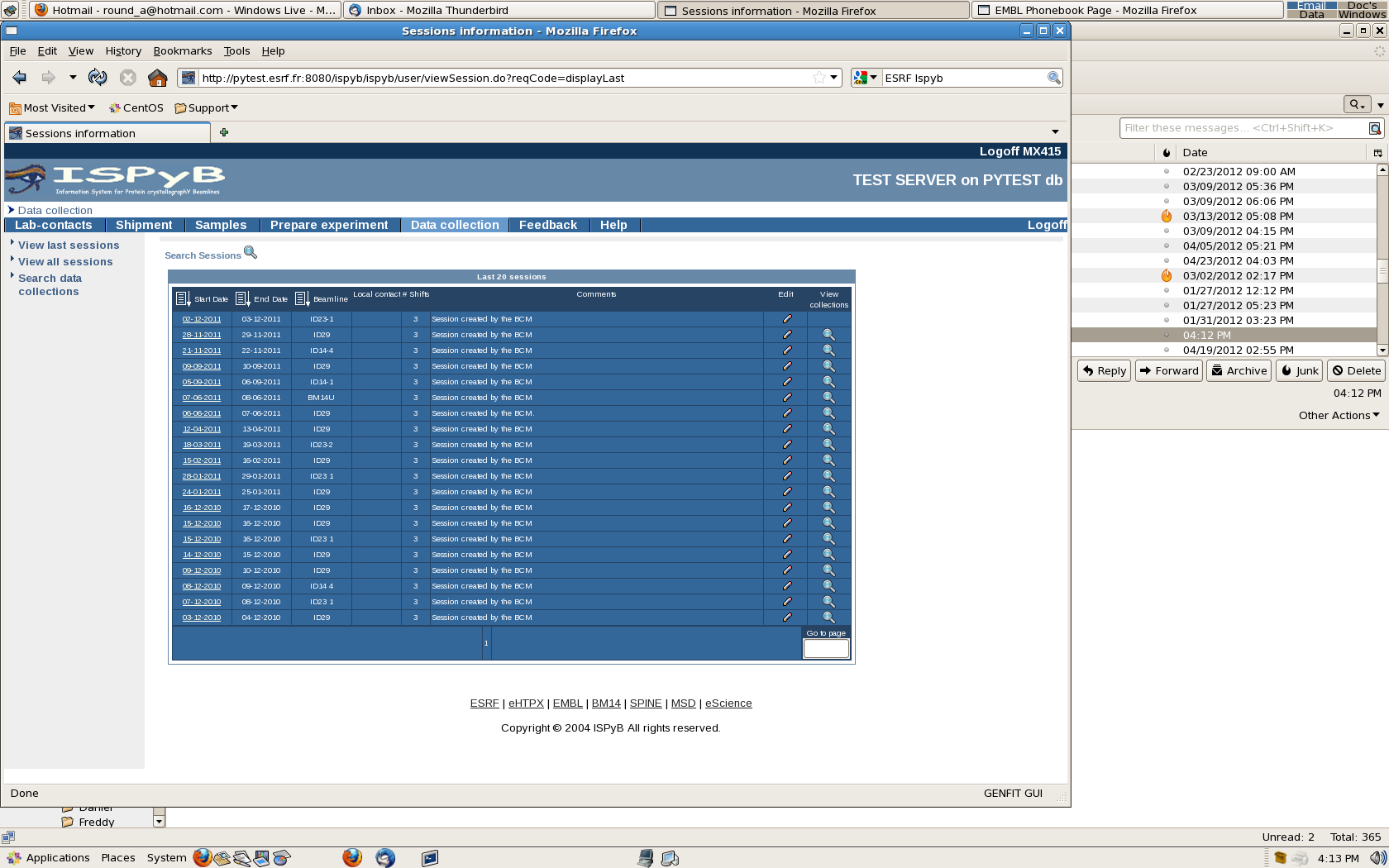


## Screenshots

### Session List

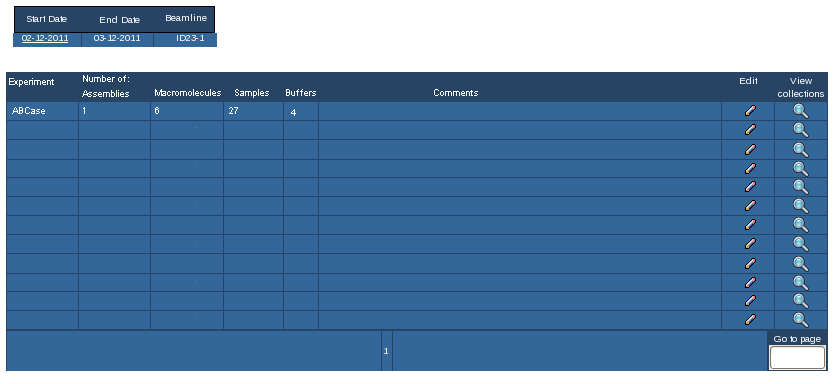
This screen shows all the sessions that are available for the given proposal. Clicking on the “Edit” button of a session brings you to the Experiment List of this session (see next screen).

Note: This session list will contain all MX and BioSAXS sessions – by selecting a BM29 session the users interface will change to best accommodate BioSAXS information displayed. It might be good to have a toggle to display MX sessions or BioSAXS sessions alone.



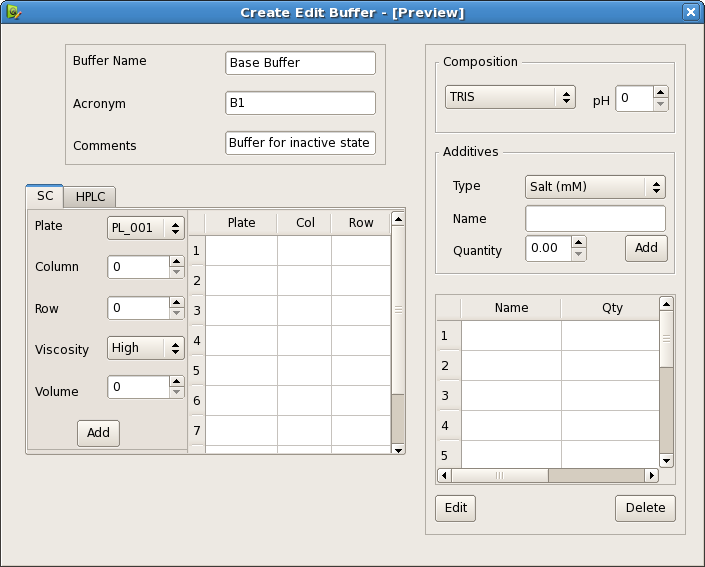
### Experiment List

This screen shows all the experiments of a given session. You can Delete, Edit an existing experiment or Create a new one (see next screen).



### Create / Edit Buffer

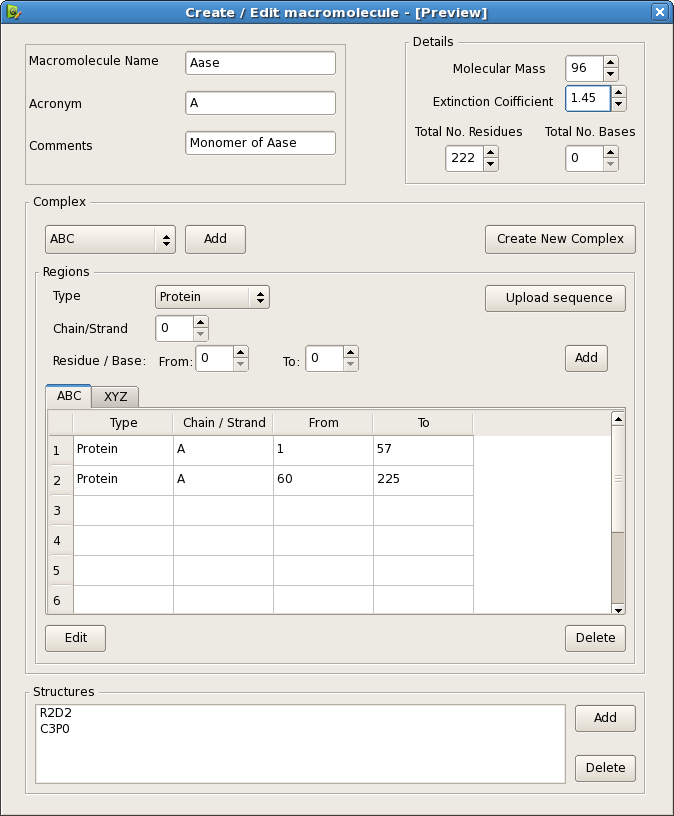
A buffer is now the specific conditions and additives in which a sample will be and is associated with. It also has the positions in which all like buffers can be found.



Multiple defined positions in SC are required

### Create / Edit Macromolecule

The macromolecule contains all the information regarding the context to the assembly and biological details.



### Create / Edit Samples

Sample contains just the specific information for the individual measurement of a macromolecule under 1 condition at ONE concentration, as the only variables changing are concentration, temperature and the position of the sample the definition of samples should allow defining multiple samples (different concentrations / temperatures quickly and easily).

As was observed from previous use a visual representation of the SC and sample positions in a plate would be very helpful.

Note: Perhaps a variable in the sample information would be useful for protonation/duteration (labeling) for contrast variation experiments but as it is irrelevant for X-rays it could also be included in the comments

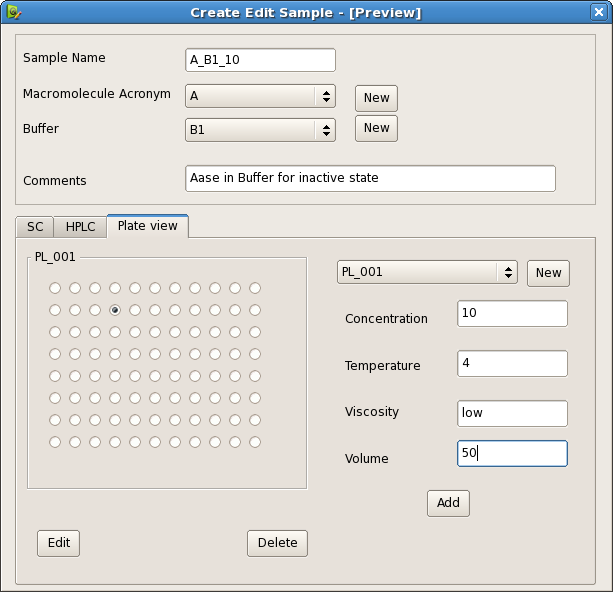
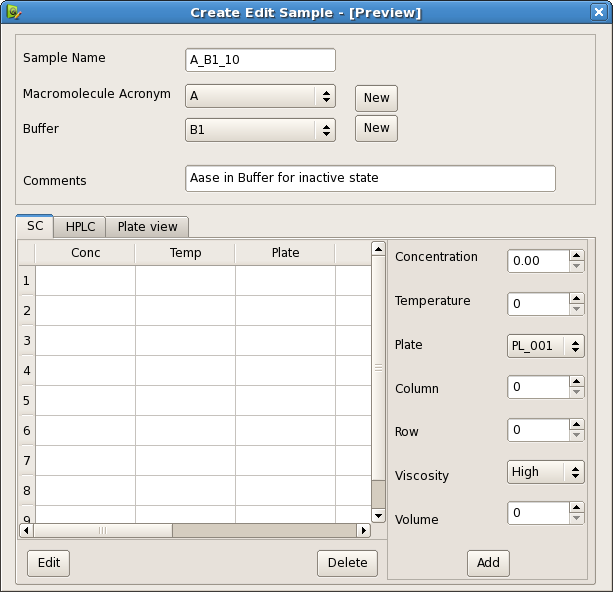
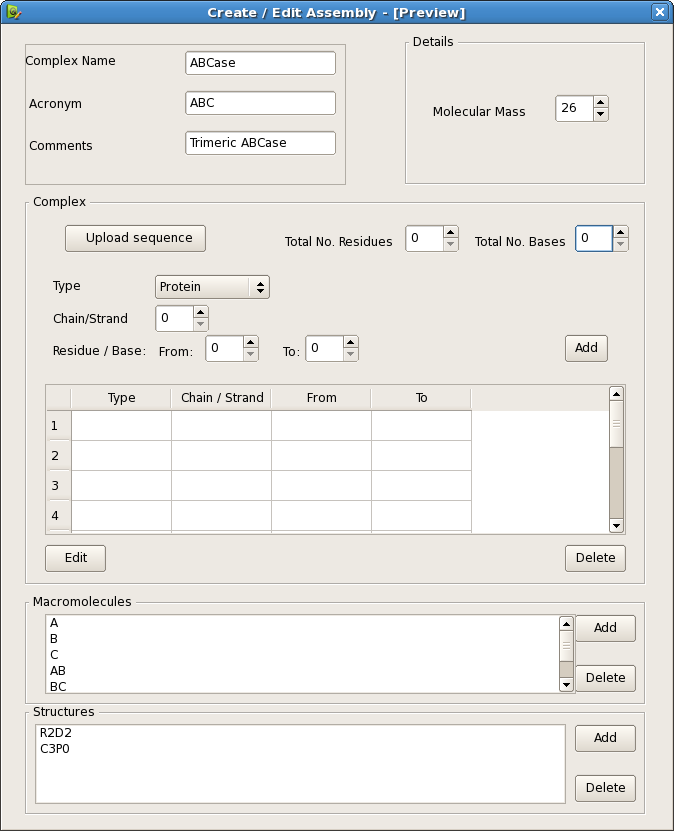


Table View Plate view

### Create / Edit Assembly

An assembly not only holds the list of all macromolecules it contains but the biological details of the whole assembly\*

\*Specific information on how the macromolecules are related (fit in) to the assembly is held in the macromolecule itself



## Wizards for streamlining user interaction with ISPyB

Using only the dialogs (presented above) to input all the information will be time consuming and in many cases the information will be the duplicated. Additionally as the users will be aware of the assemblies they have and the macromolecules they wish to measure but not always be aware of the samples needed to undertake a thorough experiment. Guiding the users through the generation of samples based on the macromolecules defined. The interaction through ISPyB (using a wizard) should allow retrieval of already defined samples macromolecules and assemblies and addition of new ones as necessary.

Work flows can be separated into 3 distinct types.

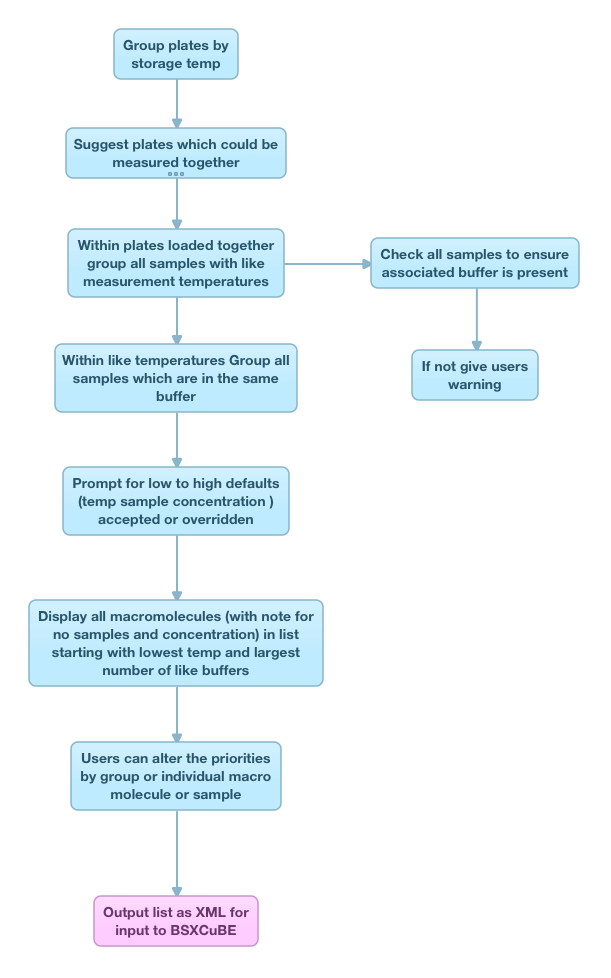
* Static
* Kinetics
* HPLC

For each case the users need to be prompted for all required information and not only the database entries filled appropriately but the suggested measurement sequence provided which should be editable to give freedom for more complex situations that are out of the scope of the wizard.

### test3.jpgStatic

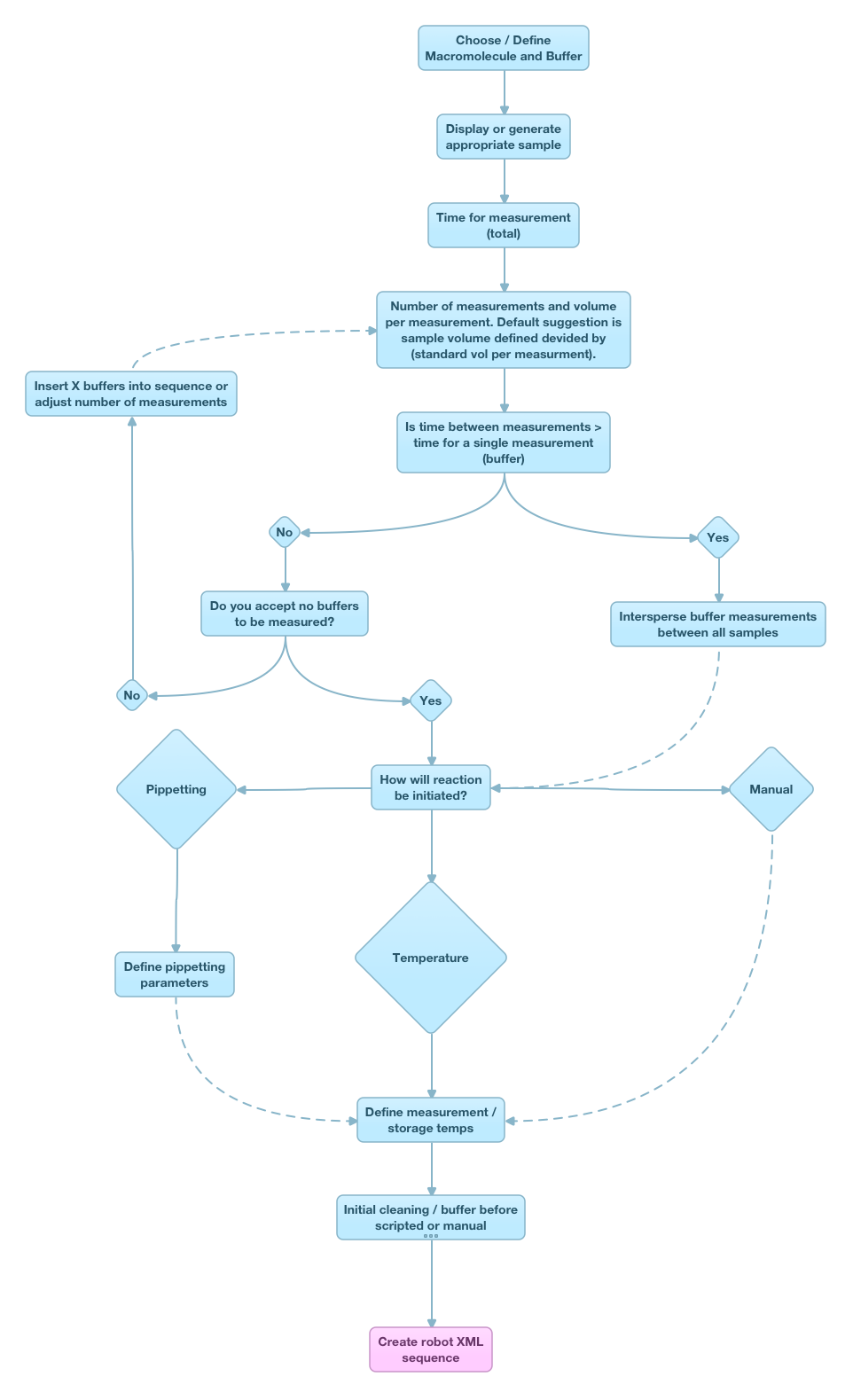
Static measurements are the current standard measurements made at the BioSAXS Beamline. To measure the proteins in solution requires buffer to be measured before and after every sample, minimum 3 samples per macromolecule and can include many macromolecules in an assembly at many different buffer conditions. With defined macromolecules and buffer conditions the 3 samples can be automatically generated with an interface for the concentration to be added (possibly later, at measurement time) and the location of each sample in the plate.

Once all the samples have been defined the sequence for the data collection can be automatically generated suggesting which plates can be loaded together (based on SC configuration) and the sequence the samples should be measured in to be most efficient.



### Kinetic

This kind of experiment is designed to follow a reaction form start to finish (minutes or hours) and although this in principle will be one sample (at by definition one condition). It could be a measurement of one aliquot for the whole experiment or it could be separate volumes from a large well and in this case, buffers could be measured as well. An additional complication is the decision for the number of runs to be acquired during the reaction time.

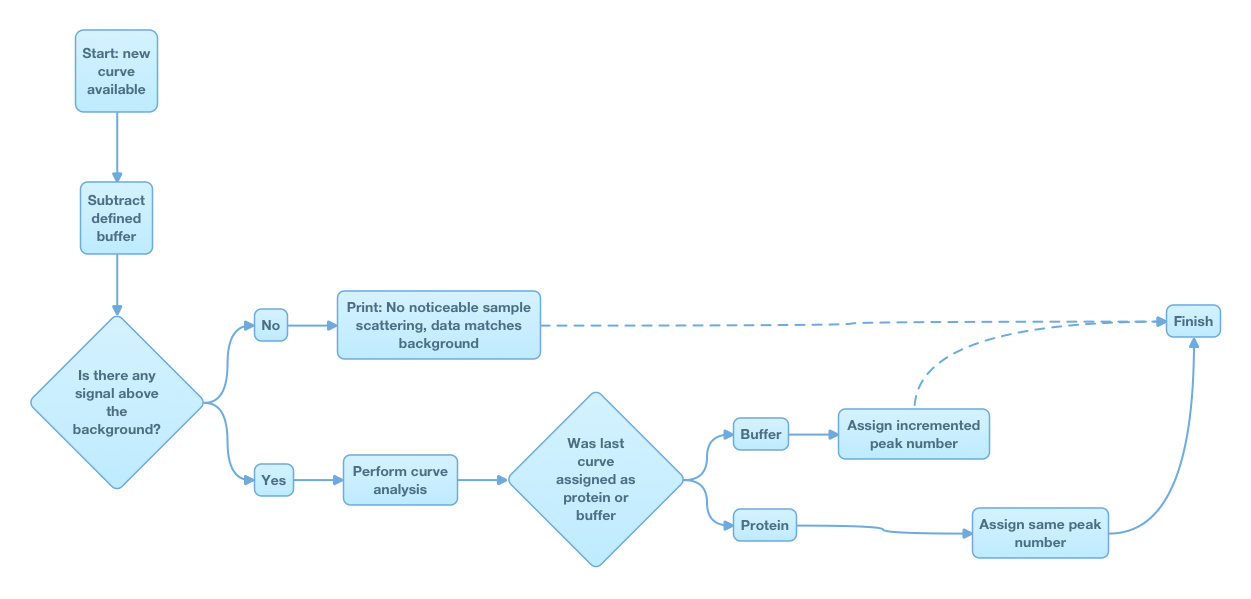


### HPLC

The exact nature of the HPLC experiments has yet to be determined but we can anticipate that the macromolecule will only need one sample as the concentration will be varied naturally as the sample elutes from the column and the varying concentration will need to be recorded (from spectrophotometer reading) with the frame for later normalization.

HPLC operation will probably not need an XML file generated in advance to define the measurement sequence but it should be used to keep the log of what was done. As it could be that the measurement cell gets dirty and the HPLC run interrupted to allow cleaning cycles using the SC.

There will be an increased need for online data analysis with the HPLC system in order to give feedback on which frames correspond to buffer or sample and which part of the assembly (or in which oligomeric state) the peak corresponds to.

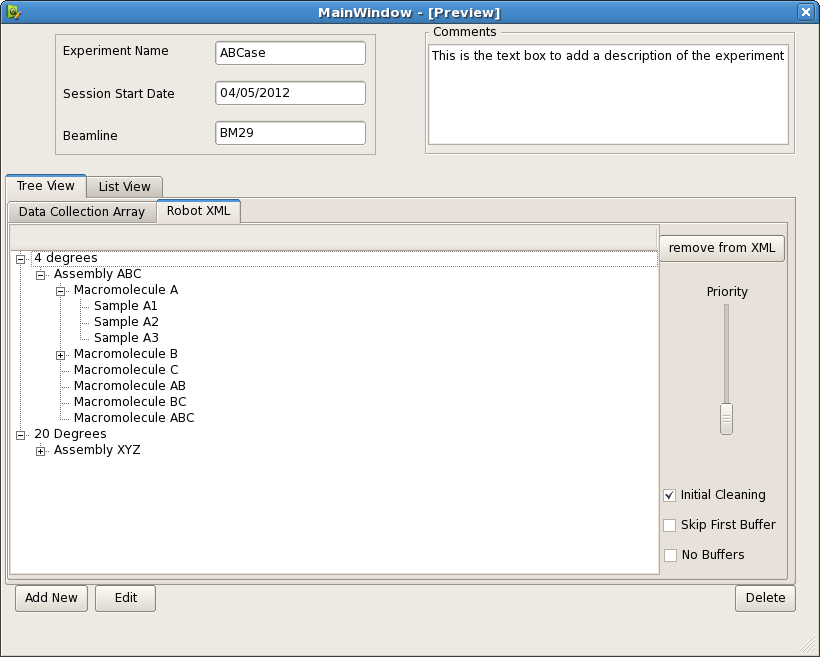
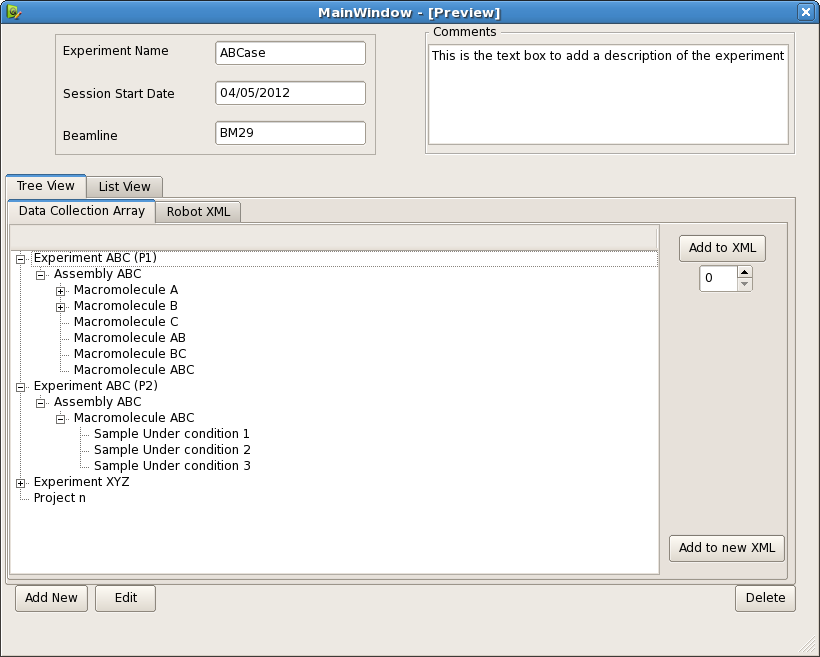


### Edit Robot XML / Data Collection Array Sequence

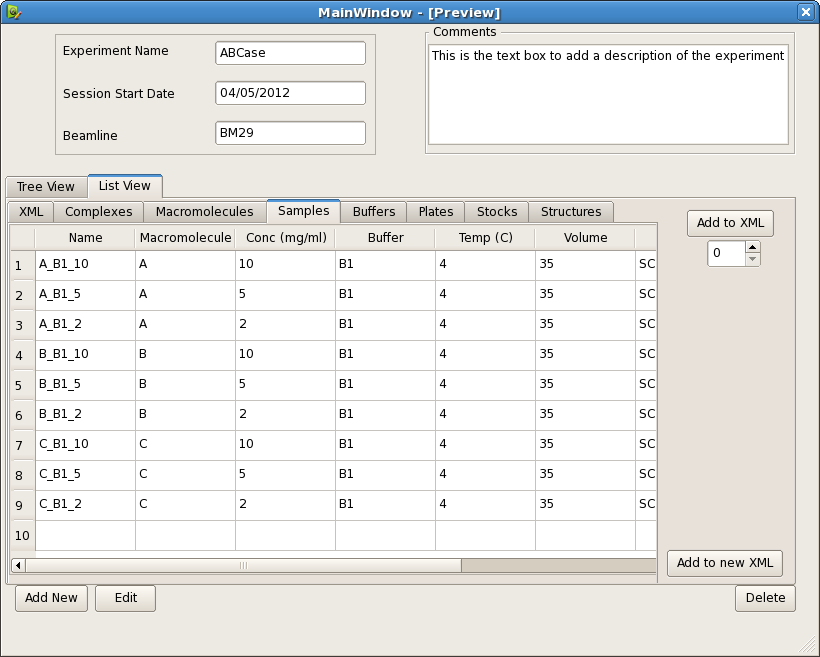
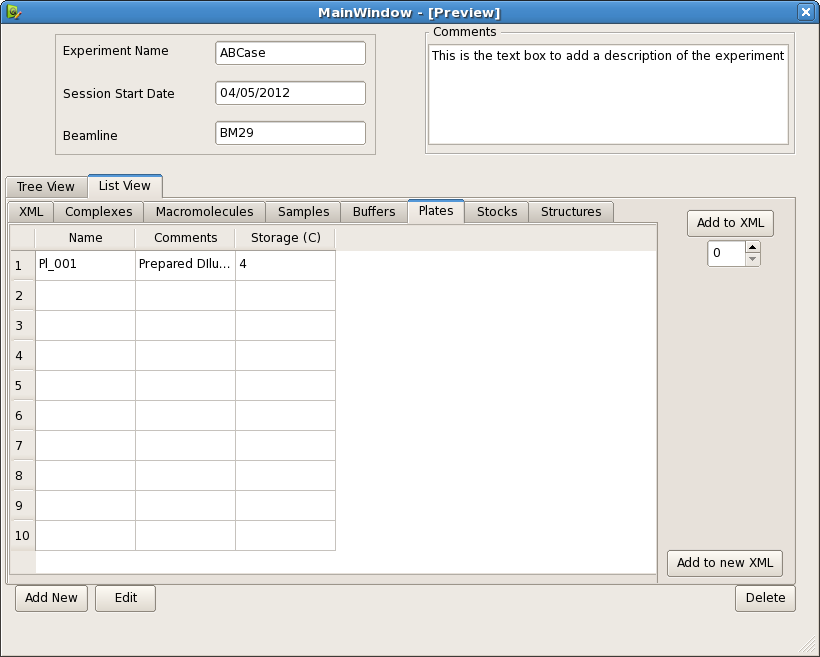
A data collection array is defined a number of data collection groups which in turn consists of a number of measurements of buffers or macromolecule solutions and as such can be listed in order of the sequence of measurements. This interface can be used to edit the information in the data collection array and to create the robot XML which can be read into BSXCuBE to control the data collection. However, the robot XML does not have to be the full data collection array as only a subset of samples my actually be in the SC. The data collection array needs to hold the information for all required samples, their buffers the macromolecules and in turn the assembly the samples belong to, but the robot XML only needs to contain the sample information for the indented data collection.

Pressing Add or edit should take the user to the appropriate dialog depending what (list, or part of the tree) is active. If nothing is active a dialogue can be offered.

Note: Any level of hierarchy can be added or removed from the XML list including individual buffers to whole plates or assemblies. Extra blank measurements of buffers with excessive additives can be measured to ensure the additive has no effect on the scattering. Thus it is possible for there to be buffers that are not related to any sample.



Tree View



List View