During Experiment

Contents

[1.1 List of terms 1](#_Toc320711520)

[1.1.1 Data classification hierarchy 1](#_Toc320711521)

[1.1.2 Data aquisition 1](#_Toc320711522)

[1.1.3 Data processing 2](#_Toc320711523)

[1.2 Definitions 2](#_Toc320711524)

[1.2.1 Measurement / Run 2](#_Toc320711525)

[1.2.2 Series 2](#_Toc320711526)

[1.2.3 Data collection 3](#_Toc320711527)

[1.2.4 Experiment 3](#_Toc320711528)

## List of terms

### 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)*

**Series** – combination of runs (buffer\_before, macromolecule at one concentration and buffer\_after).

**Data collection** – combined series for one macromolecule (minimum 3 concentrations). *In a sequential data collection, 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 series and the buffer\_before in the subsequent series.*

**Experiment** – The group of data collections that are required to answer a biological question.

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

Depending on the complexity of the system for some cases (structure validation for example) an experiment is an individual data collection. However, for others the experiment could be in a number of related parts (see example)

### Data aquisition

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

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

Acquisition parameters should normally be kept the same for all **measurments** in a **data collection** and all **data collections** in an **experiment**. However, as some **samples** my 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 **series**, 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! 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 (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

## Definitions

### Measurement / Run

* Run Number (increments with each successive **measurement** and should be unique for a **session\*1**)
* The **sample** or **buffer** name that data was collected from
* A number **frames**
* Time per **frame**
* Time
  + start
  + end
  + total duration
* Beamline parameters (non exhaustive list)
  + Energy
  + Diode readings
    - Beamstop diode (the only diode permanently in the beam should be average for each frame)
    - Other diodes can be put into the beam using pneumatic actuators and readings could be taken upstream (from sample) during the cleaning cycle of the SC
      * Diode just before sample
      * After mirror
      * After mono
      * White beam
  + Machine current
  + Transmission (%)
  + Measurement of beam (before / after / every *n* measurements / *j* times per day\*2)
    - Size
    - Shape
    - profile
* Sample parameters
  + Type
    - SC
    - HPLC
  + Measured Storage temperature (per **frame** and average)
  + Measured exposure temperature (per **frame** and average)
  + Spectrophotometer readings
  + Video feed (from SC) of sample in measurement position
    - snap shots?
      * Start
      * End
      * Intermediate (1,5 10 per second?)
    - Video?
* Scattering (identified by **frame** number or ave\*3)
  + 2D images
    - Raw
    - Intermediate
  + 1D calibrated scattering curves

\*1 Could be made unique for an experiment only and separate data by experiment. This is perhaps dependant on the architecture of ISPyB and needs to be discussed.

\*2Depends on technical solution for measurement on Beamline.

\*3 In the case of a HPLC measurement the ave will also require an incremental variable to take account of multiple elution peaks per run.

### Series

* Name of average file (including run number)of
  + **Buffer** before (B1\_4C\_016\_ave)
  + **Sample** (A\_B1\_10\_4C\_017\_ave)
  + **Buffer** after(B1\_4C\_018\_ave)
* Intermediate processing files
  + **Buffer** average
  + **Subtractions**
    - **Buffer** before only
    - **Buffer** after only
    - Average buffer
* Output **subtraction** (A\_B1\_10\_4C\_017\_Sub)
* EDNA processing results
  + Model independent parameters
    - Rg
    - I0
    - Volume
    - Dmax
    - Estimated Molecular mass
  + Ab initio models
    - PDB
    - Fit to Exp data
    - Averageing (with DAMAVER)
      * Most representative model
      * Average model
      * Filtered model

### Data collection

* List of all **subtractions** (output of each series)
* **Merged** (concentration independent scattering curve)
* EDNA processing results for **merged** curve
  + Model independent parameters
    - Rg, I0, Volume, Dmax, Estimated Molecular mass (and comparison with expected)
  + Ab initio models
    - PDB
    - Fit to Exp data
    - Averageing (with DAMAVER)
      * Most representative model
      * Average model
      * Filtered model
* Fit of any **structure(s)** for the **macromolecule**
  + Chi
    - Fit to Exp data
* Overlay of ab-initio models and best **structure** (supcomb20)

### Experiment

* List of all merged (**macromolecule**) scattering curves
* For studies on
  + A multi component **Complex** (example P1 from prepare experiment)
    - DAWB / DAWN Workflow to be developed to suggest and execute strategies depending on information available.
  + The same **macromolecule** under different conditions (example P2 from prepare experiment)
    - Superimpose DAMAVER models for each **macromolecule** under each condition
    - Comparison of all Model independent parameters (EDNA pipeline)
      * Rg
      * I0
      * Volume
      * Dmax
      * Estimated Molecular mass
    - Plot of parameters against variables altered
  + DAWB / DAWN Workflow to be developed to suggest and execute strategies depending on information available.