

# BEST PRACTICES IN DATA ANALYSIS & REPORTING

This guide aims to provide insights into different aspects of data analysis for biocalorimetry. There is not always one right answer, and various options could be applied in different situations. The guide provides suggestions in the hope of unifying the way calorimetric data is analyzed and presented in publications.

## Aspects of data

1. Baseline
2. Nomenclature of parameters
3. Qualitative vs quantitative data
4. Presentation of data
5. Normalization of data
6. Parameters for materials and methods

## 1. BASELINE

By definition, the baseline serves as a zero value. It is the position where no heat is released or absorbed. However, there will be some deviations from a zero value due to sample vial placement on the sensor and noise in the system. Depending on the experimental design, it might be important to correct for the baseline value. The baseline could either be set using a metabolically inert state of the sample during the experiment (internal) or by using an external value.

The internal baseline could be defined during the lag phase of prokaryotic experiments or at the end when the metabolic activity has returned to zero (Fig. 1). Keep in mind, however, that it is not always easy to know if metabolic activity has returned to zero or is merely continuing at very low levels.

An external baseline could be used for experiments where there is no lag time and the metabolic activity

does not reach zero before the end of the run. In those cases, a separate plate with inert media or sample that matches the reference should be run, preferably for 24 hours to establish the baseline read. Keep in mind that readings from two runs might not perfectly align. It is, therefore, beneficial to repeat the empty run a couple of times to establish the reproducibility between runs for each calScreener.

The length of the baseline read depends on the system's stability. The longer the baseline read is the better it represents the no-heat flow read. If there is a large room temperature variation, a 24-hour read should be considered to capture the variation. However, in reality, this is not often possible. In those cases, a period of

at least 30 minutes should be chosen. The baseline is calculated individually for each sample vial/calorimeter channel. It can, therefore, be a different time window for each channel. You cannot use the baseline from one channel for another. Ideally, the baseline should be outside the region of interest that you will analyze.

In some cases, runs can be analyzed without a baseline. For example, when looking at when things happen rather than total heat produced, the baseline setting does not affect time-to-peak. Similarly, microbial samples with high metabolic activity and substantial variations in sample properties might bypass the need for baseline correction.

It is important to assess a baseline relative to the ongoing measurement, considering factors such as measurement duration, heat flow levels, and acceptable result variations.

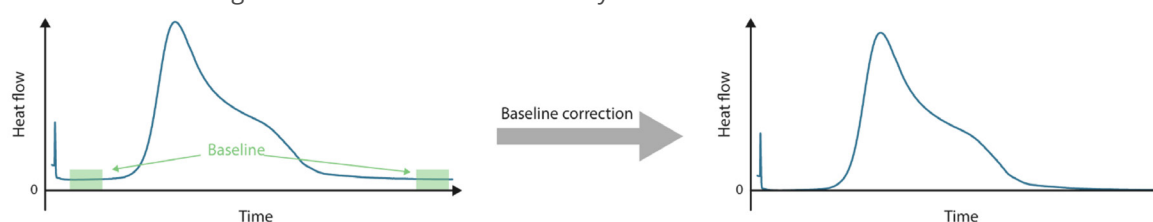


Figure 1

## 2. NOMENCLATURE OF PARAMETERS

In the realm of isothermal calorimetry, a specialized nomenclature exists to describe the various components of thermograms precisely. Understanding and employing this nomenclature is vital for researchers to accurately communicate, interpret, and compare results obtained through isothermal calorimetric experiments. This paragraph serves as a reference for various terms used in publications.

Commonly used terms include **heat flow/(thermal) power/ heat production rate**, which represents the thermal energy exchanged during a reaction. It is also referred to as metabolic activity when talking about biological systems. This is the direct reading from the calScreener, which is measured in (micro)watts ( $W=J/s$ ) (Fig. 2A).

**Total heat** is determined from the area under the curve and represents the total thermal energy released during a given time period. It is important to use the same time window when comparing the total heat between samples (Fig. 2B).

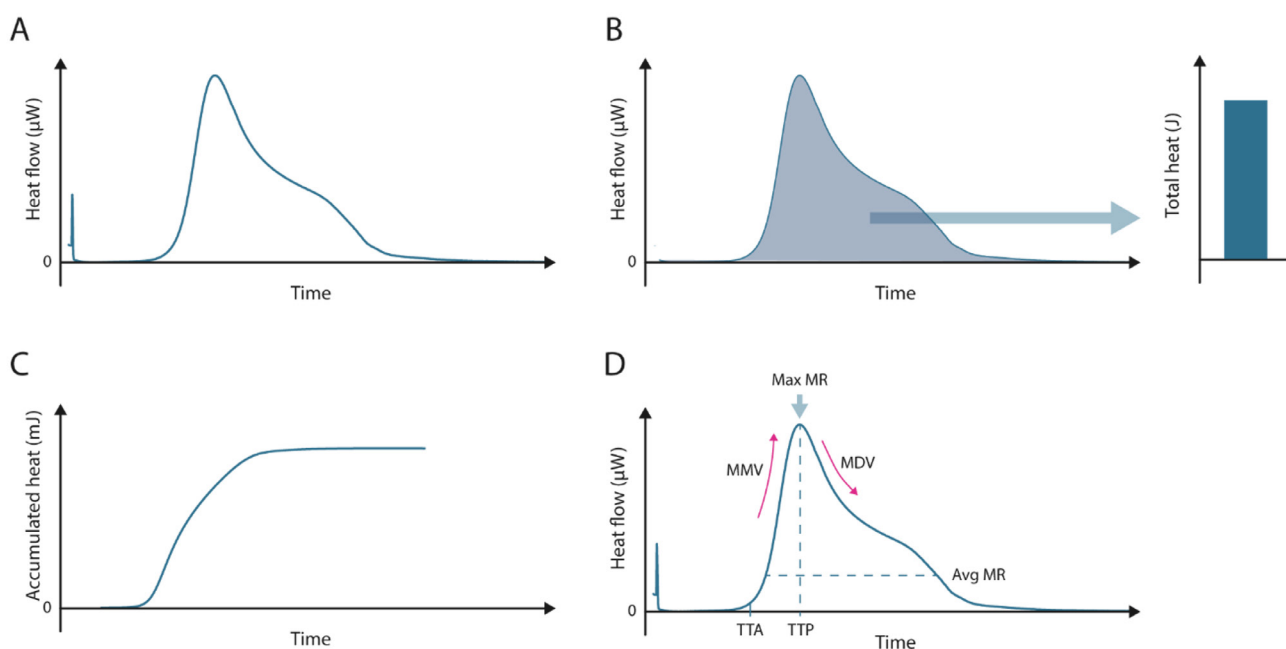
**Accumulated heat/cumulated heat** is the integral (sum of the heat flow multiplied by the time step between signal points) within the region of interest (ROI). This curve is more like a classical growth curve and could, therefore, be easier to interpret for microbiologists. Furthermore, looking at the accumulation of heat is beneficial when looking at organisms with very low thermal output. This data is shown in joules (J) (Fig. 2C),

**Max metabolic rate (Max MR)/peak power** is the maximal heat flow achieved in the thermogram. In simple terms, it is the height of the peak. This value comes in handy when investigating treatment effects on metabolism or designing an optimal growth media. The Average Metabolic Rate (Avg MR) is closely related to the previous parameter. It is calculated as the cumulative value at the end of the ROI divided by the time from the beginning of the ROI to the end of the ROI (Fig. 2D).

**Time to activity (TTA)** is often shown in experiments that aim to display the speed to detection or initial inoculum size. It is important to define how this measure is taken – a cut-off value vs percentage from the peak power. TTA often correlates with time to peak (TTP), which shows when the peak of the thermogram occurs.

**Maximum metabolic velocity (MMV)** ( $\mu W/h$ ) – is defined as the maximum positive slope of the heat flow curve, representing the rate at which heat flow increases most rapidly before reaching its peak. This value can indicate the maximal growth rate and the efficiency of nutrient turnover.

**Maximum decay velocity (MDV)** ( $\mu W/h$ ) is the opposite of the previous value. It is defined as the maximum negative slope of the heat flow curve, representing the rate at which heat flow decreases most rapidly after reaching its peak. This value can indicate the rate of decline in metabolic activity.



**Figure 2** Graphic representation of calorimetric parameters.

### 3. QUALITATIVE VS QUANTITATIVE DATA

In most cases, IMC results are qualitative and rely on comparisons with controls. For example, metabolic activity of 10  $\mu$ W in a treated sample does not provide much information unless the metabolic output from the control is known.

To obtain quantitative data, standard curves need to be produced. Often, this involves complementary methods. However, when correlations between heat flow data and various phenomena are established, IMC becomes a powerful screening tool.

Lately, principal component analysis (PCA) has gained traction. This is a useful tool for evaluating minute differences between thermograms. Grouping the samples can reveal their relationship.

The choice of analytical method is tightly bound to the research question and the assay used. An example could be an analysis of the optimal culture condition of a hard-to-grow species. The assay might rely on small adjustments to media components, atmosphere, or temperature. In such a case, a simple quantitative analysis of the Total heat output over a given time period might be sufficient.

If the researcher would like to understand the phenotypic difference between a large number of genotypically different strains or even species, a qualitative PCA plot might provide a nuanced picture of the phenotypic likeness of the various thermograms.

### 4. PRESENTATION OF DATA

Example thermograms should be included in the publications, at least in the supplementary information. Thermograms allow the knowledgeable reader to evaluate the suitability and meaning of the parameters presented.

Kinetic curves should always be shown as representative examples. Averaging thermograms reduces fine details. Instead, averages of parameters like total heat and maximum metabolic activity should be shown.

### 5. NORMALIZATION OF DATA

For certain types of samples normalization is required, for example, samples where the weight and volume between vials differ (soil, powders, biopsies, etc.). All eukaryotic experiments should be normalized to cell numbers in the end as the readout is dependent on the number of cells.

Although normalization is a useful tool when comparing different types of samples, the sample volumes, weights, and cell numbers should be kept as equal as possible as they might influence other aspects of the samples, such as available head space.

Heat flow normalization to the amount of drugs in a sample is normally not used.

### 6. PARAMETERS FOR MATERIALS AND METHODS

IMC results can be influenced by many factors, and it is therefore important to describe them in the materials and methods section. Sample volume could influence the signal intensity. Depending on the volume, there is a difference in head space above the samples which would influence oxygen availability. Headspace is also influenced by the use of plastic inserts that reduce the total volume.

In some cases, it might be necessary to note down the atmospheric conditions or the relative humidity during vial preparation.

As the nomenclature can differ, it is a good idea to define the different parameters that are presented. For example, how time to activity was defined.

When presenting enthalpies (total heat), it is important to disclose the integration method and how the baseline was defined/recorded.