The primary purpose of this dissertation was to produce a general construction methodology for two-phase experiments when the second phase experiment involved either a four- or eight-plex iTRAQ proteomics experiment, although the methodology we have produced is more widely applicable. The outcome of this research has three primary components. Foremost, an algorithm for information decomposition of data compiled from single- and two-phase experiments was developed to automate the construction of theoretical ANOVA tables to enable fast assessment of the attributes of competing conceptions of the same size. This algorithm was implemented in an R package called infoDecompuTE which is available on the Comprehensive R Archive Network (CRAN). Second, algorithms were built up for identifying optimal designs for two-phase experiments where the material in the Phase 1 experiment is arranged in a completely randomised design, a randomised complete block design or a balanced incomplete block design. Tables of optimal designs have been produced. The last facet of the research was part shows how to calculate the variance components and effective degrees of freedom when restricted maximum likelihood is used for the statistical analysis of the information.

One main concept that was not handled in detail in this thesis was unbalanced design. The design is not balanced when the number of experimental units is not equally replicated. The unbalanced design can occur very often in biological experiments and can be divided into two issues: 1) design can become unbalanced when the part of experimental dataset are lost or 2) the researcher has to fit the sample using unbalanced design due to budget. The missing value issue is very common in the MudPIT-iTRAQ$^{TM}$ experiment, where some proteins may not be detected in a whole run of the experiment. One path to overcome this matter is to calculate the missing values, notwithstanding, this procedure cannot be automated, because the construction of the ANOVA table still needs to be made based on the unbalanced experiment. For the second issue, this is also very common to see the researcher can only design an experiment where treatment replication is not identical due to budget. For instance, a biologist was trying to compare 3 treatment groups using 10 single-colour microarrays. Thus, one way to design such experiment is to have one of three treatments replicated 4 times and the other two replicated 3 times.

One main disadvantage of unbalanced design is the treatment canonical efficiency factors are not identical, this means treatment information will appear to be high in some treatment comparisons. This attribute has been remarked in Chapter 3 and 4 in some optimal two-phase experimental designs. Nonetheless, these optimal designs are unbalanced is due to the confounding of the treatment effects with a run or tag effects, they are not caused by unequal treatment replication.

In the two-phase experiment, unbalanced design can happen in three cases: 1) only the Phase 1 experiment is not balanced, 2) only the Phase 2 experiment is not balanced and 3) both Phase 1 and 2 experiments are not balanced. In constructing the ANOVA table using the information decomposition, more investigation is required to each of these events. In an unbalanced factorial design, the order of how treatment factors are fitted can affect the social organization of the ANOVA and the treatment efficiency factor. If the experiment is two-phase, then the Phase 2 block factors are fitted first before the Phase 1 block factors and then the treatment factors. If the design is unbalanced, the main complication is on the order of factors are fitted within each structure and on how does it affect the final structure of the theoretical ANOVA table.

More work is also needed in developing the objective function for optimal two-phase design where either or both Phase 1 and Phase 2 are not balanced. The main focus on the objective function should be on the treatment canonical efficiency factors as well as how the treatment comparison are made, i.e.\ the eigenvectors, because treatment canonical efficiency factors are unlikely going to be identical and some treatment comparison may not the primary interest by the researchers. Therefore, the construction of the objective function may need to take the eigenvectors into account.

The further direction of the biological experiment is striding into \emph{Next-Generation Sequencing} (NGS) for the DNA sequences, due to the rapid reduction in the sequencing cost and the increase in the quantity of information that can be gathered. In 2001, the first human genome required 15 years to sequence and cost about 3 billion US dollars. In contrast, researchers now can sequence 45 human genomes in a single day for about 1000 US dollars for each genome. The multi-plexing technology in the NGS can be carried out by adding unique index sequences, namely \emph{barcodes}, to each DNA fragment. Therefore, the researchers can attach different barcodes for different samples, allowing NGS to sequence multiple samples simultaneously. The expression levels of the sequences are then based on the amount of barcode presented. This idea is very similar to the iTRAQ$^{TM}$ tags when measuring the protein abundances. Therefore, the theory of the two-phase experiment should also be applied in the NGS allowing a better experimental design and result from the data analysis.