CHAPTER 1

INTRODUCTION

1.1 Background

Antibody is a type of globulin protein synthesized by plasma B cells in the serum and tissue fluids of human body. It protects the human body against microbial infection (Parija, 2012). There are five classes of immunoglobulin found in mammals, which are IgA, IgG, IgM, IgE, and IgD that vary at the domain structure. IgA is dimeric in structure; IgM is pentameric, whereas IgG, IgD and IgE are monomeric. Each of them has their respective roles in the immunity system (Parija, 2012). IgG, which is a small molecule that is able to penetrate through tissues easily, appears to be the most abundant immunoglobulin in human body. IgG can be divided into four subclasses; IgG1, IgG2, IgG3 and IgG4 rendering to their concentrations in serum (Zabriskie, 2009).

Recently, monoclonal antibodies (mAbs) is utilized as therapies and molecular drug targets against diseases such as cancer, chronic inflammatory diseases, transplantation, infectious diseases and cardiovascular diseases (Liu, 2014). This leads to a huge demand for monoclonal antibodies in the pharmaceutical industry due to increased number of patients for the diseases mentioned (Shukla, & Thömmes, 2010). Cell lines used for the production of monoclonal antibodies are generally produced from mammalian cells as mammalian cells are able to fold properly, assemble the desired proteins and capable to perform human posttranslational modifications (Zhang, 2010). Cell lines that are commonly used to produce monoclonal antibodies are Chinese Hamster Ovary (CHO) cells, mouse myeloma (NSO) cells, human embryonic kidney (HEK) cells, PER.C6, baby hamster kidney (BHK) cells, using hybridoma and phage display techniques (Bandaranayake, & Almo, 2013) (Liu, 2014).

Among the cell lines mentioned, CHO cell lines is the most widely used cell lines for the production of monoclonal antibodies (Kou et al., 2011). This is due to the known gene

sequence of CHO cells (Kou et al., 2011), its ease of adaptation to serum free conditions, ease of maintenance and ability to perform glycosylation (Bandaranayake, & Almo, 2013). As the demand for monoclonal antibodies rises, the productivity of CHO cell lines will be significant and various strategies are developed to increase the productivity of CHO cell lines (Kou et al., 2011).

The productivity of CHO cell lines can be categorised into two; firstly is volumetric productivity and secondly is specific productivity *(q_p)*. Volumetric productivity refers to the total amount of monoclonal antibody produced by a cell. On the contrary, specific productivity denotes the amount of desired products produced by a single cell which is measured in g/L/cell. In the case of cell line development, the decisive objective is to select the cell lines with high specific productivity from a pool of cell lines with varying specific productivity from a pool of cell lines with varying antibody production (Zhang, 2010).

In identifying and selecting cell lines with high specific productivity, the interactions and correlations between the process parameters or physiological characteristics need to be determined as it is significant for controlling the production of biopharmaceutical products by cell line (Zalai et al., 2016). Hence, several physiological characteristics of cell line are investigated in this study to determine their effect and correlation on the specific productivity of CHO cell line that produce monoclonal antibody. Physiological characteristics, related to physiology, refers to the branch of biology that affects the normal functions of living organisms and their parts (Roux, 2014). The physiological characteristics studied are total intracellular protein, biomass, cell size, cell density and cell volume.

Multivariate statistics (MVS) or multivariate analysis would be a suitable method to deal with the complex relationships between the physiological characteristics as MVS is a multitude of statistical methods that helps to analyse the potentially numerous interrelated variables by considering them together rather than separated from each other (Raykov & Marcoulides, 2012). To reduce the dimensionality of the variables, Principal Component Analysis (PCA) will be one of the statistical method applied in MVS to solve this problem.

Principal Component Analysis (PCA) is carried out on the data obtained using STATISTICA software to determine the correlation of the physiological characteristics on the specific productivity of CHO cell line. PCA is chosen as it is one of the most important and powerful method in chemometrics (Bro & Smilde, 2014) especially for large number of data sets. The main aim of PCA is to determine the correlation between variables within huge

number of data. Besides, PCA aids in data reduction where the data will be reduced to a moderately complex model structure in PCA (Wold, Esbensen & Geladi, 1987). In other words, PCA simplifies data, carries out data reduction modelling, detects outliers, selects relevant variables, classification, prediction and unmixing (Wold, Esbensen & Geladi, 1987); where in this case, PCA is used for data simplification, outliers detection and variables classification.

1.2 Problem Statement

The main challenge encountered by the biopharmaceutical industries regarding the production of monoclonal antibody from mammalian cell line is low productivity. Tremendous efforts have been made to increase the specific productivity of mammalian cell lines in the past few decades. The effort to maximum the specific productivity of mammalian cell lines can be achieved by understanding the mechanism of protein synthesis to enable cell engineering process that can improve the productivity. To date, huge data have been collected to study the correlation of specific productivity and cellular characteristics at physiological and molecular levels in order to establish the relationship between specific productivity and the characteristics. Often, the relationships are established by considering only one characteristic at a time by using either Regression analysis (David R. Lloyd et. al., 2000) or Pearson correlation coefficient analysis (Edros et. al., 2013) and the results showed a limited understanding on how the cellular markers correlate to the specific productivity. Since the production of monoclonal antibody is resulted from numbers of interrelated processes that occur simultaneously at cellular level that can be reflected by the physiological states of the cells including cell size, cell volume, cell density, biomass and total intracellular protein, determination of correlation should consider all these states at a time. This can be achieved through the application of multivariate analysis.